UNIVERSIDADE FEDERAL DE SANTA CATARINA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Códigos de Subespaço aplicados a Codificação de Rede

Dissertação submetida à
Universidade Federal de Santa Catarina
como parte dos requisitos para a obtenção
do grau de Mestre em Engenharia Elétrica

Roberto Wanderley da Nóbrega

Florianópolis, 7 de agosto de 2009.

CÓDIGOS DE SUBESPAÇO APLICADOS A CODIFICAÇÃO DE REDE

Roberto Wanderley da Nóbrega

Esta Dissertação foi julgada adequada para a obtenção do título de Mestre em Engenharia Elétrica, área de concentração Comunicações e Processamento de Sinais, e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Santa Catarina.

Bartolomeu F. Uchôa Filho, Ph.D.
Orientador

Roberto de Souza Salgado, Ph.D.
Coordenador do Programa de Pós-Graduação em Engenharia Elétrica

Banca examinadora

Leonardo Silva Resende, D.Sc.
Presidente

Cecilio José Lins Pimentel, Ph.D.

Celso Melchiades Doria, Ph.D.

Mario de Noronha Neto, D.Sc.

Em memória de
Osmundo Wanderley da Nóbrega (1904–1998)
Águeda Ferreira do Amarante (1914–2006)

Agradecimentos

Desejo expressar meu reconhecimento a todos que, de uma maneira ou outra, colaboraram na realização deste trabalho, em especial

a Bartolomeu Ferreira Uchôa Filho, pela excelente orientação e constante estímulo durante todo o mestrado; e pela extensiva revisão do texto, fornecendo críticas e sugestões sempre construtivas;

a Carlos Aurélio Faria da Rocha e Leonardo Silva Resende, pelo incentivo e orientação acadêmica; e pelo contínuo esforço na manutenção e melhoria da qualidade do Grupo de Pesquisa em Comunicações (GPqCom);

a Andrei Piccinini Legg, Bruno Sens Chang, João Luiz Rebelatto, Wilson Leonel Enriquez Lopez e demais colegas do GPqCom, pelas valiosas discussões e sugestões no decorrer do trabalho e pelos momentos de confraternização;

a Adaucto Wanderley da Nóbrega e Maria Luiza Amarante da Nóbrega, meus queridos pais, pelo estímulo e apoio incondicional desde a primeira hora; e pela paciência com que se portam diante dos meus erros e alegria que sentem perante meus acertos;

a Maria Isabel Amarante da Nóbrega Wolff e Adaucto Wanderley da Nóbrega Junior, meus estimados irmãos, pelos exemplos de dedicação e caráter que sempre nortearam meu caminho;

a Alvina da Silva, pela bondade, ternura e zelo com que me tratou durante toda a minha existência;

a Melina de Andrade Silveira, minha amada, pelo apoio, carinho e compreensão e por ser a alegria da minha vida;

a Manssur Gustavo Cassias Pereira, Lucas Barcelos de Oliveira, Renato Herartt, Francisco Antônio Machado da Silva, Fernanda de Pinho e demais amigos, pela afeição e companheirismo presentes em todos os momentos;

à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo auxílio financeiro e eficiência no apoio à pesquisa bibliográfica através de seu portal de periódicos.

Resumo da Dissertação apresentada à UFSC como parte dos requisitos necessários para obtenção do grau de Mestre em Engenharia Elétrica

CÓDIGOS DE SUBESPAÇO APLICADOS A CODIFICAÇÃO DE REDE

Roberto Wanderley da Nóbrega

7 de agosto de 2009

Orientador: Bartolomeu F. Uchôa Filho, Ph.D.
Área de concentração: Comunicações e Processamento de Sinais
Palavras-chave: codificação de rede, códigos de subespaço, controle de erros, multidifusão, não-coerência, códigos de bloco, espaço projetivo, construção multinível
Número de páginas: xvii + 147

Este trabalho aborda codificação de subespaço e sua aplicação no controle de erros em codificação de rede. É assumida uma rede de comunicação com um transmissor e vários receptores, isto é, um cenário multidifusão. O trabalho inicia revisando os principais resultados em codificação de rede multidifusão livre de erros. O foco é então mudado para codificação de rede externa, na qual a comunicação entre o nó fonte e um dado nó destino é modelada por um canal matricial sobre corpos finitos. É mostrado como esse enfoque pode ser usado para se lidar tanto com a não-coerência (desconhecimento da topologia e das operações realizadas pelos nós internos) quanto com eventuais erros nos enlaces da rede. Em particular, na ausência de erros, a não-coerência é contornada através da transmissão de subespaços vetoriais pela rede, visto que esses são preservados qualquer que seja o canal linear. Agora, na presença de erros, pode-se limitar os possíveis subespaços transmitidos a um subconjunto particular de todos os subespaços vetoriais, definindo, assim, um código de subespaço. Até agora, atenção foi dada apenas a códigos de subespaço *one-shot*, isto é, códigos nos quais o canal de comunicação supracitado é utilizado apenas uma vez. Aqui, é explorada a ideia de usar o canal mais de uma vez e códigos de subespaço *multishot* são investigados.

Abstract of Dissertation presented to UFSC as a partial fulfillment of the requirements for the degree of Master in Electrical Engineering

SUBSPACE CODES FOR NETWORK CODING

Roberto Wanderley da Nóbrega

August 7th, 2009

Advisor: Bartolomeu F. Uchôa Filho, Ph.D.
Area of concentration: Communications and Signal Processing
Keywords: network coding, subspace codes, error control, multicast, non-coherent, block codes, projective space, multilevel construction
Number of pages: xvii + 147

This work addresses subspace coding and its application to error control in network coding. A communication network with one transmitter and many receivers (that is, a multicast scenario) is assumed. The work begins by reviewing the main results in error-free multicast network coding. It then moves its focus to outer network coding, in which the communication between the source and a given sink node is modeled by a finite-field matricial channel. It is shown how this approach can be used to deal with both the non-coherence (unknown inner node operations and network topology) and the network link errors. Particularly, in the absence of errors, the non-coherence is circumvented with the transmission of vector subspaces over the network, since those are preserved regardless of the linear channel. Now, if errors are present, one can limit the possible transmitted subspaces to a particular subcollection of all vector subspaces, thus defining a subspace code. So far, attention has been given to one-shot subspace codes, that is, codes that use the aforementioned communication channel just once. Here, the idea of using the channel more than once is explored and the so called multishot subspace codes ared investigated.

Sumário

Simbologia

Símbolo	Descrição	Ref.
$\mathcal{B} \subseteq \mathcal{A}$	$\mathcal{B}$ está contido em $\mathcal{A}$, ou seja, $\mathcal{B}$ é um subconjunto de $\mathcal{A}$.	–
$\mathcal{B} \subset \mathcal{A}$	$\mathcal{B}$ é um subconjunto de $\mathcal{A}$, mas $\mathcal{B} \neq \mathcal{A}$.	–
$\mathcal{A} \times \mathcal{B}$	Produto cartesiano do conjunto $\mathcal{A}$ com o conjunto $\mathcal{B}$.	–
$\mathcal{A}^n$	n-ésima potência cartesiana do conjunto $\mathcal{A}$, isto é conjunto de todas as n-tuplas com componentes em $\mathcal{A}$.	–
$\mathcal{A}^{m \times n}$	Conjunto de todas as matrizes de dimensão $m \times n$ com elementos em $\mathcal{A}$.	–
$\lvert\mathcal{A}\rvert$	Cardinalidade do conjunto $\mathcal{A}$.	–
$2^{\mathcal{A}}$	Conjunto das partes de $\mathcal{A}$, isto é, conjunto de todos os subconjuntos de $\mathcal{A}$.	–
$\forall x \in \mathcal{X} : P$	Quantificação lógica universal. Interpretação: para todo x no conjunto $\mathcal{X}$ a afirmação P é válida.	–

Símbolo	Descrição	Ref.
$\exists x \in \mathcal{X} : P$	Quantificação lógica existencial. Interpretação: existe x no conjunto $\mathcal{X}$ tal que a afirmação P é válida.	–
$\mathbb{N}$	Números naturais, incluindo o zero.	–
$\mathbb{Z}$	Números inteiros.	–
$\mathbb{Z}_q$	Anel finito dos inteiros módulo n. Alternativamente denotado por $\mathbb{Z}/n\mathbb{Z}$.	–
$\mathbb{F}_q$	Corpo finito com q elementos. Também chamado de corpo de Galois. Alternativamente denotado por $\mathrm{GF}(q)$.	p. 115
$R[X]$	Anel polinomial. Conjunto de todos os polinômios em X com coeficientes no anel comutativo R.	p. 115
$R(X)$	Corpo de funções racionais. Conjunto de todos os quocientes de polinômios em X com coeficientes no anel comutativo R.	–
$\binom{m}{k}_q$	Binomial gaussiana. Também chamada de q-binomial.	p. 121
$f : \mathcal{X} \longrightarrow \mathcal{Y}$ $x \longmapsto y$	Mapeamento ou função com domínio $\mathcal{X}$ e contradomínio $\mathcal{Y}$. Associa o elemento $x \in \mathcal{X}$ ao elemento $y \in \mathcal{Y}$, isto é, $f(x) = y$.	–
$f(\mathcal{X}')$	Imagem de $\mathcal{X}'$, em que $\mathcal{X}' \subseteq \mathcal{X}$ é um subconjunto do domínio de f. Dada por $f(\mathcal{X}') = \{f(x) : x \in \mathcal{X}'\}$.	
$f \circ g$	Composição de funções, definida por $(f \circ g)(x) = f(g(x))$.	–
$\mathrm{In}(v)$	Conjunto dos canais de entrada do nó v.	p. 111
$\mathrm{Out}(v)$	Conjunto dos canais de saída do nó v.	p. 111
$\mathcal{P}(W)$	Espaço projetivo de W.	p. 120
$\mathcal{P}(W, k)$	Grassmanniano k-dimensional de W.	p. 120
$\dim V$	Dimensão do espaço vetorial V.	–

Símbolo	Descrição	Ref.
$V^{\perp}$	Subespaço ortogonal a V.	p. 121
$\det \mathbf{M}$	Determinante da matriz $\mathbf{M}$.	–
$\operatorname{rank} \mathbf{M}$	Posto da matriz $\mathbf{M}$.	–
$\operatorname{rankdef} \mathbf{M}$	Deficiência de posto da matriz $\mathbf{M}$.	–
$\Pr[A]$	Probabilidade do evento A.	–
$\lfloor x \rfloor$	Maior inteiro menor ou igual a x. Função *floor*.	–
$\lceil x \rceil$	Menor inteiro maior ou igual a x. Função *ceiling*.	–

CAPÍTULO 1

Introdução

Tradicionalmente, as redes de comunicação adotam a técnica de **roteamento**, que consiste na escolha de caminhos adequados para a informação que trafega pela rede. Nesse esquema, cada nó da rede funciona como um *comutador*: apenas *seleciona, replica e repassa* o que recebe (Figura 1.1). Em outras palavras, os dados são considerados *unidades atômicas imutáveis*.

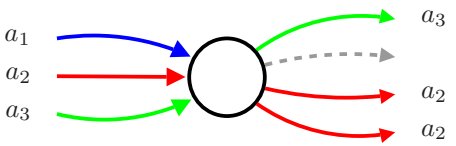


Figura 1.1: Nó da rede operando como um comutador.

A área de **codificação de rede** (do inglês *network coding*), introduzida no ano 2000 por Ahlswede, Cai, Li & Yeung em [1], confronta esse paradigma. Ao invés de limitar a operação dos nós à função de um comutador, nesse novo contexto é permitido que cada nó opere plenamente como um *codificador*: os dados na saída podem ser uma combinação arbitrária dos dados na entrada do nó (Figura 1.2). Assim, ocorre o *processamento* da informação que flui pela rede. Uma vez

que todo codificador pode operar como um comutador, roteamento é apenas um caso particular de codificação de rede.

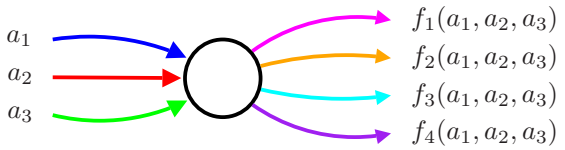
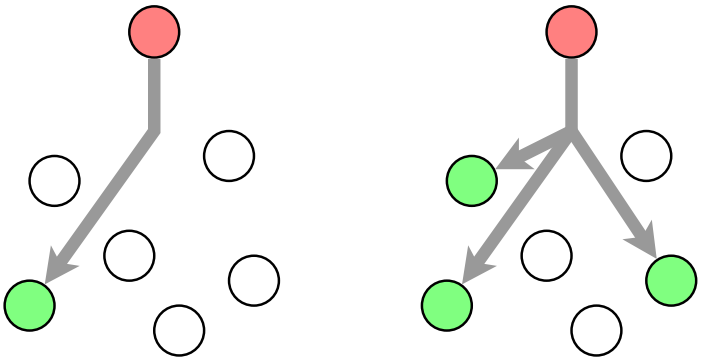


Figura 1.2: Nó da rede operando como um codificador.

Fluxo de informação em redes

O problema mais geral do fluxo de informação em redes associa a cada possível par de nós da rede uma informação diferente a ser transmitida; ainda mais, é permitido que haja dependência estatística entre tais dados. Entretanto, dois cenários bastante particulares são de interesse tanto teórico quanto prático.



(a) Unidifusão. **(b)** Multidifusão.

Figura 1.3: Representação esquemática de dois cenários particulares.

O primeiro é o caso **unidifusão**, do inglês *unicast*, no qual existe apenas um transmissor (nó fonte) e um receptor (nó destino). Já no caso **multidifusão**, do inglês *multicast*, existe um único transmissor que deseja enviar a mesma informação a cada um dos nós pertencentes a um determinado conjunto de receptores. Esses cenários estão esquematizados na Figura 1.3.

Problema do fluxo máximo

Para o caso particular de unidifusão, *roteamento é suficiente*. De fato, esse problema é equivalente ao **problema do fluxo máximo** em redes, estudado na década de 1950 por Elias, Feinstein & Shannon em [8] e por Ford Jr. & Fulkerson em [13], o qual considera o tráfego de elementos de um ponto a outro de uma rede (*e.g.*, redes de tráfego veicular, de encanamento hidráulico ou de energia elétrica).

O problema do fluxo máximo é abordado brevemente no Apêndice A.1. O resultado fundamental é o *Teorema maxflow-mincut*, que afirma que, dados quaisquer dois nós s e t da rede, o valor do máximo fluxo entre s e t coincide com o valor do mínimo corte entre s e t. Adaptado ao caso de uma rede de comunicação com canais de mesma capacidade, o Teorema maxflow-mincut afirma o seguinte.

> *O número mínimo de canais que, quando removidos, separam o nó fonte s do nó destino t é igual ao número máximo de caminhos disjuntos de s a t.*

Como exemplo, considere a Figura 1.4, em que o nó fonte s deseja transmitir informação ao nó destino t. Suponha que cada canal da rede seja capaz de transportar até um bit. (Na figura, cada canal é representado por uma aresta de um nó a outro.) A figura indica o corte mínimo, cujo valor é 3. Portanto, o Teorema maxflow-mincut afirma que existem 3 caminhos disjuntos entre s e t e não mais que isso. Assim, o melhor a ser feito é rotear 3 bits, w_1, w_2, w_3, por tais caminhos, como indicado na figura.

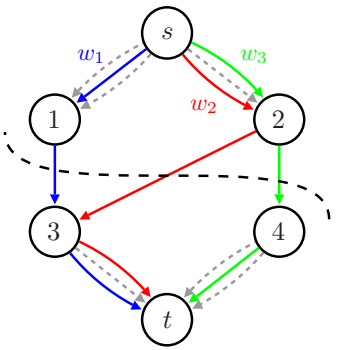


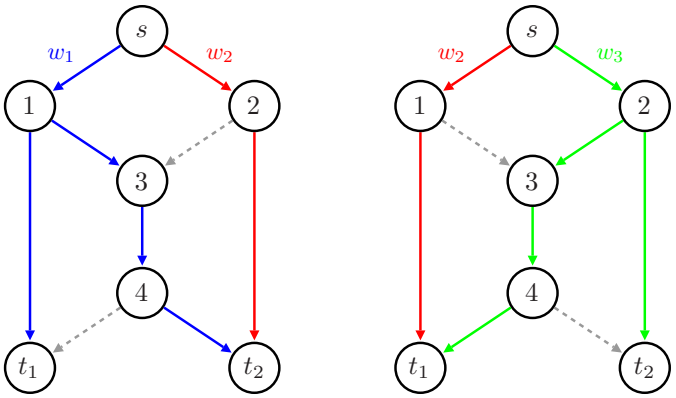
Figura 1.4: Unidifusão em uma rede de comunicação através de roteamento.

Pode esse enfoque ser generalizado a outros tipos de conexão sem perda de desempenho? A resposta é negativa: informação não precisa ser tratada da mesma maneira que veículos ou fluidos; informação pode ser replicada, combinada, processada.

Rede borboleta

Considere o "exemplo padrão" utilizado na literatura de codificação de rede: multidifusão na **rede borboleta**. A rede borboleta, introduzida em [1], é mostrada nas figuras a seguir. Nela, cada canal é capaz de transmitir um único bit por unidade de tempo, instantaneamente e livre de erros; o objetivo é a transmissão de informação do nó fonte s para os nós destino t_1 e t_2.

Se apenas roteamento é permitido, tal tarefa seria executada como na Figura 1.5, em que os recursos da rede são compartilhados no tempo. No primeiro instante o nó t_1 recebe apenas o bit w_1, enquanto o nó t_2 recebe os bits w_1 e w_2. No segundo instante o nó t_1 recebe os bits w_2 e w_3, enquanto o nó t_2 recebe apenas o bit w_3. O resultado é a transmissão de três bits em dois instantes de tempo, isto é, uma taxa de *um bit e meio* por instante de tempo.



(a) Primeiro instante. **(b)** Segundo instante.

Figura 1.5: Multidifusão na rede borboleta utilizando roteamento.

Em contraste, utilizando codificação de rede, é possível transmitir *dois bits* por instante de tempo, de acordo com a Figura 1.6. O nó 3 calcula e transmite a soma módulo 2 (equivalente à operação de XOR)

dos bits que recebe. O nó t_1, que recebe os bits w_1 e $w_1 \oplus w_2$, consegue decodificar w_2 calculando $w_1 \oplus (w_1 \oplus w_2) = w_2$. Analogamente, o nó t_2 também é capaz de decodificar w_1 e w_2.

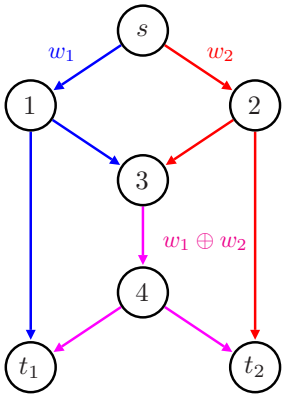


Figura 1.6: Multidifusão na rede borboleta utilizando codificação de rede.

Assim, existe um benefício, em termos de taxa de transmissão, quando se permite o processamento de informação pelos nós intermediários da rede, justificando o uso de codificação de rede.

Codificação de rede multidifusão

Esta dissertação se limita ao cenário multidifusão. Ainda que restritivo, esse caso é um dos mais estudados e fundamentais dentro da área de codificação de rede. Além disso, muitas aplicações práticas (tais como as *redes de distribuição de conteúdo* [20]) se encaixam no modelo. Codificação de rede obteve diversas conquistas no cenário multidifusão. Dentre elas, destacam-se as seguintes:

(*i*) A determinação de *condições necessárias e suficientes* para que a comunicação multidifusão possa ser bem-sucedida. Esse resultado foi obtido no trabalho original [1] de Ahlswede *et al.*

(*ii*) A *suficiência da codificação de rede linear* para a comunicação multidifusão (*i.e.*, as funções f_i ilustradas na Figura 1.2 podem ser limitadas a mapeamentos lineares sem perda de desempenho). Esse afirmação foi provada primeiramente por Li, Yeung & Cai em [39] e posteriormente, sob uma nova ótica, por Koetter & Médard [34, 35]. Infelizmente, o resultado não se estende a perfis

de conexão mais gerais, como mostrado em [7] por Dougherty, Freiling & Zeger.

(*iii*) O desenvolvimento de *algoritmos em tempo polinomial* para o projeto de códigos de rede lineares multidifusão, por Jaggi *et al.* em [29] e Sanders *et al.* em [44] (trabalhos depois combinados em [30]). Outros algoritmos também foram obtidos mais tarde por Harvey *et al.* em [23, 22].

(*iv*) A ideia da *codificação de rede linear aleatória*, na qual os mapeamentos lineares são escolhidos aleatoriamente. Proposto por Ho *et al.* em [24, 26], esse método atinge o desempenho ótimo na multidifusão com alta probabilidade, desde que o tamanho do alfabeto utilizado na comunicação seja suficientemente grande.

Problemas em codificação de rede

Uma das premissas originais da teoria da codificação de rede em [1] é a ausência de erros de transmissão nos enlaces que unem os nós da rede de comunicação. A eliminação dos erros seria obtida *canal a canal*, através da aplicação de *códigos corretores de erro* em camadas inferiores da rede. Em outras palavras, havia uma *separação* entre codificação de rede e codificação de canal.

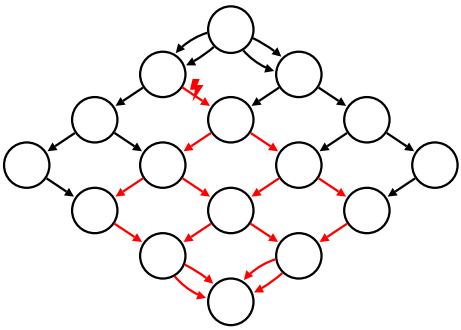


Figura 1.7: Propagação de erros em uma rede (retirado de [37]).

Contudo, além dos erros de transmissão nos canais, redes de comunicação estão sujeitas a outros tipos de adversidades, tais como perda de informação devido a atrasos ou congestionamentos, nós e enlaces saindo do ar e presença de agentes maliciosos no sistema. Em adição,

codificação de rede é especialmente suscetível ao fenômeno de *propagação de erros*, ilustrado na Figura 1.7: a adulteração da informação transmitida por um único canal pode contaminar todos os dados recebidos por um determinado nó, algo que não ocorre com roteamento.

Outro problema de ordem prática presente em diversas redes de comunicação é o desconhecimento, *a priori*, tanto da topologia quanto do código de rede sendo utilizado.

Controle de erros em codificação de rede

Em [2], Cai & Yeung abordam os campos de codificação de rede e codificação de canal em conjunto pela primeira vez. No trabalho, expandido em [55, 3], a teoria tradicional dos códigos corretores de erro é generalizada para redes de comunicação multidifusão. A ideia de **controle de erros em codificação de rede** é distribuir redundância pelos diversos canais, de modo que a informação gerada pelo nó fonte possa ser decodificada por todos os nós destino mesmo na presença de erros. Nessa linha ressaltam-se também [57, 58], de Zhang e [53], de Yang & Yeung.

Codificação de rede não-coerente

Em [5], Chou, Wu & Jain desenvolvem um esquema simples que permite a decodificação da informação recebida sem a necessidade do conhecimento antecipado do código de rede utilizado, bastando esse ser linear. É criada a área conhecida como **codificação de rede não-coerente**. Essa proposta, aliada ao método de codificação linear aleatória, permite um funcionamento totalmente descentralizado do sistema; a topologia da rede pode até mesmo ser variante no tempo. Esse mesmo trabalho ainda apresenta o **modelo de gerações**, método prático para se lidar com o *assincronismo* do sistema, presente em praticamente todas as redes de pacotes.

Canal de comunicação matricial

Codificação de rede pode ser separada entre **codificação de rede interna**, que lida com a escolha das funções f_i na Figura 1.2, para todos os nós da rede e **codificação de rede externa**, que lida com a escolha dos dados injetados na rede pelo nó fonte e a decodificação dos dados colhidos pelos nós destino.

A Figura 1.8 mostra o **canal de comunicação matricial**, que modela a comunicação na presença de erros entre o nó fonte e um dado nó destino em um cenário multidifusão. O nó fonte transmite a matriz $\mathbf{X}$; essa sofre uma transformação multiplicativa (matriz $\mathbf{G}$), que representa o código de rede interno, seguida de uma perturbação aditiva (matriz $\mathbf{Z}'$), que representa os erros; o nó destino em consideração recebe o resultado, a matriz $\mathbf{Y}$.[*] O canal matricial se aplica tanto à codificação de rede síncrona quanto ao modelo de gerações de Chou *et al.*

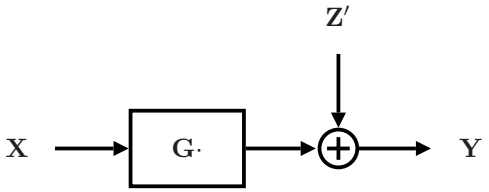


Figura 1.8: Canal de comunicação matricial.

Os problemas de controle de erros em redes e de codificação de rede não-coerente podem ser abordados em conjunto através do canal de comunicação matricial. De fato, a não-coerência se reflete no *desconhecimento da matriz* $\mathbf{G}$, enquanto os erros, como já dito, são modelados pela matriz $\mathbf{Z}'$. Assim, ambos os problemas são resolvidos limitando-se a escolha das possíveis matrizes transmitidas $\mathbf{X}$ a um determinado conjunto $\mathcal{X}$, denominado de **código matricial**.

Nessa linha, destacam-se os seguintes trabalhos:

(*i*) [32, 33], de Koetter & Kschischang, que introduz os chamados *códigos de subespaço* para o controle de erros em codificação de rede multidifusão não-coerente. Códigos de subespaço serão descritos em breve na presente introdução.

(*ii*) [46, 47], de Silva & Kschischang, que considera controle de erros em codificação de rede multidifusão coerente e não-coerente, determinando métricas adequadas para ambos os casos. O trabalho também conclui que a separação entre codificação de rede interna e externa, intrínseca ao caso não-coerente, pode ser feita sem perda de otimalidade no caso coerente, ou seja, pode-se primeiro projetar o código de rede interno de modo a solucionar

[*]Os elementos de tais matrizes pertencem a um determinado corpo finito (*cf.* Apêndice A.3), como será visto nos próximos capítulos.

o problema multidifusão sem erros (determinando, assim, a matriz $\mathbf{G}$) e, após isso, projetar o código de rede externo (com o conhecimento de $\mathbf{G}$) para o controle de erros.

(*iii*) [41], de Montanari & Urbanke e [49, 50], de Silva, Koetter & Kschischang. Enquanto os trabalhos mencionados nos pontos anteriores consideram um modelo de erro *adversário*, esses supõem um modelo de erro *estocástico* e calculam a *capacidade de informação* do canal matricial.

Códigos de subespaço

Suponha, por um instante, que a perturbação aditiva $\mathbf{Z}'$ não exista, de modo que $\mathbf{Y} = \mathbf{G} \cdot \mathbf{X}$ no canal matricial. Em [33], os autores observam que, sob essa circunstância, o *subespaço vetorial* gerado pelas linhas da matriz $\mathbf{X}$ é preservado e sugerem um método de se lidar com a não-coerência: transmitir a informação não diretamente nos elementos de $\mathbf{X}$, mas sim através de subespaços vetoriais. Agora, se a perturbação aditiva $\mathbf{Z}'$ está presente, pode-se limitar os subespaços vetoriais transmitidos a um subconjunto particular de todos os possíveis subespaços vetoriais. Tal escolha pode ser guiada por uma métrica chamada de **distância de subespaço**, que, de acordo com [33], é adequada ao problema em questão. Ao fazer isso se está definindo um **código de subespaço**.

O objeto matemático composto por todos os subespaços vetoriais de um dado espaço vetorial é chamado de **espaço projetivo**. Munido da distância de subespaço, o espaço projetivo torna-se um *espaço métrico*. Assim, os códigos de subespaço nada mais são que *códigos corretores de erro no espaço projetivo*. Além de [33], que define e estuda codificação no espaço projetivo visando sua aplicação no controle de erros em codificação de rede, aponta-se também os trabalhos [12], de Etzion & Vardy e [18], de Gabidulin & Bossert, nos quais outros resultados fundamentais no assunto são obtidos.

Códigos de subespaço *multishot*

Até então, os códigos de subespaços considerados na literatura utilizam o canal matricial apenas uma vez. O uso do canal matricial *repetidas vezes* foi sugerido no trabalho original [33], mas a ideia não foi levada adiante. Na presente dissertação são definidos *códigos de bloco* sobre o espaço projetivo, isto é, códigos nos quais a informação é codificada

em *sequências* de subespaços. Tais códigos são chamados de **códigos de subespaço *multishot***.

Organização do trabalho

O **Capítulo 1** ofereceu uma breve introdução às áreas de codificação de rede e codificação de subespaço. O restante desta dissertação é organizado como segue. O **Capítulo 2** estuda a *codificação de rede multidifusão* em seus fundamentos e aborda importantes resultados na área. O **Capítulo 3** trata da *codificação de rede externa* através do canal matricial e os problemas da não-coerência e do controle de erros são discutidos. O **Capítulo 4** apresenta os *códigos de subespaço*: definições, propriedades, limitantes e sua aplicação no controle de erros em codificação de rede não-coerente são abordados. O **Capítulo 5** introduz os *códigos de subespaço multishot* e é considerado a contribuição deste trabalho. O **Capítulo 6** conclui a dissertação, sumarizando as contribuições realizadas e sugerindo futuras investigações na área.

CAPÍTULO 2

Codificação de rede multidifusão

Como dito no Capítulo 1, um dos cenários de codificação de rede mais fundamentais, mais estudados e para o qual se obteve diversos resultados importantes é o caso *multidifusão*. Nesse cenário existe um único nó—o nó fonte—com acesso a toda a informação e essa deve ser distribuída em sua íntegra para cada um dos nós pertencentes a um determinado conjunto de nós—os nós destino.

A presente dissertação se limita a esse caso, cuja fundamentação é abordada no atual capítulo. Inicia-se introduzindo o modelo da rede e a definição de códigos de rede. São então apresentados resultados sobre a máxima taxa alcançada na multidifusão em função da topologia da rede. Codificação de rede linear e sua suficiência no problema multidifusão são abordados em seguida. Além disso, o capítulo disserta brevemente sobre o projeto de códigos de rede, incluindo *codificação de rede linear aleatória*. Por fim, são apresentados *códigos de rede convolucionais*, que consideram atrasos de transmissão nos canais da rede. Erros nos canais de comunicação são desconsiderados até o Capítulo 3, quando, enfim, serão incorporados ao modelo.

O conteúdo deste capítulo é baseado principalmente nos trabalhos de Yeung *et al.* [56], Fragouli & Soljanin [17, 16] e Koetter & Médard [35].

2.1 Modelo da rede

Inicialmente são apresentadas as definições matemáticas de uma *rede de comunicação* e de uma *conexão multidifusão* sobre essa rede. Tais definições serão adotadas durante todo o restante deste trabalho. Neste momento, recomenda-se a leitura do Apêndice A.1 para a nomenclatura e notação a respeito de grafos.

DEFINIÇÃO. Uma **rede de comunicação** é definida por um *grafo composto direcionado* $(\mathcal{V}, \mathcal{E})$, em que

(i) os *vértices* $v \in \mathcal{V}$ representam os **nós** da rede e

(ii) as *arestas* $e \in \mathcal{E}$ representam os **canais** da rede.

Além disso, é exigido que o grafo seja *acíclico*. □

Alguns comentários sobre esse modelo se fazem adequados.

Grafo composto. Nesta dissertação é adotado um modelo de canais unitários, ou seja, cada canal transporta apenas um único símbolo de informação de um mesmo alfabeto; em compensação, o grafo é composto, o que permite canais paralelos entre dois nós. Um modelo alternativo, algumas vezes utilizado, considera o grafo simples mas atribui pesos a cada canal, representando a respectiva capacidade do mesmo. Na prática, os dois modelos são equivalentes, visto que as capacidades de cada canal podem ser aproximadas por múltiplos de uma unidade comum; adotar um ou outro modelo é, portanto, questão de comodidade matemática.

Grafo direcionado. Um grafo direcionado modela uma rede com canais de comunicação unidirecionais. No caso bidirecional, modelado por um grafo não-direcionado, um dado canal $e \in \mathcal{E}$ pode simultaneamente transportar fluxos opostos, desde que a soma desses não exceda a capacidade do canal. Esta dissertação, no entanto, se limita ao caso direcionado. Uma bibliografia abrangente de codificação de rede em redes não-direcionadas pode ser encontrada em [40].

Grafo acíclico. Redes acíclicas são mais fáceis de se tratar matematicamente: essa restrição permite que os nós sejam parcialmente

ordenados na chamada *ordem de codificação* (ou ordem *upstream-to-downstream*, conhecida na teoria dos grafos como *ordem ancestral*), com a propriedade de que um nó $u \in \mathcal{V}$ precede um nó $v \in \mathcal{V}$ se e somente se existe um caminho direcionado de u até v.[*] Isto, por sua vez, permite que se considere um modelo de transmissões instantâneas livre de problemas de inconsistência e não-causalidade. A Seção 2.7, mais adiante, estuda o problema de codificação de rede em redes cíclicas.

Definição. Uma **conexão multidifusão** sobre uma dada rede de comunicação $(\mathcal{V}, \mathcal{E})$ é definida por um par $(s, \mathcal{T})$, em que

(*i*) $s \in \mathcal{V}$, tal que $\mathrm{In}(s) = \emptyset$, é o **nó fonte** e

(*ii*) $\mathcal{T} \subseteq \mathcal{V}$ é o conjunto dos **nós destino** da conexão.

Adota-se a notação $\mathcal{V}^*$ para o conjunto $\mathcal{V} - \{s\}$ contendo todos os nós não-fonte da rede. □

Exemplo 2.1. A *rede borboleta*, introduzida no Capítulo 1, é apresentada novamente na Figura 2.1. A rede é definida pelo conjunto dos nós

$$\mathcal{V} = \{s, 1, 2, 3, 4, t_1, t_2\}$$

(aí dispostos em uma ordem de codificação) e pelo conjunto dos canais

$$\mathcal{E} = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8, e_9\},$$

em que $e_1 = (s, 1)$, $e_2 = (s, 2)$, $e_3 = (1, 3)$, $e_4 = (2, 3)$, $e_5 = (1, t_1)$, $e_6 = (2, t_2)$, $e_7 = (3, 4)$, $e_8 = (4, t_1)$ e $e_9 = (4, t_2)$, em conformidade com a figura. A conexão multidifusão considerada é $(s, \mathcal{T})$, em que $\mathcal{T} = \{t_1, t_2\}$. Ou seja, se está interessado na transmissão de informação do nó fonte s para cada um dos nós destino t_1 e t_2. □

[*]Em geral, a ordem de codificação não é única—no entanto, qualquer uma delas conduz ao mesmo resultado. Em adição à ordem de codificação atribuída aos nós da rede, considera-se também um ordenamento dos canais. Esse é totalmente arbitrário, tendo como propósito uma definição consistente dos objetos matemáticos que virão a seguir.

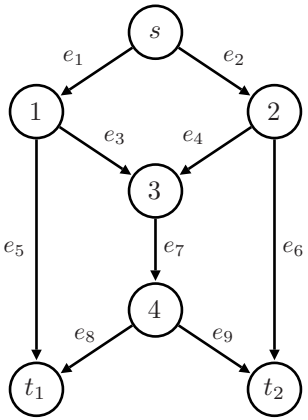


Figura 2.1: Topologia da rede borboleta.

2.2 Códigos de rede

O objetivo de uma conexão multidifusão em uma rede de comunicação é transmitir a mesma informação do nó fonte para cada um dos nós destino. Para atingir tal objetivo utiliza-se um *código de rede*, conceito que será definido nesta seção. Antes disso, o funcionamento do sistema é abordado. O enfoque aqui adotado é similar ao de Koetter & Médard [35].

2.2.1 Funcionamento do sistema

Cada canal $e \in \mathcal{E}$ da rede transporta (instantaneamente e livre de erros) um símbolo a_e de um dado alfabeto finito $\mathcal{A}$. Cada nó $v \in \mathcal{V}$ da rede pode "ler" os símbolos dos canais em $\mathrm{In}(v)$ e "escrever" símbolos nos canais em $\mathrm{Out}(v)$. O processo inicia com a geração de um símbolo de informação w, denominado de *mensagem*, a partir de um alfabeto finito $\mathcal{W}$ qualquer. A evolução do sistema prossegue de acordo com os seguintes passos.[†]

(*i*) O nó fonte s tem acesso à mensagem w e a codifica em $n_s = |\mathrm{Out}(s)|$ símbolos do alfabeto $\mathcal{A}$, dispostos na tupla $\mathbf{x} = \mathbf{a}_{\mathrm{Out}(s)} \in \mathcal{A}^{n_s}$, que são injetados na rede através dos canais em $\mathrm{Out}(s)$.

[†]A notação $\mathbf{a}_\zeta$ representa a tupla $(a_e : e \in \zeta)$, em que $\zeta \subseteq \mathcal{E}$ é um subconjunto qualquer de canais. A ordem dos elementos em $\mathbf{a}_\zeta$ é coerente com o ordenamento adotado para os canais.

(*ii*) Os demais nós $v \in \mathcal{V}^*$ operam, um nó de cada vez e seguindo a ordem de codificação, da seguinte forma: o nó v recebe os símbolos $\mathbf{a}_{\mathrm{In}(v)}$ em seus canais de entrada e, em função desses símbolos, calcula e transmite pelos seus canais de saída os símbolos $\mathbf{a}_{\mathrm{Out}(v)}$. A aciclicidade do grafo garante que todas as entradas já estejam disponíveis na vez de cada nó.

(*iii*) Cada um dos nós destino $t \in \mathcal{T}$ dá um palpite $\hat{w}_t$ sobre a mensagem w, tendo como base os $n_t = |\mathrm{In}(t)|$ símbolos que recebe, dispostos na tupla $\mathbf{y}_t = \mathbf{a}_{\mathrm{In}(t)} \in \mathcal{A}^{n_t}$.

Neste capítulo, o alfabeto da mensagem $\mathcal{W}$ é restrito à forma $\mathcal{A}^h$. Em outras palavras, a mensagem w gerada consiste em h símbolos do alfabeto $\mathcal{A}$, dispostos na tupla $\mathbf{w} = (w_1, \ldots, w_h) \in \mathcal{A}^h$. Tal formato será largado no Capítulo 3, por motivos que serão explicados lá.

2.2.2 Descrição local de códigos de rede

A definição de um código de rede é apresentada a seguir, levando em conta o funcionamento da rede recém descrito.

Definição. Sejam $(\mathcal{V}, \mathcal{E})$ uma rede de comunicação, $(s, \mathcal{T})$ uma conexão multidifusão sobre essa rede, h um inteiro positivo e $\mathcal{A}$ um alfabeto finito. Um **código de rede de taxa h sobre $\mathcal{A}$** é definido por

(*i*) um *mapeamento de codificação da fonte*

$$\begin{array}{rccc} \phi_s: & \mathcal{A}^h & \longrightarrow & \mathcal{A}^{n_s} \\ & \mathbf{w} & \longmapsto & \mathbf{x}, \end{array}$$

(*ii*) um *mapeamento de codificação local*

$$\begin{array}{rccc} f_v: & \mathcal{A}^{|\mathrm{In}(v)|} & \longrightarrow & \mathcal{A}^{|\mathrm{Out}(v)|} \\ & \mathbf{a}_{\mathrm{In}(v)} & \longmapsto & \mathbf{a}_{\mathrm{Out}(v)} \end{array}$$

para cada nó não-fonte $v \in \mathcal{V}^*$ da rede e

(*iii*) um *mapeamento de decodificação*

$$\begin{array}{rccc} \hat{\phi}_t: & \mathcal{A}^{n_t} & \longrightarrow & \mathcal{A}^h \\ & \mathbf{y}_t & \longmapsto & \hat{\mathbf{w}}_t \end{array}$$

para cada nó destino $t \in \mathcal{T}$.

O conjunto $\{f_v : v \in \mathcal{V}^*\}$ contendo todos os mapeamentos de codificação locais define o chamado **código de rede interno**. Já o conjunto $\{\phi_s, \hat{\phi}_t : t \in \mathcal{T}\}$ contendo os mapeamentos de codificação da fonte e de decodificação dos nós destino define o chamado **código de rede externo**. □

Exemplo 2.2. A Figura 2.2 ilustra o conceito de mapeamento de codificação local. Na figura, tem-se $\mathbf{a}_{\mathrm{In}(v)} = (a_1, a_2)$ e $\mathbf{a}_{\mathrm{Out}(v)} = (a_3, a_4, a_5)$.

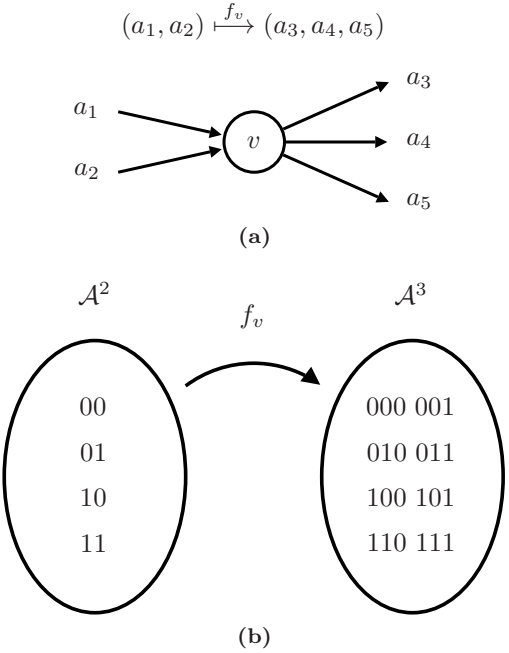


Figura 2.2: Ilustração do conceito de mapeamento de codificação local.

Qualquer mapeamento $f_v : \mathcal{A}^2 \longrightarrow \mathcal{A}^3$ é permitido. Se, por exemplo, $\mathcal{A} = \{0, 1\}$ é o alfabeto binário então existem $8^4 = 4096$ mapeamentos f_v possíveis. □

Nem todo código de rede concretiza o seu objetivo de permitir uma conexão multidifusão com taxa h.

DEFINIÇÃO. Um código de rede é dito ser **bem-sucedido** se $\hat{\mathbf{w}}_t = \mathbf{w}$ para todo $t \in \mathcal{T}$, ou seja, se cada nó destino é capaz de decodificar a mensagem $\mathbf{w}$ corretamente. □

Exemplo 2.3. Considere o funcionamento da rede borboleta como mostrado na Figura 2.3, em que $h = 2$. Considere também o alfa-

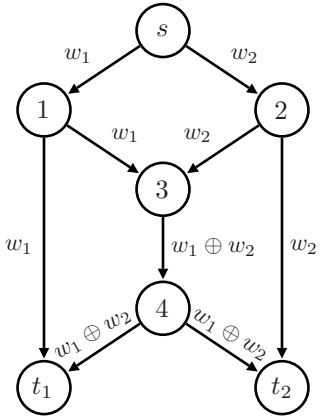


Figura 2.3: Operações realizadas na rede borboleta.

beto $\mathcal{A} = \{0, 1\}$, isto é, cada canal transporta um único bit. Assim sendo, o mapeamento de codificação da fonte é dado por

$$\begin{array}{rccc} \phi_s: & \{0,1\}^2 & \longrightarrow & \{0,1\}^2 \\ & (0,0) & \longmapsto & (0,0) \\ & (0,1) & \longmapsto & (0,1) \\ & (1,0) & \longmapsto & (1,0) \\ & (1,1) & \longmapsto & (1,1), \end{array}$$

os mapeamentos de codificação locais são dados por

$$\begin{array}{rccc} f_1 = f_2 = f_4: & \{0,1\} & \longrightarrow & \{0,1\}^2 \\ & 0 & \longmapsto & (0,0) \\ & 1 & \longmapsto & (1,1), \end{array} \qquad \begin{array}{rccc} f_3: & \{0,1\}^2 & \longrightarrow & \{0,1\} \\ & (0,0) & \longmapsto & 0 \\ & (0,1) & \longmapsto & 1 \\ & (1,0) & \longmapsto & 1 \\ & (1,1) & \longmapsto & 0, \end{array}$$

e os mapeamentos de decodificação são dados por

$$\begin{array}{rccc} \hat{\phi}_{t_1} = & \{0,1\}^2 & \longrightarrow & \{0,1\}^2 \\ & (0,0) & \longmapsto & (0,0) \\ & (0,1) & \longmapsto & (0,1) \\ & (1,0) & \longmapsto & (1,1) \\ & (1,1) & \longmapsto & (1,0), \end{array} \qquad \begin{array}{rccc} \hat{\phi}_{t_2}: & \{0,1\}^2 & \longrightarrow & \{0,1\}^2 \\ & (0,0) & \longmapsto & (0,0) \\ & (0,1) & \longmapsto & (1,0) \\ & (1,0) & \longmapsto & (1,1) \\ & (1,1) & \longmapsto & (0,1). \end{array}$$

De forma mais compacta, pode-se escrever

$$\begin{gathered} \phi_s : (w_1, w_2) \longmapsto (w_1, w_2), \\ f_1 = f_2 = f_4 : a \longmapsto (a, a), \qquad f_3 : (a_1, a_2) \longmapsto a_1 \oplus a_2, \\ \hat{\phi}_{t_1} : (a_1, a_2) \longmapsto (a_1, a_1 \oplus a_2), \qquad \hat{\phi}_{t_2} : (a_1, a_2) \longmapsto (a_1 \oplus a_2, a_1). \end{gathered}$$

Mais adiante, no Exemplo 2.5, será mostrado que o código de rede assim definido é bem-sucedido. □

2.2.3 Descrição global de códigos de rede

Existe uma maneira alternativa mas equivalente de se especificar um código interno de rede, que é através dos chamados *mapeamentos de codificação globais* (*cf.* [56, 17]). Esses mapeiam os símbolos injetados na rede pela fonte diretamente no símbolo que será de fato transportado em cada canal particular.

Definição. O **mapeamento de codificação global** de um canal $e \in \mathcal{E}$ é definido pelo mapeamento que leva $\mathbf{x}$ em a_e, isto é,

$$\begin{array}{rccc} g_e: & \mathcal{A}^{n_s} & \longrightarrow & \mathcal{A} \\ & \mathbf{x} & \longmapsto & a_e. \end{array}$$

Os mapeamentos de codificação globais $\{g_e : e \in \mathcal{E}\}$ são obtidos dos mapeamentos de codificação locais $\{f_v : v \in \mathcal{V}^*\}$ de uma maneira indutiva respeitando a ordem de codificação: para todo $v \in \mathcal{V}$,

$$g_{\mathrm{Out}(v)} = \begin{cases} f_v \circ g_{\mathrm{In}(v)}, & v \neq s, \\ \mathrm{id}, & v = s, \end{cases} \tag{2.1}$$

em que id é a função identidade $\mathcal{A}^{n_s} \longrightarrow \mathcal{A}^{n_s}$, a notação $g_\zeta(\cdot)$ equivale a $(g_e(\cdot) : e \in \zeta)$ e o símbolo $\circ$ denota composição de funções. □

Evidentemente, $g_{\mathrm{Out}(s)}(\mathbf{x}) = \mathbf{x}$. Já se $v \neq s$, tem-se que

$$\mathbf{x} \overset{g_{\mathrm{Out}(v)}}{\longmapsto} \mathbf{a}_{\mathrm{Out}(v)},$$

mas também

$$\mathbf{x} \overset{g_{\mathrm{In}(v)}}{\longmapsto} \mathbf{a}_{\mathrm{In}(v)} \overset{f_v}{\longmapsto} \mathbf{a}_{\mathrm{Out}(v)},$$

o que justifica (2.1).

Exemplo 2.4. Os mapeamentos de codificação globais da rede borboleta são obtidos dos mapeamentos de codificação locais, dados no Exemplo 2.3, através de (2.1). Seguindo a ordem de codificação, tem-se, recursivamente:

Nó s: $g_{(e_1,e_2)}(x_1, x_2) = g_{\mathrm{Out}(s)}(x_1, x_2) = \mathrm{id}(x_1, x_2) = (x_1, x_2)$,

Nó 1: $g_{(e_3,e_5)}(x_1, x_2) = g_{\mathrm{Out}(1)}(x_1, x_2) = f_1(g_{\mathrm{In}(1)}(x_1, x_2)) = f_1(g_{e_1}(x_1, x_2)) = f_1(x_1) = (x_1, x_1)$,

Nó 2: $g_{(e_4,e_6)}(x_1, x_2) = g_{\mathrm{Out}(2)}(x_1, x_2) = f_2(g_{\mathrm{In}(2)}(x_1, x_2)) = f_2(g_{e_2}(x_1, x_2)) = f_2(x_2) = (x_2, x_2)$,

Nó 3: $g_{e_7}(x_1, x_2) = g_{\mathrm{Out}(3)}(x_1, x_2) = f_3(g_{\mathrm{In}(3)}(x_1, x_2)) = f_3(g_{(e_3,e_4)}(x_1, x_2)) = f_3(x_1, x_2) = x_1 \oplus x_2$,

Nó 4: $g_{(e_8,e_9)}(x_1, x_2) = g_{\mathrm{Out}(4)}(x_1, x_2) = f_4(g_{\mathrm{In}(4)}(x_1, x_2)) = f_4(g_{e_7}(x_1, x_2)) = f_4(x_1 \oplus x_2) = (x_1 \oplus x_2, x_1 \oplus x_2)$.

Sumarizando,

$$\begin{aligned} g_{e_1}(x_1, x_2) = g_{e_3}(x_1, x_2) = g_{e_5}(x_1, x_2) = x_1, \\ g_{e_2}(x_1, x_2) = g_{e_4}(x_1, x_2) = g_{e_6}(x_1, x_2) = x_2, \\ g_{e_7}(x_1, x_2) = g_{e_8}(x_1, x_2) = g_{e_9}(x_1, x_2) = x_1 \oplus x_2, \end{aligned}$$

resultado condizente com a Figura 2.3 e com o fato de que $(x_1, x_2) = \phi_s(w_1, w_2) = (w_1, w_2)$. □

Observação. Tanto os mapeamentos de codificação locais $\{f_v : v \in \mathcal{V}^*\}$ quanto os mapeamentos de codificação globais $\{g_e : e \in \mathcal{E}\}$ definem plenamente um código de rede interno. No entanto, é importante ressaltar que, enquanto qualquer escolha de mapeamento local é "permitida" (ou seja, as funções $\{f_v : v \in \mathcal{V}^*\}$ podem ser escolhidas arbitrariamente), nem toda escolha de mapeamento global o é (ou seja, as funções $\{g_e : e \in \mathcal{E}\}$ *não* podem ser escolhidas arbitrariamente). Isto se dá pois os valores de $g_{\mathrm{Out}(v)}(\mathbf{x})$ têm de ser, obrigatoriamente, computáveis a partir dos valores de $g_{\mathrm{In}(v)}(\mathbf{x})$. Por exemplo, na rede borboleta, o nó 1 não tem acesso à segunda coordenada da tupla $\mathbf{x}$ e, portanto, os mapeamentos g_{e_3} e g_{e_5} não podem fazer uso de tal valor.

2.2.4 Funções de transferência

Esquematicamente, pode-se representar o encadeamento do sistema por

$$\mathbf{w} \overset{\phi_s}{\longmapsto} \mathbf{x} \overset{g_{s,t}}{\longmapsto} \mathbf{y}_t \overset{\hat{\phi}_t}{\longmapsto} \hat{\mathbf{w}}_t, \tag{2.2}$$

em que $\phi_s : \mathcal{A}^h \longrightarrow \mathcal{A}^{n_s}$ é o mapeamento de codificação do nó fonte s, $\hat{\phi}_t : \mathcal{A}^{n_t} \longrightarrow \mathcal{A}^h$ é o mapeamento de decodificação do nó destino $t \in \mathcal{T}$ e $g_{s,t} : \mathcal{A}^{n_s} \longrightarrow \mathcal{A}^{n_t}$, denominado de **função de transferência** (de s para t), é um mapeamento tal que

$$\mathbf{y}_t = g_{s,t}(\mathbf{x}).$$

Os mapeamentos de codificação globais facilitam a tarefa de se obter a função de transferência—de fato, $g_{s,t} = g_{\mathrm{In}(t)}$.

Indo mais além, pode-se escrever

$$\hat{\mathbf{w}}_t = h_{s,t}(\mathbf{w}),$$

em que $h_{s,t}$ é definida pela composição $\hat{\phi}_t \circ g_{s,t} \circ \phi_s$. Fica claro que para que um código de rede seja bem-sucedido é necessário e suficiente que a função $h_{s,t}$ seja o mapeamento identidade $\mathcal{A}^h \longrightarrow \mathcal{A}^h$, para cada $t \in \mathcal{T}$.

Exemplo 2.5. No caso da rede borboleta, a função de transferência de s para t_1é dada por

$$g_{s,t_1}(x_1, x_2) = g_{\mathrm{In}(t_1)}(x_1, x_2) = g_{(e_5,e_8)}(x_1, x_2) = (x_1, x_1 \oplus x_2)$$

e a função de transferência de s para t_2 é dada por

$$g_{s,t_2}(x_1, x_2) = g_{\mathrm{In}(t_2)}(x_1, x_2) = g_{(e_6,e_9)}(x_1, x_2) = (x_2, x_1 \oplus x_2),$$

nas quais foram utilizados resultados do Exemplo 2.4. Portanto, os mapeamentos que levam $\mathbf{w}$ para $\hat{\mathbf{w}}_{t_1}$ e $\hat{\mathbf{w}}_{t_2}$ são dados, respectivamente, por

$$\begin{aligned} h_{s,t_1}(w_1, w_2) &= \hat{\phi}_{t_1}(g_{s,t_1}(\phi_s(w_1, w_2))) \\ &= \hat{\phi}_{t_1}(g_{s,t_1}(w_1, w_2)) \\ &= \hat{\phi}_{t_1}(w_1, w_1 \oplus w_2) \\ &= (w_1, w_1 \oplus w_1 \oplus w_2) \\ &= (w_1, w_2), \end{aligned}$$

$$\begin{aligned} h_{s,t_2}(w_1, w_2) &= \hat{\phi}_{t_2}(g_{s,t_2}(\phi_s(w_1, w_2))) \\ &= \hat{\phi}_{t_2}(g_{s,t_2}(w_1, w_2)) \\ &= \hat{\phi}_{t_2}(w_2, w_1 \oplus w_2) \\ &= (w_2 \oplus w_1 \oplus w_2, w_2) \\ &= (w_1, w_2), \end{aligned}$$

em que foram utilizados resultados do Exemplo 2.3. Como h_{s,t_1} e h_{s,t_2} coincidem com o mapeamento identidade, o código de rede é bem-sucedido. □

2.3 Taxas alcançáveis

Um dos principais problemas estudados pela teoria da codificação de rede é a determinação da máxima taxa que pode ser alcançada em uma dada rede de comunicação. Esse problema foi resolvido para o caso multidifusão pela primeira vez no trabalho original de Ahlswede *et al.* [1] através de um enfoque baseado em teoria da informação.

Definição. Dada uma rede de comunicação $(\mathcal{V}, \mathcal{E})$ e uma conexão multidifusão $(s, \mathcal{T})$ sobre essa rede, uma taxa $h \in \mathbb{N}$ é dita ser **alcançável** se, para algum alfabeto finito $\mathcal{A}$, existir código de rede bem-sucedido de taxa h sobre $\mathcal{A}$. □

Observação. Se uma taxa h é alcançável, isto não significa que essa taxa possa ser alcançada com um alfabeto $\mathcal{A}$ de tamanho qualquer. A Figura 2.4, de Yeung *et al.* [56], apresenta um problema de codificação multidifusão no qual a taxa $h = 2$ não pode ser alcançada utilizando um alfabeto binário, mas pode ser utilizando um alfabeto ternário.

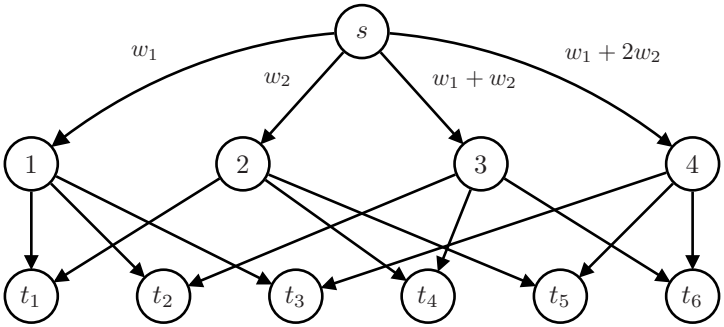


Figura 2.4: Problema multidifusão sem solução utilizando um alfabeto binário.

Para uma discussão sobre o tamanho do alfabeto em códigos de rede, incluindo limitantes superiores e inferiores, veja Rasala Lehman [43, Chapter 2] e Fragouli & Soljanin [17, Chapter 7].

Por um instante, a atenção será direcionada a uma *conexão unidifusão* sobre a rede $(\mathcal{V}, \mathcal{E})$, definida por um par (s, t), em que $s \in \mathcal{V}$ é o nó fonte e $t \in \mathcal{V}$ é o (único) nó destino da conexão. Nesse caso, a maior taxa alcançável é dada pelo valor de um fluxo máximo entre s e t, denotado por $\text{maxflow}(s, t)$, que, pelo Teorema maxflow-mincut (*cf.* Apêndice A.1), é igual ao valor de um corte mínimo entre s e t, denotado por $\text{mincut}(s, t)$. Assim, para alcançar uma taxa $h \in \mathbb{N}$ tal que

$$h \leq \text{mincut}(s, t),$$

roteamento é suficiente: de fato, basta transmitir cada coordenada da mensagem $\mathbf{w} \in \mathcal{A}^h$ por um dos h caminhos disjuntos entre s e t. (A existência de h caminhos disjuntos entre s e t é garantida também pelo Teorema maxflow-mincut.)

Voltando ao caso multidifusão, é intuitivo que qualquer taxa alcançável em uma conexão multidifusão $(s, \mathcal{T})$ não pode ser superior à menor entre as taxas que seriam alcançadas individualmente em cada conexão unidifusão em $\{(s, t) : t \in \mathcal{T}\}$. O que não é intuitivo, mas

igualmente verdadeiro, é a recíproca: se uma taxa é igual ou inferior à menor entre as taxas individualmente alcançáveis em cada conexão unidifusão em $\{(s,t) : t \in \mathcal{T}\}$, então essa taxa é alcançável também na conexão multidifusão $(s, \mathcal{T})$. Para isto, porém, roteamento é insuficiente e codificação de rede se faz necessária.

Em outras palavras, o mínimo entre os desempenhos alcançados individualmente (com a rede utilizando todos os seus recursos em prol de um nó destino específico) também pode ser alcançado em conjunto através do uso de um código de rede, desde que se permita um alfabeto suficientemente grande. O teorema a seguir, de Ahlswede *et al.* [1], formaliza esse resultado.

Teorema 2.1. *Dada uma rede de comunicação* $(\mathcal{V}, \mathcal{E})$ *e uma conexão multidifusão* $(s, \mathcal{T})$ *sobre essa rede, uma taxa* $h \in \mathbb{N}$ *é alcançável se e somente se*

$$h \leq \min\{\text{mincut}(s,t) : t \in \mathcal{T}\}. \tag{2.3}$$

A validade desse teorema é assegurada na Seção 2.5, que também mostra que codificação de rede linear, a ser discutida a seguir, é sempre suficiente para conexões multidifusão.

2.4 Codificação de rede linear

A área de *codificação de rede linear* foi introduzida em [39] por Li, Yeung & Cai. Nessa, os mapeamentos de codificação dos códigos de rede estão limitados a funções lineares. Em outras palavras, os nós da rede efetuam combinações lineares dos símbolos que recebem. Naquele mesmo trabalho, os autores provam que codificação de rede linear é suficiente para o caso multidifusão, isto é, dado qualquer cenário multidifusão, se uma taxa é alcançável então essa taxa é alcançável utilizando-se um código de rede linear.

2.4.1 Códigos de rede lineares

A linearidade aqui adotada será sobre *corpos finitos* (*cf.* Apêndice A.3), de modo que o alfabeto $\mathcal{A}$ é considerado um corpo finito $\mathbb{F}_q$, no qual q é uma potência de primo. Lembrando que um mapeamento f qualquer é linear se

$$f(c_1 x_1 + c_2 x_2) = c_1 f(x_1) + c_2 f(x_2),$$

para todo c_1, c_2, x_1, x_2. Daí a utilização de um corpo finito: essa definição, compulsoriamente, necessita das operações de *produto* e *soma*.

DEFINIÇÃO. Um *código de rede interno* sobre um alfabeto $\mathcal{A} = \mathbb{F}_q$ é dito **linear** se os mapeamentos de codificação locais $\{f_v : v \in \mathcal{V}^*\}$ são todos lineares. Nesse caso, existe uma matriz $\mathbf{F}_v \in \mathbb{F}_q^{|\mathrm{Out}(v)|\times|\mathrm{In}(v)|}$ tal que

$$\mathbf{a}_{\mathrm{Out}(v)} = f_v(\mathbf{a}_{\mathrm{In}(v)}) = \mathbf{F}_v \cdot \mathbf{a}_{\mathrm{In}(v)}$$

para todo $\mathbf{a}_{\mathrm{In}(v)} \in \mathbb{F}_q^{|\mathrm{In}(v)|}$. A matriz $\mathbf{F}_v$ é denominada de **matriz de codificação local** do nó v. □

Uma definição alternativa considera os mapeamentos globais.

DEFINIÇÃO. Um *código de rede interno* sobre um alfabeto $\mathcal{A} = \mathbb{F}_q$ é dito **linear** se os mapeamentos globais $\{g_e : e \in \mathcal{E}\}$ são todos lineares. Nesse caso, existe uma matriz linha $\mathbf{G}_e \in \mathbb{F}_q^{1\times n_s}$ tal que

$$a_e = g_e(\mathbf{x}) = \mathbf{G}_e \cdot \mathbf{x}$$

para todo $\mathbf{x} \in \mathbb{F}_q^{n_s}$. A matriz linha $\mathbf{G}_e$ é denominada de **matriz de codificação global** do canal e. □

As definições são de fato equivalentes, como mostra o teorema a seguir, de Yeung *et al.* [56].

Teorema 2.2. *Os mapeamentos locais $\{f_v : v \in \mathcal{V}^*\}$ são todos lineares se e somente se os mapeamentos globais $\{g_e : e \in \mathcal{E}\}$ são todos lineares.*

Demonstração. Suponha que os mapeamentos locais $\{f_v : v \in \mathcal{V}^*\}$ sejam todos lineares. Para mostrar que os mapeamentos globais $\{g_e : e \in \mathcal{E}\}$ são todos lineares é suficiente provar que $g_{\mathrm{Out}(v)}$ é linear para cada $v \in \mathcal{V}$. Prova-se que o mapeamento $g_{\mathrm{Out}(v)}$ é linear por indução sobre o conjunto $\mathcal{V}$ ordenado de acordo com a ordem de codificação:

(*i*) $g_{\mathrm{Out}(s)} = \mathrm{id}$ é obviamente linear.

(*ii*) Seja $v \in \mathcal{V}$ um nó qualquer e $\mathcal{V}' \subseteq \mathcal{V}$ o conjunto de todos os nós que precedem (estritamente) v na ordem de codificação. Suponha que $g_{\mathrm{Out}(v')}$ seja linear para todo $v' \in \mathcal{V}'$. Então $g_{\mathrm{In}(v)}$ é linear, pois, pela propriedade da ordem de codificação, cada canal

em $\mathrm{In}(v)$ está também em $\mathrm{Out}(v')$, para algum $v' \in \mathcal{V}'$. Como a composição de funções lineares é também uma função linear, a relação (2.1) mostra que $g_{\mathrm{Out}(v)}$ é linear.

Suponha agora que os mapeamentos globais $\{g_e : e \in \mathcal{E}\}$ sejam todos lineares. Seja

$$\mathbf{a}_i = g_{\mathrm{In}(v)}(\mathbf{x}_i) \tag{2.4}$$

para $i = 1, 2$. Se $c_1, c_2 \in \mathbb{F}_q$, tem-se, para todo $v \in \mathcal{V}^*$,

$$\begin{aligned} f_v(c_1\mathbf{a}_1 + c_2\mathbf{a}_2) &\overset{(a)}{=} f_v\left(c_1 g_{\mathrm{In}(v)}(\mathbf{x}_1) + c_2 g_{\mathrm{In}(v)}(\mathbf{x}_2)\right) \\ &\overset{(b)}{=} f_v\left(g_{\mathrm{In}(v)}(c_1\mathbf{x}_1 + c_2\mathbf{x}_2)\right) \\ &\overset{(c)}{=} g_{\mathrm{Out}(v)}(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) \\ &\overset{(d)}{=} c_1 g_{\mathrm{Out}(v)}(\mathbf{x}_1) + c_2 g_{\mathrm{Out}(v)}(\mathbf{x}_2) \\ &\overset{(e)}{=} c_1 f_v\left(g_{\mathrm{In}(v)}(\mathbf{x}_1)\right) + c_2 f_v\left(g_{\mathrm{In}(v)}(\mathbf{x}_2)\right) \\ &\overset{(f)}{=} c_1 f_v(\mathbf{a}_1) + c_2 f_v(\mathbf{a}_2), \end{aligned}$$

em que (a) e (f) seguem de (2.4); (b) e (d) seguem da hipótese de que os mapeamentos g_e (e portanto $g_{\mathrm{In}(v)}$ e $g_{\mathrm{Out}(v)}$) são lineares; e (c) e (e) seguem de (2.1). Portanto, f_v é linear para todo $v \in \mathcal{V}^*$. □

Exemplo 2.6. No Exemplo 2.2, no qual foi considerado o corpo binário $\mathbb{F}_2 = \{0, 1\}$ como o alfabeto $\mathcal{A}$, foi visto que existem 4096 possíveis mapeamentos locais (lineares e não-lineares) para o nó v mostrado na Figura 2.2. Quantos possíveis mapeamentos locais lineares existem nesse mesmo caso? A resposta coincide com o número de matrizes binárias de dimensão 2×3, que é igual a $2^6 = 64$. □

Para códigos de rede lineares, a relação (2.1) se torna

$$\mathbf{G}_{\mathrm{Out}(v)} = \begin{cases} \mathbf{F}_v \cdot \mathbf{G}_{\mathrm{In}(v)}, & v \neq s, \\ \mathbf{I}, & v = s, \end{cases} \tag{2.5}$$

para todo nó $v \in \mathcal{V}$, em que $\mathbf{I}$ é a matriz identidade $n_s \times n_s$ e a notação $\mathbf{G}_\zeta$ equivale à matriz cujas linhas são dadas pelas matrizes $\mathbf{G}_e$, $e \in \zeta$.

De maneira análoga, define-se a linearidade do código externo.

DEFINIÇÃO. Um *código de rede externo* é dito **linear** se os mapeamentos que o definem são todos lineares. Se for esse o caso, definem-se a **matriz de codificação da fonte**, denotada por $\mathbf{\Phi}_s$, e a **matriz de decodificação de um nó destino** $t \in \mathcal{T}$, denotada por $\hat{\mathbf{\Phi}}_t$, como aquelas que representam os mapeamentos ϕ_s e $\hat{\phi}_t$, respectivamente. □

Exemplo 2.7. (Adaptado de [56, Chapter 2]) O código de rede que vem sendo considerado nos exemplos da rede borboleta é, de fato, linear, com $\mathcal{A} = \mathbb{F}_2$. Do Exemplo 2.3 tem-se que as matrizes de codificação locais, juntamente com as matrizes de codificação e decodificação são dadas por

$$\mathbf{\Phi}_s = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

$$\mathbf{F}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \mathbf{F}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

$$\mathbf{F}_3 = \begin{bmatrix} 1 & 1 \end{bmatrix}, \qquad \mathbf{F}_4 = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

$$\hat{\mathbf{\Phi}}_{t_1} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \qquad \hat{\mathbf{\Phi}}_{t_2} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

As matrizes de codificação globais podem ser obtidas imediatamente do Exemplo 2.4, sendo dadas por

$$\mathbf{G}_{e_1} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \qquad \mathbf{G}_{e_2} = \begin{bmatrix} 0 & 1 \end{bmatrix},$$

$$\mathbf{G}_{e_3} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \qquad \mathbf{G}_{e_4} = \begin{bmatrix} 0 & 1 \end{bmatrix},$$

$$\mathbf{G}_{e_5} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \qquad \mathbf{G}_{e_6} = \begin{bmatrix} 0 & 1 \end{bmatrix},$$

$$\mathbf{G}_{e_7} = \begin{bmatrix} 1 & 1 \end{bmatrix}, \qquad \mathbf{G}_{e_8} = \begin{bmatrix} 1 & 1 \end{bmatrix}, \qquad \mathbf{G}_{e_9} = \begin{bmatrix} 1 & 1 \end{bmatrix},$$

resultado que também pode ser alcançado recursivamente a partir das matrizes de codificação locais através da relação (2.5). □

2.4.2 Matrizes de transferência

No caso linear, a cadeia (2.2) pode ser reescrita como

$$\mathbf{w} \overset{\mathbf{\Phi}_s}{\longmapsto} \mathbf{x} \overset{\mathbf{G}_{s,t}\cdot}{\longmapsto} \mathbf{y}_t \overset{\hat{\mathbf{\Phi}}_t}{\longmapsto} \hat{\mathbf{w}}_t, \tag{2.6}$$

em que $\mathbf{\Phi}_s \in \mathbb{F}_q^{n_s \times h}$ é a matriz de codificação do nó fonte s, $\hat{\mathbf{\Phi}}_t \in \mathbb{F}_q^{h \times n_t}$ é a matriz de decodificação do nó destino $t \in \mathcal{T}$ e $\mathbf{G}_{s,t} \in \mathbb{F}_q^{n_t \times n_s}$, denominada de **matriz de transferência** (de $\mathbf{x}$ para $\mathbf{y}_t$), é dada por $\mathbf{G}_{s,t} = \mathbf{G}_{\mathrm{In}(t)}$ e é tal que

$$\mathbf{y}_t = \mathbf{G}_{s,t} \cdot \mathbf{x}. \tag{2.7}$$

Analogamente ao caso não-linear, define-se também a matriz $\mathbf{H}_{s,t} \in \mathbb{F}_q^{h \times h}$, que leva $\mathbf{w}$ em $\hat{\mathbf{w}}_t$, pelo produto $\mathbf{H}_{s,t} = \hat{\mathbf{\Phi}}_t \cdot \mathbf{G}_{s,t} \cdot \mathbf{\Phi}_s$. Assim,

$$\hat{\mathbf{w}}_t = \mathbf{H}_{s,t} \cdot \mathbf{w}$$

e um código de rede linear é bem-sucedido se e somente se $\mathbf{H}_{s,t}$ é a matriz identidade $h \times h$, para todo $t \in \mathcal{T}$.

Exemplo 2.8. Do Exemplo 2.7, deduz-se que as matrizes de transferência de $\mathbf{x}$ para $\mathbf{y}_{t_1}$ e de $\mathbf{x}$ para $\mathbf{y}_{t_2}$ são, respectivamente,

$$\mathbf{G}_{s,t_1} = \mathbf{G}_{\mathrm{In}(t_1)} = \begin{bmatrix} \mathbf{G}_{e_5} \\ \mathbf{G}_{e_8} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix},$$

$$\mathbf{G}_{s,t_2} = \mathbf{G}_{\mathrm{In}(t_2)} = \begin{bmatrix} \mathbf{G}_{e_6} \\ \mathbf{G}_{e_9} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix},$$

resultado que também pode ser obtido através do Exemplo 2.4. Além disso,

$$\mathbf{H}_{s,t_1} = \hat{\mathbf{\Phi}}_{t_1} \cdot \mathbf{G}_{s,t_1} \cdot \mathbf{\Phi}_s = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

$$\mathbf{H}_{s,t_2} = \hat{\mathbf{\Phi}}_{t_2} \cdot \mathbf{G}_{s,t_2} \cdot \mathbf{\Phi}_s = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

resultado também coerente com o Exemplo 2.4. □

2.5 Suficiência da codificação de rede linear

Esta seção utiliza a abordagem algébrica de Koetter & Médard [35] para provar o Teorema 2.1. Nessa abordagem são utilizados resultados da álgebra abstrata acerca de polinômios multivariáveis (*cf.* Apêndice A.4) sobre corpos finitos.

A ideia fundamental é representar um código de rede por um conjunto de variáveis que assumem valores no corpo $\mathbb{F}_q$ mas deixar tais valores em aberto. De fato, um código de rede linear é definido inteiramente pelas hn_s variáveis

$$\alpha_{i,d}, \quad i \in \{1, \ldots, h\},\ d \in \mathrm{Out}(s),$$

que definem a matriz de codificação da fonte $\mathbf{\Phi}_s$; pelas $\sum_{v \in \mathcal{V}^*} |\mathrm{In}(v)| \cdot |\mathrm{Out}(v)|$ variáveis

$$\beta_{d,e}, \quad d \in \mathrm{In}(v),\ e \in \mathrm{Out}(v),\ v \in \mathcal{V}^*,$$

que definem as matrizes de codificação locais em $\{\mathbf{F}_v : v \in \mathcal{V}^*\}$; e pelas $\sum_{t \in \mathcal{T}} n_t h$ variáveis

$$\gamma_{e,i}, \quad e \in \mathrm{In}(t),\ i \in \{1, \ldots, h\},\ t \in \mathcal{T},$$

que definem as matrizes de decodificação em $\{\hat{\mathbf{\Phi}}_t : t \in \mathcal{T}\}$. Tais variáveis são referidas conjuntamente pelo símbolo Ξ.

Exemplo 2.9. (Adaptado de [56, Chapter 2]) No Exemplo 2.7, se deixarmos as entradas das matrizes em aberto, obtém-se

$$\mathbf{\Phi}_s = \begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{bmatrix},$$

$$\mathbf{F}_1 = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \qquad \mathbf{F}_2 = \begin{bmatrix} \beta_3 \\ \beta_4 \end{bmatrix},$$

$$\mathbf{F}_3 = \begin{bmatrix} \beta_5 & \beta_6 \end{bmatrix}, \qquad \mathbf{F}_4 = \begin{bmatrix} \beta_7 \\ \beta_8 \end{bmatrix},$$

$$\hat{\mathbf{\Phi}}_{t_1} = \begin{bmatrix} \gamma_1 & \gamma_2 \\ \gamma_3 & \gamma_4 \end{bmatrix}, \qquad \hat{\mathbf{\Phi}}_{t_2} = \begin{bmatrix} \gamma_5 & \gamma_6 \\ \gamma_7 & \gamma_8 \end{bmatrix},$$

de modo que as vinte variáveis

$$\Xi = \{\alpha_1, \ldots, \alpha_4, \beta_1, \ldots, \beta_8, \gamma_1, \ldots, \gamma_8\}$$

definem completamente o código linear de rede. Assim, existe um total de q^{20} códigos de rede lineares diferentes permitidos (nem todos bem-sucedidos). □

O teorema a seguir, além de ter como consequência imediata o Teorema 2.1, garante que códigos de rede lineares são suficientes para o problema de codificação de rede multicast. Esse resultado foi primeiramente apresentado por Li *et al.* em [39]. De quebra, o teorema também fornece um resultado sobre o tamanho suficiente para o alfabeto. A prova aqui apresentada baseia-se em Koetter & Médard [35].

Teorema 2.3. *Sejam $(\mathcal{V}, \mathcal{E})$ uma rede de comunicação e $(s, \mathcal{T})$ uma multidifusão sobre essa rede. Dadas uma taxa $h \in \mathbb{N}$ e uma potência de primo q, se*

$$h \leq \min\{\text{mincut}(s,t) : t \in \mathcal{T}\},$$

$$q > h\,|\mathcal{T}|$$

então existe código de rede linear bem-sucedido de taxa h sobre $\mathbb{F}_q$.

Demonstração. É preciso provar que, sob as duas hipóteses acima, existem matrizes $\mathbf{\Phi}_s$, $\mathbf{F}_v : v \in \mathcal{V}^*$ e $\hat{\mathbf{\Phi}}_t : t \in \mathcal{T}$ com elementos em $\mathbb{F}_q$ tais que, para todo $t \in \mathcal{T}$, tem-se $\mathbf{H}_{s,t} = \hat{\mathbf{\Phi}}_t \cdot \mathbf{G}_{s,t} \cdot \mathbf{\Phi}_s = \mathbf{I}$, em que $\mathbf{I}$ é a matriz identidade $h \times h$. Sem perda de generalidade, tal condição é equivalente a[‡]

$$\exists \xi \in \mathbb{F}_q^{|\Xi|} : \forall t \in \mathcal{T} : [\det \mathbf{H}_{s,t}]_{\Xi \leftarrow \xi} \neq 0, \tag{2.8}$$

visto que, se for esse o caso, as matrizes em $\{\hat{\mathbf{\Phi}}_t : t \in \mathcal{T}\}$ podem ser pré-multiplicadas pelas inversas das matrizes em $\{\mathbf{H}_{s,t} : t \in \mathcal{T}\}$ para se obter a matriz identidade.

Sejam $P_t(\Xi) = \det \mathbf{H}_{s,t}$, para todo $t \in \mathcal{T}$ e seja $P(\Xi) = \prod_{t \in \mathcal{T}} P_t(\Xi)$. A seguir são feitas três afirmações a respeito dessas quantidades.

Afirmação 1. *Dado qualquer $t \in \mathcal{T}$, tem-se que $P_t(\Xi)$ é um polinômio nas variáveis Ξ.*

Demonstração. De acordo com (2.5), os elementos das matrizes $\mathbf{G}_e : e \in \mathcal{E}$, são polinômios nas variáveis $\beta_{d,e}$. Consequentemente, os elementos das matrizes $\mathbf{H}_{s,t} = \hat{\mathbf{\Phi}}_t \cdot \mathbf{G}_{s,t} \cdot \mathbf{\Phi}_s : t \in \mathcal{T}$, são polinômios nas variáveis $\Xi = \{\alpha_{i,d}, \beta_{d,e}, \gamma_{e,i}\}$ e $P_t(\Xi) = \det \mathbf{H}_{s,t} : t \in \mathcal{T}$, também são polinômios nessas mesmas variáveis.

[‡] A notação $[E(\Xi)]_{\Xi \leftarrow \xi}$ significa a expressão simbólica $E(\Xi)$ avaliada no ponto ξ.

Afirmação 2. *Dado qualquer $t \in \mathcal{T}$, tem-se que $P_t(\Xi)$ não é o polinômio nulo.*

Demonstração. Fixe qualquer $t \in \mathcal{T}$. É suficiente provar que a função polinomial $P_t(\cdot)$ associada ao polinômio $P_t(\Xi)$ não é a função identicamente nula. Isto vem do fato de que existe escolha ξ_t de Ξ tal que $P_t(\xi_t) = 1$. De fato, a hipótese de que $h \leq \text{mincut}(s,t)$ juntamente com o Teorema maxflow-mincut garantem a existência de h caminhos disjuntos entre o nó fonte s e o nó destino t; basta, então, escolher as variáveis em ξ_t com os valores 1 ou 0 que reflitam um roteamento por tais caminhos. Isto faz com que $\mathbf{H}_{s,t} = \mathbf{I}$ para esse t em particular e portanto $P_t(\xi_t) = \det \mathbf{H}_{s,t} = 1$.

Observação. Note que essa escolha de ξ_t *não* assegura que $P_{t'}(\xi_t) \neq 0$ para os demais nós destino $t' \neq t$. Isso não é nenhum impedimento, pois, para provar a Afirmação 2 é necessário apenas mostrar que, dado qualquer $t \in \mathcal{T}$, tem-se que $P_t(\cdot)$ não é a função identicamente nula.

Afirmação 3. *$P(\Xi)$ é um polinômio não-nulo nas variáveis Ξ; o grau de cada variável em $P(\Xi)$ é no máximo $h\,|\mathcal{T}|$.*

Demonstração. A primeira parte dessa afirmação é consequência direta da afirmação anterior e do fato de que um produto de polinômios não-nulos é também um polinômio não-nulo. Em adição, como o grau de uma dada variável em cada um dos polinômios em $\{P_t(\Xi) : t \in \mathcal{T}\}$, é no máximo h, então o grau dessa mesma variável no polinômio $P(\Xi)$ não pode ser superior a $h\,|\mathcal{T}|$. Portanto, a segunda parte da afirmação também procede.

Tendo em vista essa última afirmação e a hipótese de que $q > h\,|\mathcal{T}|$, pode-se aplicar o Lema dos zeros esparsos (*cf.* Apêndice A.4) ao polinômio $P(\Xi)$. Com efeito, esse último garante que

$$\exists \xi \in \mathbb{F}_q^{|\Xi|} : P(\xi) \neq 0,$$

o que é equivalente a

$$\exists \xi \in \mathbb{F}_q^{|\Xi|} : \forall t \in \mathcal{T} : P_t(\xi) \neq 0,$$

o que, por sua vez, equivale a (2.8) e o teorema está provado. □

Observação. O Teorema 2.3 assegura a existência de códigos de rede lineares bem-sucedidos sobre corpos finitos $\mathbb{F}_q$ com $q > h\,|\mathcal{T}|$, mas *não*

exclui a existência de códigos de rede lineares bem-sucedidos com $q \leq h\,|\mathcal{T}|$. Por exemplo, na rede borboleta que vem sendo estudada até agora, tem-se $h = 2$ e $|\mathcal{T}| = 2$; apesar disto, o código de rede apresentado é bem-sucedido mesmo com $q = 2$.

2.6 Projeto de códigos de rede lineares

A prova do Teorema 2.3 juntamente com o Lema dos zeros esparsos sugere um método de se projetar códigos de rede lineares para o caso multidifusão: basta considerar um corpo finito de tamanho $q > h\,|\mathcal{T}|$, calcular os coeficientes dos polinômios $P_t(\Xi) = \det \mathbf{H}_{s,t}$ e, utilizando as ideias da prova do Lema dos zeros esparsos, determinar valores para as variáveis Ξ de modo que o polinômio produto $P(\Xi) = \prod_{t\in\mathcal{T}} P_t(\Xi)$ seja não-nulo. De acordo com [56, Section 2.3], uma possível implementação dessa última etapa é uma busca exaustiva.

Felizmente, foram desenvolvidos algoritmos para a obtenção de códigos de rede lineares bem-sucedidos com *complexidade polinomial* no número de canais e nós da rede. Esta seção não tem a pretensão de descrever em detalhes tais esquemas, mas apenas citar os métodos mais conhecidos na literatura. Mais detalhes podem ser obtidos em [56, Sections 2.3-2.5], [17, Chapter 5], [25, Section 2.4] e [30].

Os algoritmos para o projeto de códigos de rede podem ser classificados como *centralizados* ou *descentralizados*, de acordo com a necessidade ou não do conhecimento da topologia do sistema (*cf.* [17]). Algoritmos centralizados são executados previamente por um agente com conhecimento pleno da topologia da rede; os mapeamentos de codificação são, então, informados aos nós da rede. Em contraste, no caso decentralizado, cada nó calcula localmente seu mapeamento de codificação, sendo desnecessário o conhecimento da topologia da rede. Cada abordagem apresenta suas vantagens e desvantagens:

(i) Algoritmos centralizados, em geral, são recomendados para redes com topologia fixa[§]. Apesar de mais lentos que os decentralizados, exigem um alfabeto de tamanho reduzido e sempre fornecem códigos de rede bem-sucedidos.

[§]Embora a propriedade de *estaticidade* de um código de rede (introduzida em [35]) permita o uso de algoritmos centralizados mesmo em configurações de rede sujeitas a variações.

(*ii*) Algoritmos decentralizados, por outro lado, são mais adequados a redes com topologia variável (*e.g.*, com nós saindo e entrando). São, normalmente, mais rápidos que os centralizados. Tais métodos, em muitos casos, não garantem a validade do código de rede obtido; em geral, há um compromisso entre o tamanho do alfabeto e a probabilidade de o código ser bem-sucedido.

2.6.1 Algoritmo LIF

O *algoritmo LIF* (do inglês *linear information flow*) foi desenvolvido independentemente por Jaggi *et al.* em [29] e Sanders *et al.* em [44], trabalhos mais tarde combinados em [30]. Foi o primeiro algoritmo centralizado em tempo polinomial para o projeto de códigos de rede lineares.

O procedimento consiste em percorrer os nós da rede sequencialmente, obedecendo a uma ordem de codificação. Cada nó visitado recebe um mapeamento de codificação tal que a "propriedade de multidifusão" permanece satisfeita no subconjunto composto pelo nó atual e pelos outros nós já visitados. É portanto, um algoritmo *greedy*.

Possui duas variantes: *DLIF* (do inglês *deterministic* LIF), com tempo de execução $O(|\mathcal{E}|\,|\mathcal{T}|\,h(|\mathcal{T}|+h))$, sendo $q > |\mathcal{T}|$ suficiente; e *RLIF* (do inglês *randomized* LIF), com tempo de execução $O(|\mathcal{E}|\,|\mathcal{T}|\,h^2)$, sendo $q > 2\,|\mathcal{T}|$ suficiente. Mais detalhes são apresentados em [30].

2.6.2 Completamento de matrizes

Em [22, 23], Harvey *et al.* apresentam um algoritmo centralizado e com tempo de execução polinomial para o projeto de códigos de rede lineares. O algoritmo faz uso do conceito de *matrizes mistas*, matrizes cujas entradas podem ser tanto *constantes numéricas* quanto *variáveis simbólicas*.

De fato, utilizando a abordagem algébrica discutida na Seção 2.5, as matrizes de transferência relativas a cada nó podem ser expressas por matrizes mistas, contendo constantes numéricas (dependentes da topologia da rede) e variáveis simbólicas (cujas escolhas determinam o código de rede). Assim, o problema de codificação de rede linear se traduz em uma escolha adequada de valores para as variáveis simbólicas—procedimento denominado de *completamento de matrizes*—que resulta em matrizes de transferência todas com posto completo.

O algoritmo é executado em um tempo $O(|\mathcal{T}|\,|\mathcal{E}|^3 \log |\mathcal{E}|)$ e utiliza um corpo finito de tamanho $q > |\mathcal{T}|$.

2.6.3 Códigos de rede aleatórios

Codificação de rede linear aleatória foi proposta primeiramente por Ho *et al.* em [24] (veja também [26]). Nela, cada nó intermediário efetua *combinações lineares aleatórias* dos símbolos que recebe. No contexto deste capítulo, isso significa que os elementos das matrizes de codificação locais são escolhidos aleatoriamente. Uma vez que não exige conhecimento da topologia da rede por parte de cada nó, esse algoritmo pode ser implementado de maneira descentralizada.

O seguinte teorema (*cf.* [17]) justifica a ideia de codificação de rede linear aleatória ao mostrar que a probabilidade de um código ser bem-sucedido tende a 1 quando o tamanho q do corpo tende a infinito.

Teorema 2.4. *Sejam $(\mathcal{V}, \mathcal{E})$ uma rede de comunicação e $(s, \mathcal{T})$ uma multidifusão sobre essa rede. Dada uma taxa $h \in \mathbb{N}$ e uma potência de primo q. Assuma*

$$h \leq \min\{\text{mincut}(s,t) : t \in \mathcal{T}\},$$

$$q > h\,|\mathcal{T}|$$

e considere um código de rede linear de taxa h sobre $\mathbb{F}_q$ com os elementos das matrizes de codificação locais escolhidos aleatoriamente de maneira uniforme e independente. Então, o código de rede assim construído é bem-sucedido com probabilidade $(1 - h\,|\mathcal{T}|\,/q)^{|\Xi|}$ ou maior.

Demonstração. Segue diretamente da aplicação do Lema da probabilidade dos zeros (*cf.* Apêndice A.4) à abordagem da Seção 2.5. □

2.6.4 Códigos de rede descentralizados não-aleatórios

Existem ainda algoritmos descentralizados não-aleatórios para o projeto de códigos de rede. Apesar de não exigirem conhecimento sobre a topologia da rede, tais algoritmos operam apenas sobre classes de problemas bem específicos (por exemplo, multidifusão com dois nós destino). Mais detalhes e referências podem ser obtidos em [17, Chapter 5].

2.7 Codificação de rede convolucional

2.7.1 Problema de atrasos e ciclos

Como discutido no início deste capítulo, uma rede acíclica permite que se adote um modelo de transmissões que desconsidera atrasos, evitando, ao mesmo tempo, problemas de inconsistência e não-causalidade. Essa abordagem, apesar de matematicamente mais cômoda, pode não ser adequada em muitos casos práticos:

(*i*) Atrasos estão presentes em todas as redes de comunicação reais; por exemplo, atrasos de processamento e de transmissão. Enquanto em algumas aplicações pode-se desconsiderar tais atrasos, em muitas outras é necessário um modelo que os leve em conta.

(*ii*) Na prática, a maioria das redes possuem ciclos—canais bidirecionais, por exemplo, implicam em uma rede cíclica. Nesse caso, mesmo que os atrasos sejam negligenciáveis, em alguns casos é mandatória a introdução de atrasos em cada ciclo para evitar problemas de inconsistência e não-causalidade[¶], como mostra o exemplo a seguir.

Exemplo 2.10. (Adaptado de [17, Chapter 6]) Considere a rede cíclica mostrada na Figura 2.5, na qual o conjunto dos nós é dado por

$$\mathcal{V} = \{s, t_1, t_2\}$$

e o conjunto dos canais é dado por

$$\mathcal{E} = \{e_1, e_2, e_3, e_4\},$$

em que $e_1 = (s, t_1)$, $e_2 = (s, t_2)$, $e_3 = (t_1, t_2)$ e $e_4 = (t_2, t_1)$. A fonte gera um símbolo $\mathbf{w} = (w_1, w_2)$. Seja o mapeamento de codificação da fonte dado pelo mapeamento identidade e os mapeamentos de codifica-

[¶]Em [17], as autoras mostram que a aciclicidade do grafo é uma condição suficiente mas não necessária para que se possa desprezar os atrasos e apresentam um critério necessário e suficiente para que o problema dos ciclos seja levado em conta.

ção locais dos nós t_1 e t_2 dados pela operação de soma, isto é,

$$\begin{aligned} a_{e_1} &= w_1, \\ a_{e_2} &= w_2, \\ a_{e_3} &= a_{e_1} + a_{e_4}, \\ a_{e_4} &= a_{e_2} + a_{e_3}. \end{aligned} \tag{2.9}$$

Conjuntamente, essas quatro equações implicam em $w_1 + w_2 = 0$, o que é inconsistente, visto que w_1 e w_2 devem ser independentes.

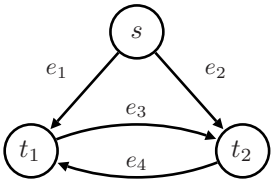


Figura 2.5: Exemplo de uma rede cíclica.

Curiosamente, o código de rede refletindo um roteamento (símbolo w_1 roteado por e_1 e e_3; símbolo w_2 roteado por e_2 e e_4) não apresenta esse tipo de problema. □

2.7.2 Códigos de rede convolucionais

Uma generalização da codificação de rede linear "instantânea", que vem sendo considerada até então, é a chamada *codificação de rede convolucional*, que introduz uma dimensão temporal (discreta) no modelo. Codificação de rede convolucional foi proposta independentemente por Erez & Feder em [9] e Fragouli & Soljanin em [15].

Nesse novo cenário, as quantidades $\mathbf{w}$, $a_e : e \in \mathcal{E}$ e $\hat{\mathbf{w}}_t : t \in \mathcal{T}$ definidas anteriormente são dependentes da variável temporal k; adotam-se as notações $\mathbf{w}[k] \in \mathbb{F}_q^h$, $a_e[k] \in \mathbb{F}_q$ e $\hat{\mathbf{w}}_t[k] \in \mathbb{F}_q^h$, para os valores daquelas grandezas no instante k. Em adição, são assumidas conhecidas as condições iniciais ($k \leq 0$) de cada uma dessas variáveis.

O mapeamento de codificação local de um dado nó não-fonte $v \in \mathcal{V}^*$ é completamente especificado por uma *equação de diferença* linear, aqui considerada invariante no tempo. A saída do nó no instante presente pode depender das entradas no instante presente e nos instantes passados, bem como da própria saída nos instantes passados.

Codificação de rede convolucional, por si só, *não* resolve o problema de inconsistência e não-causalidade. Isso é resolvido inserindo atrasos em ciclos "problemáticos". Uma maneira segura, simplificada e comumente adotada é associar *atrasos unitários* ao mapeamento local de *todos* os nós não-fonte da rede. (Relevam-se os atrasos nos mapeamentos de codificação da fonte e de decodificação.)

Assim sendo, distinguem-se três casos.

Caso sem memória. O mapeamento de codificação local utiliza apenas a entrada do nó no instante presente e, portanto, continua sendo definido por uma matriz $\mathbf{F}_v \in \mathbb{F}_q^{|\mathrm{Out}(v)|\times|\mathrm{In}(v)|}$. Tem-se

$$\mathbf{a}_{\mathrm{Out}(v)}[k+1] = \mathbf{F}_v \cdot \mathbf{a}_{\mathrm{In}(v)}[k],$$

para $k \geq 0$.

Caso FIR. O mapeamento de codificação local utiliza a entrada do nó no instante atual e nos $\mu-1$ instantes anteriores, sendo, portanto, definido por μ matrizes $\mathbf{F}_v[i] \in \mathbb{F}_q^{|\mathrm{Out}(v)|\times|\mathrm{In}(v)|} : i = 0, \ldots, \mu-1$. Tem-se

$$\mathbf{a}_{\mathrm{Out}(v)}[k+1] = \sum_{i=0}^{\mu-1} \mathbf{F}_v[i] \cdot \mathbf{a}_{\mathrm{In}(v)}[k-i],$$

para $k \geq 0$. Se recai no caso anterior quando $\mu = 1$.

Caso IIR. O mapeamento de codificação local pode utilizar valores de suas entradas atuais e anteriores, bem como valores das saídas anteriores. Tem-se

$$\mathbf{a}_{\mathrm{Out}(v)}[k+1] = \sum_{i=0}^{\mu-1} \mathbf{F}_v[i] \cdot \mathbf{a}_{\mathrm{In}(v)}[k-i] - \sum_{i=1}^{\mu'-1} \mathbf{F}'_v[i] \cdot \mathbf{a}_{\mathrm{Out}(v)}[k-i],$$

para $k \geq 0$. Além das μ matrizes do caso FIR, são necessárias $\mu'-1$ matrizes $\mathbf{F}'_v[i] \in \mathbb{F}_q^{|\mathrm{Out}(v)|\times|\mathrm{Out}(v)|} : i = 1, \ldots, \mu'-1$ para descrevê-lo. Esse é o caso mais geral e recai no caso anterior quando $\mu' = 1$.

Analogamente, para o mapeamento de codificação da fonte, tem-se

$$\mathbf{x}[k] = \sum_{i=0}^{\mu-1} \mathbf{\Phi}_s[i] \cdot \mathbf{w}[k-i] - \sum_{i=1}^{\mu'-1} \mathbf{\Phi}'_s[i] \cdot \mathbf{x}[k-i],$$

em que $\mathbf{x}[k] = \mathbf{a}_{\mathrm{Out}(s)}[k]$ e, para o mapeamento de decodificação dos nós destino, tem-se

$$\hat{\mathbf{w}}_t[k] = \sum_{i=0}^{\mu-1} \hat{\mathbf{\Phi}}_t[i] \cdot \mathbf{y}_t[k-i] - \sum_{i=1}^{\mu'-1} \hat{\mathbf{\Phi}}'_t[i] \cdot \hat{\mathbf{w}}_t[k-i],$$

em que $\mathbf{y}_t[k] = \mathbf{a}_{\mathrm{In}(t)}[k]$. Evidentemente, pode-se particularizar essas equações para o caso FIR ou sem memória.

Observação. Note os atrasos unitários de processamento inseridos nas três primeiras equações (mapeamentos de codificação locais dos nós não-fonte). Como já mencionado, esses atrasos não são considerados nas duas equações anteriores (mapeamentos de codificação da fonte e de decodificação dos nós destino).

Se, para todo nó destino, a sequência estimada por esse nó é igual à sequência mensagem emitida pela fonte, a menos de um defasamento temporal, o código de rede assim definido é dito **bem-sucedido**. Simbolicamente, o código de rede é válido se e somente se

$$\forall t \in \mathcal{T} : \exists \Delta_t \in \mathbb{N} : \forall k \geq 0 : \hat{\mathbf{w}}_t[k + \Delta_t] = \mathbf{w}[k].$$

Exemplo 2.11. Considere a rede cíclica do Exemplo 2.10. Considere também o código de rede especificado por (2.9); adaptado ao contexto de codificação de rede convolucional (caso sem memória), tem-se

$$\begin{aligned} a_{e_1}[k] &= w_1[k], \\ a_{e_2}[k] &= w_2[k], \\ a_{e_3}[k+1] &= a_{e_1}[k] + a_{e_4}[k], \\ a_{e_4}[k+1] &= a_{e_2}[k] + a_{e_3}[k]. \end{aligned} \tag{2.10}$$

A evolução temporal do sistema pode ser representada por uma treliça, como na Figura 2.6. Nesse exemplo, a saída de cada nó não-fonte é a soma de suas entradas no instante anterior, enquanto a saída do nó fonte são os símbolos de informação gerados no mesmo instante.

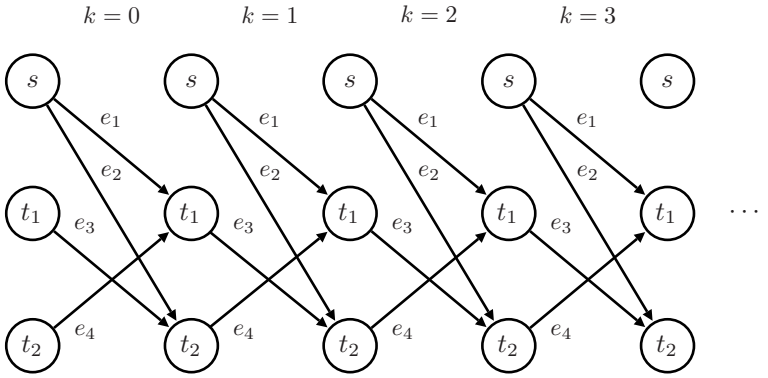


Figura 2.6: Treliça derivada da rede da Figura 2.5.

A seguinte tabela mostra a evolução dos símbolos de cada canal da rede nos primeiros instantes, supondo condições iniciais nulas.

k	$a_{e_1}[k]$	$a_{e_2}[k]$	$a_{e_3}[k]$	$a_{e_4}[k]$
0	$w_1[0]$	$w_2[0]$	0	0
1	$w_1[1]$	$w_2[1]$	$w_1[0]$	$w_2[0]$
2	$w_1[2]$	$w_2[2]$	$w_2[0]+w_1[1]$	$w_1[0]+w_2[1]$
3	$w_1[3]$	$w_2[3]$	$w_1[0]+w_2[1]+w_1[2]$	$w_2[0]+w_1[1]+w_2[2]$

A partir das equações (2.10) é possível obter relações de recorrência que expressam as sequências das mensagens em função das sequências de entrada do nó t_1:

$$\begin{aligned} w_1[k] &= a_{e_1}[k], \\ w_2[k] &= -a_{e_1}[k-1]+a_{e_4}[k+1]-a_{e_4}[k-1]. \end{aligned}$$

Assim, uma decodificação com atraso $\Delta_{t_1}=1$ é possível para o nó t_1. Resultado análogo pode ser alcançado para o nó t_2. □

2.7.3 Representação pela transformada z

Uma representação conveniente de uma sequência qualquer $\{a[k]\}_{k=0}^{\infty}$ é dada pela sua *transformada* z, definida por

$$a(z) = \sum_{k=0}^{\infty} a[k] z^{-k}.$$

Sob essa ótica, a fonte tem acesso a uma h-tupla $\mathbf{w}(z)$ de séries. Similarmente, cada canal $e \in \mathcal{E}$ transporta uma série $a_e(z)$ e cada nó destino $t \in \mathcal{T}$ efetua um palpite $\hat{\mathbf{w}}_t(z)$. Esse enfoque é equivalente à proposta inicial de Koetter & Médard em [35] para se lidar com o problema de ciclos e atrasos.

Assim, o mapeamento de codificação local do nó não-fonte $v \in \mathcal{V}^*$ é representado sucintamente por uma única matriz $\mathbf{F}_v(z)$. As entradas dessa matriz são

(*i*) elementos de $\mathbb{F}_q$, no caso de mapeamentos sem-memória;

(*ii*) elementos de $\mathbb{F}_q\left[z^{-1}\right]$, isto é, polinômios em z^{-1}, no caso de mapeamentos FIR; ou

(*iii*) elementos de $\mathbb{F}_q\left(z^{-1}\right)$, isto é, razões entre polinômios em z^{-1}, no caso de mapeamentos IIR.

Visto que $\mathbb{F}_q \subset \mathbb{F}_q\left[z^{-1}\right] \subset \mathbb{F}_q\left(z^{-1}\right)$, há uma ordem crescente de generalidade. Seja qual for o caso, a equação que relaciona as entradas e saídas do nó não-fonte $v \in \mathcal{V}^*$ é

$$\mathbf{a}_{\mathrm{Out}(v)}(z) = z^{-1}\mathbf{F}_v(z) \cdot \mathbf{a}_{\mathrm{In}(v)}(z), \tag{2.11}$$

em que o fator z^{-1} no lado direito da equação representa o atraso unitário de processamento em cada nó. Similarmente, define-se as equações de codificação da fonte e de decodificação dos nós destino (não incluindo o fator z^{-1}).

O código de rede é bem-sucedido se e somente se

$$\forall t \in \mathcal{T} : \exists \Delta_t \in \mathbb{N} : \hat{\mathbf{w}}_t(z) = z^{-\Delta_t}\mathbf{w}(z). \tag{2.12}$$

Exemplo 2.12. Aplicando a equação (2.11) ao código de rede do Exemplo 2.11, obtém-se

$$\begin{aligned}\mathbf{a}_{\mathrm{Out}(s)}(z) &= \mathbf{F}_s(z)\cdot\mathbf{w}(z),\\ \mathbf{a}_{\mathrm{Out}(t_1)}(z) &= z^{-1}\mathbf{F}_{t_1}(z)\cdot\mathbf{a}_{\mathrm{In}(t_1)},\\ \mathbf{a}_{\mathrm{Out}(t_2)}(z) &= z^{-1}\mathbf{F}_{t_1}(z)\cdot\mathbf{a}_{\mathrm{In}(t_1)},\end{aligned}$$

em que

$$\mathbf{\Phi}_s(z)=\begin{bmatrix}1&0\\0&1\end{bmatrix},\qquad \mathbf{F}_{t_1}(z)=\mathbf{F}_{t_2}(z)=\begin{bmatrix}1&1\end{bmatrix}$$

não dependem de z pois são mapeamentos sem memória. Essas equações se traduzem em

$$\begin{aligned}a_{e_1}(z) &= w_1(z),\\ a_{e_2}(z) &= w_2(z),\\ a_{e_3}(z) &= z^{-1}a_{e_1}(z)+z^{-1}a_{e_4}(z),\\ a_{e_4}(z) &= z^{-1}a_{e_2}(z)+z^{-1}a_{e_3}(z),\end{aligned}$$

resultado que também pode ser obtido aplicando a transformada z às equações (2.10). Assim, caso se queira obter $w_1(z)$ e $w_2(z)$ em função das entradas $a_{e_1}(z)$ e $a_{e_4}(z)$ do nó t_1, por exemplo, basta resolver as equações acima, resultando em

$$\begin{aligned}w_1(z) &= a_{e_1}(z),\\ w_2(z) &= -z^{-1}a_{e_1}(z)+\left(z-z^{-1}\right)a_{e_4}(z).\end{aligned}$$

Portanto, um mapeamento de decodificação para t_1 que torna o código de rede bem-sucedido é

$$\hat{\mathbf{\Phi}}_{t_1}(z)=\begin{bmatrix}z^{-1}&0\\-z^{-2}&1-z^{-2}\end{bmatrix},$$

fornecendo $\Delta_{t_1}=1$ na condição (2.12). □

Os exemplos anteriores, por simplicidade, só levaram em conta o caso de mapeamentos de codificação locais sem memória. No entanto, o uso de mapeamentos com memória pode apresentar vantagens, tais como a possibilidade de um menor corpo [17, Subsection 6.1.1] e a simplificação

do processo de decodificação nos nós destino [56, Section 3.2]. Adicionalmente, ressalta-se que a abordagem da transformada z permite a generalização de muitos resultados da codificação instantânea para a codificação convolucional, visto que ambos os casos são análogos, a menos de uma mudança de domínio—de $\mathbb{F}_q$ para $\mathbb{F}_q(z^{-1})$.

A linha de codificação convolucional não será seguida no restante desta dissertação. Mais detalhes podem ser obtidos nos trabalhos de Erez & Feder [9, 10], Fragouli & Soljanin [15] e Li & Yeung [38].

CAPÍTULO 3

Codificação de rede externa

O Capítulo 2 serviu como base para apresentar os conceitos e resultados fundamentais da teoria da codificação de rede multidifusão. De certa forma, o capítulo focou na codificação de rede *interna*. O presente capítulo, por outro lado, aborda a codificação de rede *externa*, isto é, a comunicação fim-a-fim entre o nó fonte e um determinado nó destino. Uma vez definido o código de rede interno, o efeito desse sobre a comunicação fim-a-fim pode ser representado sucintamente pela matriz de transferência, que leva os dados transmitidos pelo nó fonte nos dados recebidos pelo nó destino em consideração.

Inicia-se apresentando uma generalização conveniente da codificação de rede linear estudada anteriormente: *codificação de rede vetorial linear*. Após isso é obtido o *canal matricial* que representa adequadamente a comunicação fim-a-fim e são definidos os chamados *códigos matriciais externos*. Esse modelo permite a solução de dois problemas: o desconhecimento do código de rede interno (codificação de rede não-coerente) e a presença de erros nos enlaces da rede (controle de erros em codificação de rede); tais tópicos são abordados em seguida. O capítulo termina com o *modelo de gerações*, um esquema prático para contornar o problema do assincronismo presente em diversas redes de comunicação reais; o canal matricial se aplica notavelmente a esse modelo.

3.1 Codificação de rede vetorial linear

No cenário de codificação de rede linear considerado até o capítulo anterior, o sistema operava sobre *escalares* pertencentes a um determinado corpo finito $\mathbb{F}_q$. Neste capítulo, tal modelo é generalizado para os chamados *códigos de rede vetoriais lineares*, nos quais as operações são executadas sobre *vetores* de um espaço vetorial da forma $\mathbb{F}_q^m$. Nesse caso, um vetor é uma m-tupla com elementos em um corpo finito $\mathbb{F}_q$.

3.1.1 Funcionamento do sistema

O funcionamento do sistema é análogo ao caso escalar linear e é descrito a seguir. Neste capítulo, o nó destino considerado é sempre o mesmo, de modo que alguns subscritos são descartados por conveniência.

(*i*) O nó fonte s tem acesso a um símbolo de informação $w \in \mathcal{W}$ e os codifica, através do *mapeamento de codificação da fonte* ϕ, em $n_s = |\mathrm{Out}(s)|$ vetores de $\mathbb{F}_q^m$, denotados por $\mathbf{x}_1, \ldots, \mathbf{x}_{n_s}$:

$$\begin{aligned} \phi : \mathcal{W} &\longrightarrow (\mathbb{F}_q^m)^{n_s} \\ w &\longmapsto (\mathbf{x}_1, \ldots, \mathbf{x}_{n_s}). \end{aligned}$$

(*ii*) Cada nó não-fonte $v \in \mathcal{V}^*$ recebe $|\mathrm{In}(v)|$ vetores em seus canais de entrada e calcula, através da *matriz de codificação local* $\mathbf{F}_v \in \mathbb{F}_q^{|\mathrm{Out}(v)| \times |\mathrm{In}(v)|}$, os $|\mathrm{Out}(v)|$ vetores de seus canais de saída:

$$\begin{aligned} \mathbf{F}_v \cdot : (\mathbb{F}_q^m)^{|\mathrm{In}(v)|} &\longrightarrow (\mathbb{F}_q^m)^{|\mathrm{Out}(v)|} \\ (\mathbf{a}_e : e \in \mathrm{In}(v)) &\longmapsto (\mathbf{a}_e : e \in \mathrm{Out}(v)). \end{aligned}$$

(*iii*) O nó destino $t \in \mathcal{T}$ em consideração recebe em sua entrada $n_t = |\mathrm{In}(t)|$ vetores de $\mathbb{F}_q^m$, denotados por $\mathbf{y}_1, \ldots, \mathbf{y}_{n_t}$, e, através do *mapeamento de decodificação* $\hat{\phi}$, faz uma estimativa $\hat{w} \in \mathcal{W}$ do símbolo de informação w:

$$\begin{aligned} \hat{\phi} : (\mathbb{F}_q^m)^{n_t} &\longrightarrow \mathcal{W} \\ (\mathbf{y}_1, \ldots, \mathbf{y}_{n_t}) &\longmapsto \hat{w}. \end{aligned}$$

Assim como no caso linear escalar, o efeito da codificação de rede interna pode ser visto como um sistema MIMO (do inglês *multiple input, multiple output*) sobre um corpo finito, como mostrado na Figura 3.1.

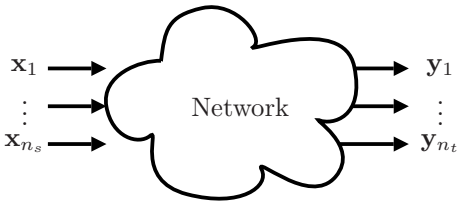


Figura 3.1: Rede de comunicação vista como um sistema MIMO.

3.1.2 Representação matricial

Alternativa e convenientemente, pode-se dispor os n_s vetores (m-tuplas) transmitidos $\mathbf{x}_1, \ldots, \mathbf{x}_{n_s}$ nas linhas de uma matriz $\mathbf{X} \in \mathbb{F}_q^{n_s \times m}$ e os n_t vetores (m-tuplas) recebidos $\mathbf{y}_1, \ldots, \mathbf{y}_{n_t}$ nas linhas de uma matriz $\mathbf{Y} \in \mathbb{F}_q^{n_t \times m}$. Feito isso, a relação (2.7) é adaptada ao caso vetorial linear para

$$\mathbf{Y} = \mathbf{G} \cdot \mathbf{X}, \tag{3.1}$$

em que $\mathbf{G} \in \mathbb{F}_q^{n_t \times n_s}$ é a *matriz de transferência*, que leva a matriz de entrada $\mathbf{X}$ na matriz de saída $\mathbf{Y}$; a cadeia do sistema (2.6) torna-se

$$w \overset{\phi}{\longmapsto} \mathbf{X} \overset{\mathbf{G}\cdot}{\longmapsto} \mathbf{Y} \overset{\hat{\phi}}{\longmapsto} \hat{w},$$

em que $\phi : \mathcal{W} \longrightarrow \mathbb{F}_q^{n_s \times m}$ e $\hat{\phi} : \mathbb{F}_q^{n_t \times m} \longrightarrow \mathcal{W}$ são, respectivamente, o mapeamento de codificação da fonte e o mapeamento de decodificação do nó destino em consideração.

3.2 Códigos de rede externos

Por simplicidade, é comum assumir $n_t = n_s = h$ e assim será feito[*] neste trabalho. Quando é esse o caso, as matrizes $\mathbf{X}$ e $\mathbf{Y}$ têm ambas dimensão $h \times m$, enquanto a matriz $\mathbf{G}$ torna-se quadrada com dimensão $h \times h$.

Se um código de rede é bem-sucedido, então o mapeamento de codificação da fonte, ϕ, é necessariamente injetivo (caso contrário, existiriam $w_1, w_2 \in \mathcal{W}$ distintos tais que $\mathbf{X} = \phi(w_1) = \phi(w_2)$ e, portanto, w_1 e w_2 acarretariam no mesmo $\mathbf{Y}$ e no mesmo $\hat{w}$).

[*] *cf.* [41, 50].

Essa observação sugere que o mapeamento $\phi : \mathcal{W} \longrightarrow \mathbb{F}_q^{h\times m}$ pode ser, alternativamente, representado por sua imagem, isto é, pelo subconjunto $\mathcal{X} = f(\mathcal{W})$ de $\mathbb{F}_q^{h\times m}$. A injetividade implica $|\mathcal{W}| = |\mathcal{X}|$. O mapeamento de decodificação $\hat{\phi}$ ainda é necessário, mas agora efetua um palpite $\hat{\mathbf{X}}$ sobre a matriz $\mathbf{X}$ a partir da matriz $\mathbf{Y}$. Com isso em mente, adota-se a seguinte definição.

DEFINIÇÃO. Um **código de rede externo** é definido por um par $(\mathcal{X}, \hat{\phi})$, em que $\mathcal{X} \subseteq \mathbb{F}_q^{h\times m}$ é chamado de **código matricial** e $\hat{\phi} : \mathbb{F}_q^{h\times m} \longrightarrow \mathcal{X}$ é denominado de **decodificador**. A **taxa** desse código externo é definida por

$$R(\mathcal{X}) = \frac{\log_q |\mathcal{X}|}{hm}$$

e corresponde à razão entre o número de símbolos q-ários representando informação e o número total de símbolos q-ários transmitidos. □

Uma vez que a cadeia do sistema torna-se

$$\mathbf{X} \overset{\mathbf{G}\cdot}{\longmapsto} \mathbf{Y} \overset{\hat{\phi}}{\longmapsto} \hat{\mathbf{X}},$$

é adequada a redefinição do termo *bem-sucedido* sob a ótica da codificação de rede externa.

DEFINIÇÃO. Dada matriz de transferência $\mathbf{G}$ representando o código de rede interno, um código de rede externo $(\mathcal{X}, \hat{\phi})$ é dito ser **bem-sucedido** (sobre $\mathbf{G}$) se, para todo $\mathbf{X} \in \mathcal{X}$, tem-se $\hat{\mathbf{X}} = \mathbf{X}$. □

Portanto, ao se utilizar um código externo com taxa $R(\mathcal{X})$ é possível transmitir informação a uma taxa de $h \cdot R(\mathcal{X})$ símbolos q-ários por uso da rede, desde que esse seja bem-sucedido.

O Teorema 2.3 adaptado ao caso em questão afirma que se q for suficientemente grande e se h for menor que o valor do mínimo corte entre a fonte e o nó destino em consideração então é possível escolher matriz $\mathbf{G}$ inversível. Isso, por sua vez, permite que se escolha um código externo $(\mathcal{X}, \hat{\phi})$ bem-sucedido com

$$R(\mathcal{X}) = 1,$$

em que o código matricial $\mathcal{X}$ é o conjunto de todas as matrizes $h \times m$ com elementos em $\mathbb{F}_q$ e o decodificador $\hat{\phi}$ calcula a matriz $\mathbf{X}$ a

partir da matriz $\mathbf{Y}$ através de (3.1), isto é, $\mathcal{X} = \mathbb{F}_q^{h\times m}$ e $\hat{\phi} : \mathbf{Y} \longmapsto \mathbf{G}^{-1} \cdot \mathbf{Y}$. Assim, o teorema garante que é sempre possível obter código de rede interno $\mathbf{G}$ e código de rede externo $(\mathcal{X}, \hat{\phi})$ bem-sucedido que desenvolvem uma taxa de símbolos h.

3.3 Codificação de rede não-coerente

Esta seção aborda a chamada *codificação de rede não-coerente*, na qual considera-se que o código de rede interno utilizado (seja esse aleatório ou não) é desconhecido pelos nós fonte e destino. Codificação de rede não-coerente aliada a codificação de rede linear aleatória é bastante atrativa do ponto de vista prático, pois permite um funcionamento decentralizado da rede de comunicação. A topologia da rede pode ser desconhecida e até mesmo variante no tempo.

É utilizado o modelo da seção anterior. Nesse contexto, a não-coerência se traduz no desconhecimento da matriz de transferência $\mathbf{G} \in \mathbb{F}_q^{h\times h}$. Outra premissa adotada é a inversibilidade dessa matriz: isso é sempre possível de ser obtido (de acordo com Teorema 2.3) ou ocorre com alta probabilidade em codificação de rede linear aleatória (de acordo com o Teorema 2.4) quando a condição de mincut é satisfeita e o tamanho do corpo é suficientemente grande.

No caso *coerente*, o conhecimento da matriz $\mathbf{G}$ é garantido. Sendo assim, de acordo com a seção anterior, é sempre possível obter código externo $(\mathcal{X}, \hat{\phi})$ bem-sucedido sobre $\mathbf{G}$ com taxa $R(\mathcal{X}) = 1$. Já no caso *não-coerente*, nem o nó fonte nem o nó destino têm conhecimento *a priori* da matriz $\mathbf{G}$. Mesmo nessa circunstância a comunicação ainda é possível. A seguir são apresentados dois métodos, cada um fornecendo um código externo $(\mathcal{X}, \hat{\phi})$ que é bem-sucedido *simultaneamente* sobre todas as matrizes $\mathbf{G}$ inversíveis.

3.3.1 Matriz identidade no cabeçalho

O primeiro método para lidar com o problema da não-coerência foi proposto por Chou, Wu & Jain em [5] e consiste na inclusão de um *cabeçalho* em cada vetor emitido pela fonte; o cabeçalho do i-ésimo vetor injetado é a h-tupla contendo 1 na posição i e 0 nas demais posições. É necessário assumir $m > h$. Assim, a matriz transmitida

pela fonte é da forma

$$\begin{aligned}\mathbf{X} &= \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & x'_{1,1} & x'_{1,2} & \cdots & x'_{1,(m-h)} \\ 0 & 1 & 0 & \cdots & 0 & x'_{2,1} & x'_{2,2} & \cdots & x'_{2,(m-h)} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & x'_{h,1} & x'_{h,2} & \cdots & x'_{h(m-h)} \end{bmatrix} \\ &= \left[\begin{array}{c|c} \mathbf{I} & \mathbf{X}' \end{array}\right], \end{aligned} \tag{3.2}$$

em que $\mathbf{I}$ é a matriz identidade $h \times h$ e $x'_{i,j}$ são símbolos de informação que compõem a matriz $\mathbf{X}' \in \mathbb{F}_q^{h\times(m-h)}$.

O nó destino, através da relação linear (3.1), recebe a matriz

$$\begin{aligned}\mathbf{Y} &= \mathbf{G}\cdot\mathbf{X} \\ &= \mathbf{G}\cdot\left[\begin{array}{c|c} \mathbf{I} & \mathbf{X}' \end{array}\right] \\ &= \left[\begin{array}{c|c} \mathbf{G} & \mathbf{G}\cdot\mathbf{X}' \end{array}\right], \end{aligned}$$

tendo, portanto, acesso à matriz $\mathbf{G}$ de transferência e à matriz $\mathbf{Y}' = \mathbf{G}\cdot\mathbf{X}'$ contendo os símbolos de informação transformados. Assim, esse nó é capaz de recuperar os símbolos de informação originais $\mathbf{X}'$.[†]

O preço que se paga ao utilizar esse esquema é um *overhead*: dos hm símbolos transmitidos, h^2 deles são desperdiçados nos cabeçalhos. Se o código matricial $\mathcal{X} \subseteq \mathbb{F}_q^{h\times m}$ for o conjunto de todas as matrizes da forma (3.2), é possível transmitir um total de

$$|\mathcal{X}| = q^{h(m-h)}$$

matrizes diferentes e, portanto, a taxa do código é

$$R(\mathcal{X}) = 1 - \frac{h}{m}.$$

3.3.2 Transmissão de subespaços

Este segundo esquema foi sugerido por Koetter & Kschischang em [33] e se baseia na seguinte observação: supondo que a matriz $\mathbf{G} \in \mathbb{F}_q^{h\times h}$ seja inversível, o subespaço vetorial de $\mathbb{F}_q^m$ gerado pelas linhas da matriz $\mathbf{X} \in \mathbb{F}_q^{h\times m}$ é *preservado* após a pré-multiplicação por $\mathbf{G}$ em (3.1).

[†]Fazendo um paralelo com outros sistemas de comunicação, a matriz identidade $\mathbf{I}$ inserida no cabeçalho faz o papel das "portadoras piloto" utilizadas para estimar o canal de comunicação, aqui representado pela matriz $\mathbf{G}$.

Simbolicamente,

$$\langle \mathbf{X} \rangle = \langle \mathbf{Y} \rangle ,$$

qualquer que seja a matriz $\mathbf{G}$ inversível. Portanto, a ideia é codificar a informação não mais diretamente nos elementos da matriz $\mathbf{X}$, mas sim em $\langle \mathbf{X} \rangle$, isto é, no subespaço gerado por ela. Na prática, transmite-se um conjunto contendo vetores que geram o subespaço.

De acordo com a equação (A.3), do Apêndice A.5, se $\mathcal{C}$ for o conjunto de todos os possíveis subespaços vetoriais de $\mathbb{F}_q^m$ com dimensão h ou menos e $\mathcal{X} \subseteq \mathbb{F}_q^{h \times m}$ for um código matricial tal que $\{\langle \mathbf{X} \rangle : \mathbf{X} \in \mathcal{X}\} = \mathcal{C}$ e $|\mathcal{X}| = |\mathcal{C}|$, tem-se

$$|\mathcal{X}| = \sum_{k=0}^{h} \binom{m}{k}_q ,$$

em que $\binom{m}{k}_q$ é definido em (A.2).

O caso anterior pode ser visto como um caso particular desse: cada matriz da forma (3.2) dá origem a um subespaço h-dimensional diferente. No entanto, existem subespaços h-dimensionais que não são gerados por matrizes da forma (3.2). Ainda mais, matrizes da forma (3.2) não são capazes de gerar nenhum subespaço de dimensão estritamente menor que h.

Essa proposta é, de fato, o melhor que se pode fazer no caso em que a matriz $\mathbf{G}$ é selecionada aleatoria e uniformemente entre todas as matrizes $h \times h$ inversíveis, como provam Silva, Koetter & Kschischang em [50]. Esse trabalho também mostra que os dois métodos apresentados coincidem assintoticamente em dois contextos (ambos sujeitos a $m \geq 2h$): quando o tamanho q do corpo tende a infinito e quando os valores de m e h tendem juntamente a infinito, mas mantendo h/m constante.

3.3.3 Comparação

A Figura 3.2 compara a taxa dos códigos obtidos utilizando ambos os métodos. As curvas são construídas em função do comprimento m do vetor. Os parâmetros utilizados para as curvas foram $q = 2$ e $h = 15$. Note que o método da matriz identidade no cabeçalho só é válido a partir de $m > h$. Assintoticamente, quando m tende a infinito, todas as curvas tendem à eficiência unitária.

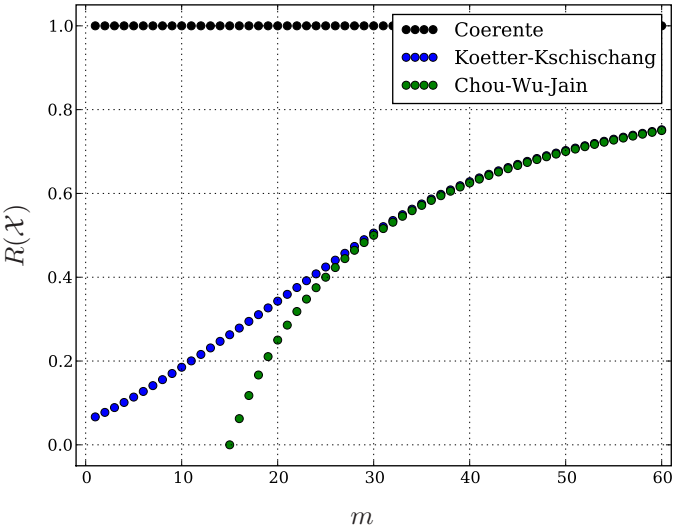


Figura 3.2: Taxas dos códigos descritos, para $q = 2$ e $h = 15$.

3.4 Codificação de rede com erros

Como visto no Capítulo 1, erros em codificação de rede foram abordados sob diversos pontos de vista. Aqui, adota-se o modelo de Silva & Kschischang [47] que estende o modelo matricial da Seção 3.1 para contemplar erros adversários.

Nessa seção, viu-se que, na ausência de erros, o nó fonte injeta uma matriz mensagem $\mathbf{X}$ de um alfabeto $\mathcal{X} \subseteq \mathbb{F}_q^{n_s \times m}$ e o nó destino em consideração coleta uma matriz $\mathbf{Y} \in \mathbb{F}_q^{n_t \times m}$. As matrizes $\mathbf{X}$ e $\mathbf{Y}$ estão relacionadas pela equação (3.1), aqui reescrita por conveniência,

$$\mathbf{Y} = \mathbf{G} \cdot \mathbf{X},$$

em que $\mathbf{G} \in \mathbb{F}_q^{n_t \times n_s}$ é a matriz de transferência do nó fonte para o nó destino. Através de um mapeamento $\hat{\phi} : \mathbb{F}_q^{n_t \times m} \longrightarrow \mathcal{W}$, o destino efetua um palpite $\hat{\mathbf{X}} = \hat{\phi}(\mathbf{Y})$ sobre a mensagem original $\mathbf{X}$. Permanece-se com $n_s = n_t = h$.

Em cima do exposto, constrói-se o modelo com erros. O erro é encarado como uma segunda entrada do sistema, como mostrado na

Figura 3.3. É considerada a injeção de n_e vetores de erros, dispostos convenientemente em uma "matriz de erros" denotada por $\mathbf{Z} \in \mathbb{F}_q^{n_e \times m}$. Assim, a equação que rege o sistema passa a ser

$$\mathbf{Y} = \mathbf{G} \cdot \mathbf{X} + \mathbf{H} \cdot \mathbf{Z}, \tag{3.3}$$

em que $\mathbf{H} \in \mathbb{F}_q^{n_t \times n_e}$ é a matriz de transferência da "fonte de erros" para o nó destino. O modelo com erros é estudado a seguir, para os casos coerente e não-coerente.

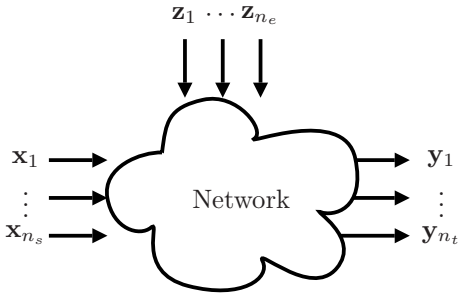


Figura 3.3: Rede de comunicação vista como um sistema MIMO com erros.

Caso coerente. No caso coerente, é assumido que o projetista tem controle (ou, pelo menos, conhecimento) sobre o código de rede interno. Em outras palavras, a matriz de transferência $\mathbf{G}$ é escolhida (dentre as permitidas pela topologia da rede) ou conhecida pelo projetista. O modelo adversário adotado é bastante pessimista e segue as seguintes hipóteses:

(*i*) O adversário conhece $\mathbf{X}$ e $\mathbf{G}$ e escolhe as matrizes $\mathbf{H}$ e $\mathbf{Z}$.

(*ii*) O projetista conhece $\mathbf{G}$, mas desconhece as matrizes $\mathbf{H}$ e $\mathbf{Z}$.

(*iii*) Para que o problema seja interessante, o número n_e de vetores de erros deve ser limitado por um inteiro $\tau \in \mathbb{N}$ (pois caso contrário, bastaria ao adversário escolher $n_e = n_s$, $\mathbf{H} = \mathbf{G}$ e $\mathbf{Z} = -\mathbf{X}$ para impossibilitar a comunicação). O parâmetro τ é conhecido por todos.

Caso não-coerente. No cenário não-coerente, o código de rede interno utilizado não é conhecido. Em termos do modelo vetorial linear, a

matriz de transferência $\mathbf{G}$ em (3.3) é desconhecida pelo projetista. Em adição, é permitido que essa matriz tenha uma deficiência de posto, o que contempla possíveis códigos internos mal-sucedidos. O modelo adversário adotado é análogo ao caso coerente, com a diferença que, agora, a matriz $\mathbf{G}$, além de desconhecida pelo projetista, é escolhida pelo adversário:

(*i*) O adversário conhece $\mathbf{X}$ e escolhe as matrizes $\mathbf{G}$, $\mathbf{H}$ e $\mathbf{Z}$.

(*ii*) O projetista desconhece as matrizes $\mathbf{G}$, $\mathbf{H}$ e $\mathbf{Z}$.

(*iii*) Além do limitante τ para o número de vetores de erros, considera-se também que e a deficiência de posto da matriz de transferência $\mathbf{G}$ é limitada por $\rho \in \mathbb{N}$ (pois caso contrário, bastaria ao adversário escolher $\mathbf{G} = \mathbf{0}$ para impossibilitar a comunicação). Os parâmetros ρ e τ são conhecidos por todos.

A solução do problema em questão se traduz no projeto de um código externo $(\mathcal{X}, \hat{\phi})$ que permita a comunicação mesmo na presença das adversidades. Desse modo, o código matricial $\mathcal{X}$ deve ser tal que seja possível descobrir qual foi a palavra-código $\mathbf{X} \in \mathcal{X}$ transmitida, através do decodificador $\hat{\phi}$, tendo conhecimento da matriz $\mathbf{Y}$ recebida. A escolha de $(\mathcal{X}, \hat{\phi})$ deve ser feita em função de $\mathbf{G}$ e τ, no caso coerente, ou ρ e τ, no caso não-coerente. Nesse contexto, destaca-se o trabalho [47], de Silva & Kschischang, que obtém métricas adequadas que guiam a escolha do código, para os casos coerente e não-coerente.

Uma forma de se obter códigos matriciais para o caso não-coerente é utilizar a abordagem de subespaços discutida na Subseção 3.3.2, em que foi proposta a transmissão de informação através de subespaços vetoriais de $\mathbb{F}_q^{h\times m}$. Naquela ocasião, utilizou-se todos os subespaços de $\mathbb{F}_q^{h\times m}$, visto que os erros eram inexistentes (*i.e.*, $\tau = 0$) e a matriz $\mathbf{G}$ era inversível (*i.e.*, $\rho = 0$). Quando essas duas hipóteses não são necessariamente verdadeiras, a ideia—apresentada e justificada por Koetter & Kschischang em [33]—é não utilizar todos os subespaços vetoriais disponíveis, mas apenas parte deles. Se está definindo, assim, um *código de subespaço*. Códigos de subespaço são o assunto do próximo capítulo.

3.5 Modelo de gerações

Além do problema de atrasos e ciclos discutido anteriormente na Seção 2.7, em muitas redes de comunicação reais os pacotes não são trans-

mitidos de maneira síncrona, como vem sendo assumido até então. Fatores como tráfego concorrente e congestionamento, por exemplo, são importantes e devem ser levados em conta em várias redes de comunicação, em particular as redes de pacotes.

Em [5], essa questão é confrontada por Chou, Wu & Jain através do *modelo de gerações*, esquema que dispensa a exigência de aciclicidade da rede além de permitir uma operação assíncrona da mesma. Ainda mais, como o modelo de gerações faz uso das ideias de codificação de rede não-coerente (apresentadas na Seção 3.3) e a assume operações aleatórias em cada nó (descritas em seguida), torna-se possível um funcionamento descentralizado do sistema. Por todas essas razões, o modelo de gerações é bastante atrativo em aplicações práticas.

No modelo proposto, a comunicação se dá através de sucessivas *gerações.* No início de cada geração, seguindo um dos métodos propostos na Seção 3.3, o nó fonte gera h pacotes $\mathbf{x}_1, \ldots, \mathbf{x}_h$, cada qual de comprimento m. O inteiro h, nesse contexto, é chamado de *tamanho da geração*. O nó fonte passa então a transmitir por seus canais de saída combinações lineares aleatórias dos pacotes $\mathbf{x}_1, \ldots, \mathbf{x}_h$.

Em cada nó da rede, um ponteiro indicando a geração atual sendo considerada nesse nó é utilizado. Quando um nó não-fonte recebe um novo pacote da geração atual ou de gerações futuras, esse é armazenado em um *buffer*; pacotes recebidos que pertencem a gerações passadas são descartados. (É assumido que todo pacote carrega um rótulo indicando a qual geração pertence.) Quando tiver oportunidade, o nó emite um pacote que consiste em uma combinação linear aleatória de todos os pacotes armazenados em seu *buffer* que pertencem à geração atual. Periodicamente, o ponteiro desse nó é avançado e pacotes da geração anterior são eliminados do seu *buffer*. Note que os ponteiros de todos os nós da rede avançam assíncrona e independentemente.

Note também que cada pacote recebido por um nó pode ou não trazer nova informação: se o pacote está fora do subespaço vetorial gerado pelos outros pacotes armazenados no *buffer* que são da mesma geração, ele é dito *inovativo* e representa nova informação. Pacotes não-inovativos não necessitam ser armazenados no *buffer* e podem ser descartados. Logo, quando um nó tem em seu *buffer* h pacotes (obrigatoriamente linearmente independentes) de uma mesma geração, ele está apto a decodificar a mensagem dessa geração.

No caso do modelo de gerações, o inteiro h não representa a taxa alcançada, devido à natureza assíncrona do sistema. A condição de

mincut (2.3) sequer precisa ser satisfeita no grafo que representa a rede; essa condição, entretanto, é satisfeita em uma treliça que representa a evolução temporal do sistema (*cf.* [54, Chapter 19]). Por fim, o trabalho original [5] apresenta mais detalhes, tais como estratégias de renovação de geração e monitoramento de taxas de inovação, a fim de melhorar o desempenho do sistema.

CAPÍTULO 4

Codificação de subespaço

No Capítulo 3 foi definido o *canal matricial* para modelar a comunicação entre o nó fonte e um nó destino qualquer de uma rede de comunicação multidifusão que opera com codificação de rede vetorial linear. Naquele mesmo capítulo, especificamente para o caso não-coerente, foi também introduzida a proposta de se transmitir informação não diretamente nos elementos da matriz injetada na rede, mas sim no subespaço vetorial gerado pelas linhas dessa matriz, através de *codificação de subespaço*.

Apesar de codificação de subespaço ter surgido como uma aplicação no controle de erros em codificação de rede não-coerente, este capítulo, com exceção da última seção, trata esses dois tópicos de maneira independente. Dessa forma, codificação de rede e o canal matricial são deixados temporariamente de lado, trabalhando-se diretamente com subespaços vetoriais.

Inicia-se considerando um canal de comunicação, o *canal de subespaço*, no qual se transmite um subespaço vetorial e se recebe outro. Em seguida, é definida uma métrica natural que mede a "distância" entre dois subespaços—a *distância de subespaço*—e são estudados códigos corretores de erros no espaço métrico definido por essa distância, os *códigos de subespaço*. São determinados limitantes inferiores e superiores que relacionam a distância mínima com o tamanho desses códigos.

Finalmente, na última seção, a conexão entre codificação de subespaço e controle de erros em codificação de rede é apresentada formalmente.

O capítulo é baseado principalmente em [33], de Koetter & Kschischang.

4.1 Canal de subespaço

Seja $q \in \mathbb{N}$ uma potência de primo e $m \in \mathbb{N}$ um número inteiro positivo. Seja $\mathbb{F}_q^m$ o espaço vetorial composto por todas as m-tuplas com elementos em $\mathbb{F}_q$. Na definição abaixo, de Koetter & Kschischang [33], a notação $\mathcal{P}(\mathbb{F}_q^m)$ representa o conjunto de todos os subespaços vetoriais de $\mathbb{F}_q^m$, objeto conhecido como *espaço projetivo* (*cf.* Apêndice A.5).

Definição. O **canal de subespaço** é um canal de comunicação discreto sem memória definido por

$$\left(\mathcal{P}(\mathbb{F}_q^m), p(\cdot|\cdot), \mathcal{P}(\mathbb{F}_q^m)\right),$$

ou seja, com alfabetos de entrada e saída dados pelo espaço projetivo $\mathcal{P}(\mathbb{F}_q^m)$ e probabilidades de transição dadas por $p(\cdot|\cdot)$. □

As probabilidades de transição $p(\cdot|\cdot)$ definem completamente o canal e devem refletir o modelo de erro adotado para o sistema. No entanto, uma vez que este trabalho considera um modelo de erro *adversário* (em detrimento de um modelo de erro *probabilístico*), a definição das probabilidades de transição não é importante. Por outro lado, é necessária a definição de uma *distância* entre os elementos do alfabeto. Esse enfoque resulta na abordagem combinatorial da *teoria da codificação* (em contraste com a abordagem estocástica da *teoria da informação*) para resolver o problema em questão.

4.2 Distância de subespaço

No canal de subespaço, transmite-se um subespaço $U \in \mathcal{P}(\mathbb{F}_q^m)$ e recebe-se outro subespaço $V \in \mathcal{P}(\mathbb{F}_q^m)$. Caso ocorra um erro, $U \neq V$. A ideia de uma distância $d_{\mathrm{S}}(\cdot,\cdot)$ entre dois subespaços é que $d_{\mathrm{S}}(U,V)$ deve mensurar o "tamanho" ou **peso** do erro que ocorreu. Esta seção tem como objetivo definir uma função $d_{\mathrm{S}}(\cdot,\cdot)$ adequada para o canal de subespaço.

4.2.1 Motivação

É natural definir a distância entre dois subespaços como sendo o número mínimo de *inserções* e *remoções* de dimensões para chegar de um subespaço a outro. Para justificar a definição da distância que será apresentada na subseção seguinte, é feito uso de uma ferramenta matemática conhecida como *diagrama de Hasse*, que representa graficamente *qualquer* conjunto finito parcialmente ordenado (*cf.* Apêndice A.2); aqui, será feita uma particularização para o caso em questão.

Com efeito, o espaço $\mathcal{P}(\mathbb{F}_q^m)$ pode ser parcialmente ordenado por $\preceq$, em que $V_1 \preceq V_2$ se e somente se V_1 é subespaço de V_2. Portanto, é possível construir o diagrama de Hasse de $\mathcal{P}(\mathbb{F}_q^m)$ com respeito a $\preceq$. Tal diagrama consiste em um grafo não-direcionado[*] cujos vértices são os elementos de $\mathcal{P}(\mathbb{F}_q^m)$. Dois vértices V_1 e V_2 estão conectados se e somente se V_1 é subespaço de V_2 e $\dim V_2 = \dim V_1 + 1$ ou vice-versa.

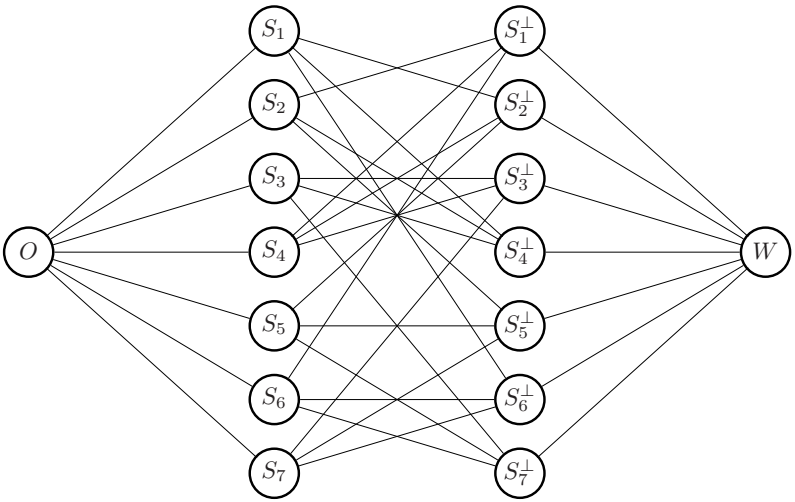


Figura 4.1: Diagrama de Hasse de $\mathcal{P}(\mathbb{F}_2^3)$.

[*]O diagrama de Hasse é originalmente um grafo direcionado, mas no presente caso é mais conveniente considerá-lo não-direcionado.

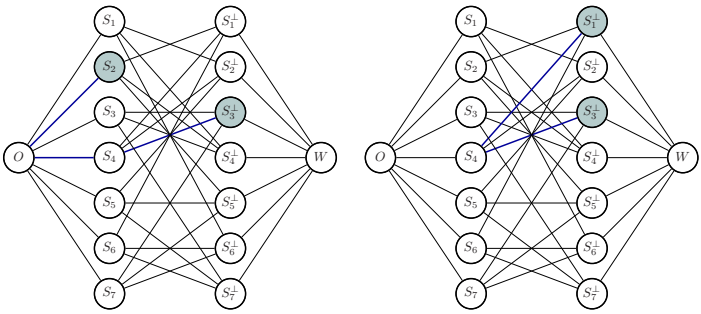
Exemplo 4.1. Os subespaços de $W = \mathbb{F}_2^3$ estão listados abaixo.

$$
\begin{array}{lcllcl}
O & = & \{000\} & W & = & \{000, 001, \ldots, 110, 111\} \\
S_1 & = & \{000, 001\} & S_1^\perp & = & \{000, 010, 100, 110\} \\
S_2 & = & \{000, 010\} & S_2^\perp & = & \{000, 001, 100, 101\} \\
S_3 & = & \{000, 011\} & S_3^\perp & = & \{000, 011, 100, 111\} \\
S_4 & = & \{000, 100\} & S_4^\perp & = & \{000, 001, 010, 011\} \\
S_5 & = & \{000, 101\} & S_5^\perp & = & \{000, 010, 101, 111\} \\
S_6 & = & \{000, 110\} & S_6^\perp & = & \{000, 001, 110, 111\} \\
S_7 & = & \{000, 111\} & S_7^\perp & = & \{000, 011, 101, 110\}
\end{array}
$$

O diagrama de Hasse de $\mathcal{P}(\mathbb{F}_2^3)$ está mostrado na Figura 4.1. □

A partir do diagrama de Hasse, é natural definir uma distância entre dois subespaços V_1, V_2 de $\mathcal{P}(\mathbb{F}_q^m)$ como o comprimento de uma *geodésica* (caminho de menor distância) ligando V_1 e V_2.

Exemplo 4.2. A Figura 4.2 mostra duas geodésicas entre elementos de $\mathcal{P}(\mathbb{F}_2^3)$.



(a) De S_2 a $S_3^\perp$; comprimento 3. **(b)** De $S_1^\perp$ a $S_3^\perp$; comprimento 2.

Figura 4.2: Geodésicas ligando elementos de $\mathcal{P}(\mathbb{F}_2^3)$.

Note que podem existir outros caminhos ligando os elementos; no entanto, os caminhos apresentados são de comprimento mínimo. □

4.2.2 Definição e propriedades

Algebricamente, o comprimento de uma geodésica ligando dois subespaços pode ser obtido de acordo com a definição a seguir, de Koetter & Kschischang [33].

DEFINIÇÃO. A **distância (de subespaço)** entre dois subespaços V_1, V_2 do espaço projetivo $\mathcal{P}(\mathbb{F}_q^m)$ é definida por

$$d_{\mathrm{S}}(V_1, V_2) = \dim(V_1 \dotplus V_2) - \dim(V_1 \cap V_2), \tag{4.1}$$

em que $\dotplus$ e $\cap$ representam, respectivamente, a soma e a interseção de subespaços. □

Utilizando (A.1) é possível reescrever a distância como

$$\begin{aligned} d_{\mathrm{S}}(V_1, V_2) &= \dim V_1 + \dim V_2 - 2\dim(V_1 \cap V_2) \\ &= 2\dim(V_1 \dotplus V_2) - \dim V_1 - \dim V_2. \end{aligned} \tag{4.2}$$

Exemplo 4.3. A definição (4.1) aplicada ao caso da Figura 4.2 fornece

$$\begin{aligned} d_{\mathrm{S}}(S_2, S_3^{\perp}) &= \dim(S_2 \dotplus S_3^{\perp}) - \dim(S_2 \cap S_3^{\perp}) \\ &= \dim W - \dim O \\ &= 3 - 0 \\ &= 3, \end{aligned}$$

$$\begin{aligned} d_{\mathrm{S}}(S_1^{\perp}, S_3^{\perp}) &= \dim(S_1^{\perp} \dotplus S_3^{\perp}) - \dim(S_1^{\perp} \cap S_3^{\perp}) \\ &= \dim W - \dim S_4 \\ &= 3 - 1 \\ &= 2, \end{aligned}$$

em que foram utilizados os detalhamentos do Exemplo 4.1. Tais resultados estão coerentes com o Exemplo 4.2. □

O teorema a seguir, também de Koetter & Kschischang [33], afirma que a distância definida por (4.1) faz com que o espaço projetivo $\mathcal{P}(\mathbb{F}_q^m)$ seja, de fato, um *espaço métrico* (*cf.* Apêndice A.6).

Teorema 4.1. *A função $d_{\mathrm{S}}(\cdot, \cdot)$ é uma métrica em $\mathcal{P}(\mathbb{F}_q^m)$.*

Demonstração. Sejam $V_1, V_2, V \in \mathcal{P}(\mathbb{F}_q^m)$. É preciso demonstrar que as três propriedades apresentadas no Apêndice A.6 são válidas.

(i) A primeira propriedade, $d_{\mathrm{S}}(V_1, V_2) = 0 \iff V_1 = V_2$, segue do fato de que $\dim(V_1 \cap V_2) \leq \dim(V_1 \dotplus V_2)$, com igualdade se e somente se $V_1 = V_2$.

(ii) A propriedade da simetria, $d_{\mathrm{S}}(V_1, V_2) = d_{\mathrm{S}}(V_2, V_1)$, é consequência da comutatividade das operações de soma e interseção de subespaços.

(iii) Por fim, a propriedade da desigualdade triangular, $d_{\mathrm{S}}(V_1, V_2) \leq d_{\mathrm{S}}(V_1, V) + d_{\mathrm{S}}(V, V_2)$, vem de

$$\begin{aligned}
&\tfrac{1}{2}\left[d_{\mathrm{S}}(V_1, V_2) - d_{\mathrm{S}}(V_1, V) - d_{\mathrm{S}}(V, V_2)\right] \\
&\quad \overset{(a)}{=} \dim(V_1 \cap V) + \dim(V_2 \cap V) \\
&\qquad - \dim V - \dim(V_1 \cap V_2) \\
&\quad \overset{(b)}{=} \dim\left[(V_1 \cap V) \dotplus (V \dotplus V_2)\right] + \dim\left[(V_1 \cap V) \cap (V \cap V_2)\right] \\
&\qquad - \dim V - \dim(V_1 \cap V_2) \\
&\quad \overset{(c)}{=} \left\{\dim\left[(V_1 \cap V) \dotplus (V \dotplus V_2)\right] - \dim V\right\} \\
&\qquad + \left\{\dim(V_1 \cap V_2 \cap V) - \dim(V_1 \cap V_2)\right\},
\end{aligned}$$

em que (a) é consequência de (4.2), (b) é consequência de (A.1) e em (c) os termos foram apenas reescritos de maneira conveniente. Mas ambas as quantidades entre chaves são não-positivas (a primeira porque V é subespaço de $(V_1 \cap V) \dotplus (V \dotplus V_2)$ e a segunda porque $V_1 \cap V_2$ é subespaço de $V_1 \cap V_2 \cap V$) e, portanto, a desigualdade triangular é válida.

Portanto, $d_{\mathrm{S}}(\cdot, \cdot)$ é uma métrica em $\mathcal{P}(\mathbb{F}_q^m)$. □

Como discutido em [18], a distância $d_{\mathrm{S}}(\cdot, \cdot)$ não é invariante à soma $\dotplus$ ou à interseção de subespaços $\cap$. Por exemplo, seja $O \in \mathcal{P}(\mathbb{F}_q^m)$ o subespaço contendo apenas o vetor nulo e $W \in \mathcal{P}(\mathbb{F}_q^m)$ o subespaço completo $\mathbb{F}_q^m$. Desse modo, tem-se que $d_{\mathrm{S}}(O, W) = m$, mas

$$d_{\mathrm{S}}(O \dotplus V, W \dotplus V) = d_{\mathrm{S}}(V, W) = m - \dim V \leq m = d_{\mathrm{S}}(O, W),$$

$$d_{\mathrm{S}}(O \cap V, W \cap V) = d_{\mathrm{S}}(O, V) = \dim V \leq m = d_{\mathrm{S}}(O, W),$$

em que $V \in \mathcal{P}(\mathbb{F}_q^m)$ é um subespaço qualquer. Em geral, vale

$$d_{\mathrm{S}}(V_1, V_2) \geq d_{\mathrm{S}}(V_1 \dot{+} V, V_2 \dot{+} V),$$

$$d_{\mathrm{S}}(V_1, V_2) \geq d_{\mathrm{S}}(V_1 \cap V, V_2 \cap V),$$

para quaisquer $V_1, V_2 \in \mathcal{P}(\mathbb{F}_q^m)$.

4.3 Códigos de subespaço

Com o espaço projetivo munido de uma métrica, é possível aplicar a *teoria da codificação para controle de erros* ao presente problema—define-se, assim, a chamada **codificação de subespaço**. O Apêndice B apresenta conceitos relativos à teoria da codificação em espaços métricos finitos arbitrários. Aqui, essa teoria é particularizada para o espaço projetivo com a métrica de subespaço.

A teoria da codificação é adequada para o modelo de erro adversário, o qual considera que o erro está limitado a um máximo peso. Assim, restringindo-se os possíveis subespaços transmitidos a um subconjunto bem escolhido, é possível detectar e até mesmo corrigir eventuais erros de transmissão que possam ocorrer.

Definição. Um **código de subespaço** $\mathcal{C}$ é um subconjunto não-vazio de $\mathcal{P}(\mathbb{F}_q^m)$. Os elementos de $\mathcal{C}$ são subespaços vetoriais de $\mathbb{F}_q^m$ e são chamados de **palavras-código**. □

Parâmetros de interesse de um código de subespaço são definidos a seguir.

Definição. Seja $\mathcal{C} \subseteq \mathcal{P}(\mathbb{F}_q^m)$ um código de subespaço.

(*i*) O **tamanho** ou **cardinalidade** do código $\mathcal{C}$ é definido por $|\mathcal{C}|$, isto é, o número de palavras-código de $\mathcal{C}$.

(*ii*) A **distância mínima** do código $\mathcal{C}$ é definida por

$$d_{\mathrm{S}}(\mathcal{C}) = \min\{d_{\mathrm{S}}(V, U) : V, U \in \mathcal{C}, V \neq U\},$$

isto é, a menor das distâncias entre pares distintos de palavras-código de $\mathcal{C}$.

Tem-se $1 \leq |\mathcal{C}| \leq \left|\mathcal{P}(\mathbb{F}_q^m)\right|$ e $1 \leq d_{\mathrm{S}}(\mathcal{C}) \leq m$. □

Exemplo 4.4. Considere o Exemplo 4.1, no qual foi apresentado o espaço projetivo $\mathcal{P}(\mathbb{F}_2^3)$. Seja o código $\mathcal{C} = \{S_1, \ldots, S_7, W\}$. O tamanho do código é $|\mathcal{C}| = 8$ e a distância mínima do código é $d_S(\mathcal{C}) = 2$. A Figura 4.3a mostra a representação gráfica do código $\mathcal{C}$. Note que para chegar de um ponto a outro do código é sempre necessário pelo menos dois desvios no diagrama de Hasse, o que confirma o valor da distância mínima. □

A cada código de subespaço $\mathcal{C}$ corresponde outro, denominado de seu *complemento ortogonal* e denotado por $\mathcal{C}^\perp$.

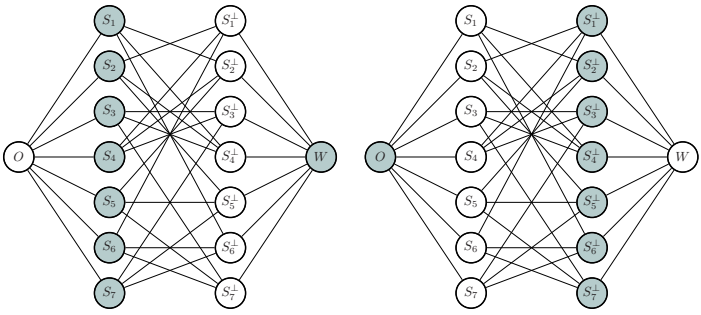
DEFINIÇÃO. O **complemento ortogonal** de um código $\mathcal{C}$ é dado por

$$\mathcal{C}^\perp = \{V^\perp : V \in \mathcal{C}\},$$

em que $V^\perp$ é o subespaço ortogonal a V, definido no Apêndice A.5. □

O complemento ortogonal tem a propriedade de conservar tanto o tamanho quanto a distância mínima do código.

Exemplo 4.5. O código complementar ao código do Exemplo 4.4 é $\mathcal{C}^\perp = \{O, S_1^\perp, \ldots, S_7^\perp\}$. O tamanho do código é $\left|\mathcal{C}^\perp\right| = 8$ e a distância mínima do código é $d_S(\mathcal{C}^\perp) = 2$. A Figura 4.3b mostra a representação gráfica do código $\mathcal{C}^\perp$. □



(a) $\mathcal{C} = \{S_1, \ldots, S_7, W\}$ **(b)** $\mathcal{C}^\perp = \{O, S_1^\perp, \ldots, S_7^\perp\}$

Figura 4.3: Representação gráfica de códigos no espaço projetivo $\mathcal{P}(\mathbb{F}_2^3)$.

4.4 Limitantes

Foca-se agora no problema de encontrar o maior tamanho possível de um código de subespaço dados os parâmetros q e m e fixada uma distância mínima d. Para tanto, define-se a notação $A_q(m, d)$, que indica a maior cardinalidade de um código em $\mathcal{P}(\mathbb{F}_q^m)$ com distância mínima d, ou seja,

$$A_q(m, d) = \max\{|\mathcal{C}| : \mathcal{C} \subseteq \mathcal{P}(\mathbb{F}_q^m), d_S(\mathcal{C}) = d\}. \tag{4.3}$$

Assim como no caso clássico (códigos no hipercubo de Hamming), a tarefa de encontrar o valor exato de $A_q(m, d)$ não é trivial. Por isso, esta seção se contenta com a determinação de limitantes inferiores e superiores para essa quantidade.

4.4.1 Esferas no espaço projetivo

Na Seção B.2 do Apêndice B são abordados os chamados *limitantes esféricos*, que se aplicam sobre qualquer espaço métrico. Aqui, esses limitantes são particularizados para o caso em questão. É, portanto, necessária a noção de *esferas* situadas no espaço projetivo. No caso de $\mathcal{P}(\mathbb{F}_q^m)$ com métrica $d_S(\cdot, \cdot)$ definida por (4.1), tem-se o seguinte.

Definição. A **esfera** de centro V_0 e raio r é dada por

$$\mathcal{B}_{(q,m)}(V_0, r) = \{V \in \mathcal{P}(\mathbb{F}_q^m) : d_S(V, V_0) \leq r\}$$

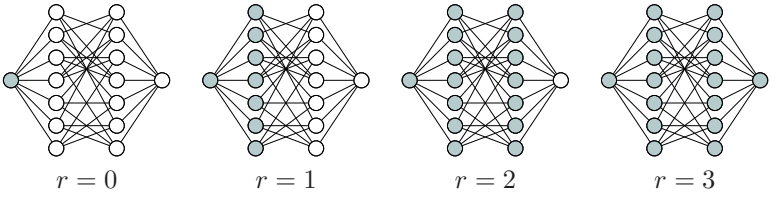
e o **volume** dessa esfera é definido por

$$\mathrm{Vol}_{(q,m)}(V_0, r) = \left|\mathcal{B}_{(q,m)}(V_0, r)\right|,$$

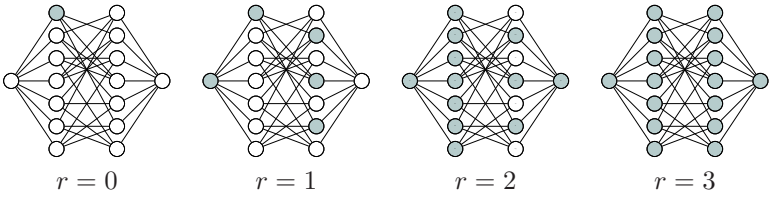
isto é, o número de pontos nela situados. □

Exemplo 4.6. A Figura 4.4 mostra esferas no espaço projetivo $\mathcal{P}(\mathbb{F}_q^m)$ com $q = 2$ e $m = 3$ centradas em O e S_1 e com raios variando de 0 a 3. Note que o volume das esferas depende do centro das mesmas; em outras palavras, o espaço projetivo com a distância de subespaço não é um espaço métrico *regular*. □

O seguinte lema determina o volume de *cascas* no espaço projetivo e facilita a obtenção do volume da esfera no Teorema 4.3. A *casca* de



(a) Esferas centradas em O.



(b) Esferas centradas em S_1.

Figura 4.4: Esferas no espaço projetivo $\mathcal{P}(\mathbb{F}_2^3)$.

centro V_0 e raio j é definida como o conjunto de pontos cuja distância até o subespaço V_0 é exatamente j, isto é, $\{V \in \mathcal{P}(\mathbb{F}_q^m) : d_{\mathrm{S}}(V, V_0) = j\}$. Esse resultado foi apresentado pela primeira vez em [12] por Etzion & Vardy. A prova do lema adapta o método utilizado em [33] para resolver um problema similar.

Lema 4.2. *O volume da casca de centro $V_0 \in \mathcal{P}(\mathbb{F}_q^m)$ e raio $j \in \{0, \ldots, m\}$ é dado por*

$$\mathrm{Vol}_{(q,m)}^{\mathrm{Casca}}(V_0, j) = \sum_{i=0}^{j} \begin{bmatrix} m-k \\ j-i \end{bmatrix}_q \begin{bmatrix} k \\ i \end{bmatrix}_q q^{i(j-i)}, \tag{4.4}$$

em que $k = \dim V_0$.

Demonstração. O volume da casca de centro V_0, em que $\dim V_0 = k$ e raio j é dado pelo número de subespaços $V \in \mathcal{P}(\mathbb{F}_q^m)$ tais que $d_{\mathrm{S}}(V_0, V) = j$. São consideradas geodésicas de V_0 para V no diagrama de Hasse da seguinte natureza: i movimentos para a esquerda, partindo de V_0 e atingindo um ponto intermediário $U \in \mathcal{P}(\mathbb{F}_q^m)$, e $j-i$ movimentos para a direita, chegando finalmente em V. Em seguida é computado o número de pontos V distintos, para $i = 0, \ldots, j$.

Fixado i, existem

$$\binom{k}{k-i}_q = \binom{k}{i}_q$$

possíveis pontos U distintos tais que U é subespaço $(k-i)$-dimensional de V_0.

Para chegar no ponto V a partir do ponto U é adicionado um subespaço $Z \in \mathcal{P}(\mathbb{F}_q^m)$ tal que Z é $(j-i)$-dimensional e a interseção de Z com V_0 seja o espaço nulo. Quantos possíveis subespaços Z dessa natureza existem? A resposta é

$$\frac{\prod_{\alpha=0}^{j-i-1} \left(q^m - q^{k+\alpha}\right)}{\prod_{\alpha=0}^{j-i-1} \left(q^{k-i+j-i} - q^{k-i+\alpha}\right)}, \tag{4.5}$$

visto que

(*i*) o *numerador* é a quantidade total de bases ordenadas distintas para subespaços Z: o primeiro vetor da base pode ser qualquer um dos q^m vetores em $\mathbb{F}_q^m$, exceto os q^k vetores em V_0; o segundo vetor da base pode ser qualquer um dos q^m vetores em $\mathbb{F}_q^m$, exceto os q^{k+1} vetores no espaço gerado por ($V_0 \cup$ "vetor já escolhido"); e assim por diante, até completar as $j-i$ dimensões de Z; e,

(*ii*) fixado um desses subespaços Z, o *denominador* é o número de possíveis bases ordenadas distintas que o geram: o primeiro vetor da base pode ser qualquer um dos $q^{k-i+j-i}$ vetores em V, exceto os q^{k-i} vetores em U; o segundo vetor da base pode ser qualquer um dos $q^{k-i+j-i}$ vetores em V, exceto os q^{k-i+1} vetores no espaço gerado por ($V \cup$ "vetor já escolhido"); e assim por diante, até completar as $j-i$ dimensões de Z.

A quantidade (4.5) pode ser simplificada para

$$\binom{m-k}{j-i}_q q^{i(j-i)}$$

e o lema segue imediatamente. □

Teorema 4.3. *O volume da esfera de centro $V_0 \in \mathcal{P}(\mathbb{F}_q^m)$ e raio $r \in \{0, \ldots, m\}$ é dado por*

$$\begin{aligned} \mathrm{Vol}_{(q,m)}(V_0, r) &= \sum_{j=0}^{r} \mathrm{Vol}^{\mathrm{Casca}}_{(q,m)}(V_0, j) \\ &= \sum_{j=0}^{r} \sum_{i=0}^{j} \binom{m-k}{j-i}_q \binom{k}{i}_q q^{i(j-i)}, \end{aligned} \tag{4.6}$$

em que $k = \dim V_0$.

Note que o volume depende apenas da dimensão k do subespaço V_0, e não diretamente de V_0. Portanto, a notação $\mathrm{Vol}_{(q,m)}(k, r)$ é utilizada para representar a quantidade calculada por (4.6).

O volume possui a seguinte propriedade, consequência da simetria da binomial Gaussiana:

$$\mathrm{Vol}_{(q,m)}(k, r) = \mathrm{Vol}_{(q,m)}(m - k, r).$$

Além disso, é possível mostrar que $\mathrm{Vol}_{(q,m)}(\cdot, r)$ é monotonicamente decrescente de 0 a $\lfloor m/2 \rfloor$, isto é, se $0 \leq k_1 \leq k_2 \leq \lfloor m/2 \rfloor$ então $\mathrm{Vol}_{(q,m)}(k_1, r) \geq \mathrm{Vol}_{(q,m)}(k_2, r)$. Em particular,

$$\mathrm{Vol}_{(q,m)}(\lfloor m/2 \rfloor, r) \leq \mathrm{Vol}_{(q,m)}(k, r) \leq \mathrm{Vol}_{(q,m)}(0, r),$$

para todo k tal que $0 \leq k \leq m$. Assim, dado um raio $r \in \{0, \ldots, m\}$, os *volumes mínimo*, *médio* e *máximo* da esfera são, respectivamente,

$$\mathrm{Vol}^{\min}_{(q,m)}(r) = \mathrm{Vol}_{(q,m)}(\lfloor m/2 \rfloor, r), \tag{4.7}$$

$$\mathrm{Vol}^{\mathrm{med}}_{(q,m)}(r) = \frac{1}{\left|\mathcal{P}(\mathbb{F}_q^m)\right|} \sum_{k=0}^{m} \binom{m}{k}_q \mathrm{Vol}_{(q,m)}(k, r), \tag{4.8}$$

$$\mathrm{Vol}^{\max}_{(q,m)}(r) = \mathrm{Vol}_{(q,m)}(0, r), \tag{4.9}$$

em que a função $\mathrm{Vol}_{(q,m)}(\cdot, r)$ é dada por (4.6).

Exemplo 4.7. Este exemplo ilustra a aplicação do Teorema 4.3. Os volumes de esferas no espaço projetivo $\mathcal{P}(\mathbb{F}_2^3)$ em função da dimensão k do centro são apresentados na tabela a seguir, para diversos valores de raio r.

	$k=0$	$k=1$	$k=2$	$k=3$
$r=0$	1	1	1	1
$r=1$	8	5	5	8
$r=2$	15	12	12	15
$r=3$	16	16	16	16

Observe que os volumes para $k=0$ e $k=1$ concordam com os resultados apresentados anteriormente no Exemplo 4.6. □

Exemplo 4.8. A Figura 4.5 apresenta o volume de esferas em $\mathcal{P}(\mathbb{F}_q^m)$ para $q=2$ e $m=20$ em função da dimensão k do centro, para valores de raio r variando de 0 a m. No eixo vertical está o logaritmo do

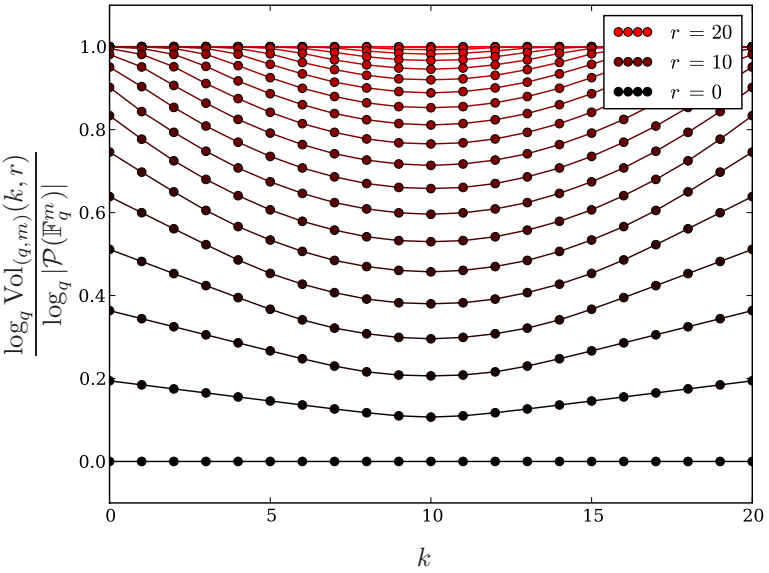


Figura 4.5: Volume de esferas no espaço projetivo $\mathcal{P}(\mathbb{F}_q^m)$ para $q=2$ e $m=20$.

volume, normalizado por $\log_q \left|\mathcal{P}(\mathbb{F}_q^m)\right|$. A propriedade da simetria e a monotonicidade decrescente de 0 até $\lfloor m/2 \rfloor$ e crescente de $\lfloor m/2 \rfloor$ até m podem ser observadas. □

4.4.2 Limitantes de Hamming e Gilbert-Varshamov

O seguinte teorema é fruto da particularização dos resultados da Seção B.2 do Apêndice B. O limitante superior fornece o chamado *limitante de Hamming*, enquanto o limitante inferior é conhecido como o *limitante de Gilbert-Varshamov*.

Teorema 4.4. *Os seguintes limitantes sobre o tamanho ótimo de códigos no espaço projetivo são válidos:*

$$\frac{\left|\mathcal{P}(\mathbb{F}_q^m)\right|}{\mathrm{Vol}^{\mathrm{med}}_{(q,m)}(d-1)} \leq A_q(m,d) \leq \frac{\left|\mathcal{P}(\mathbb{F}_q^m)\right|}{\mathrm{Vol}^{\mathrm{min}}_{(q,m)}(\lfloor (d-1)/2 \rfloor)},$$

em que $\mathrm{Vol}^{\mathrm{med}}_{(q,m)}(\cdot)$ *é dado por* (4.8) *e* $\mathrm{Vol}^{\mathrm{min}}_{(q,m)}(\cdot)$ *é dado por* (4.7).

4.4.3 Puncionamento de códigos de subespaço

O próximo limitante faz uso do conceito de *puncionamento* de tuplas, subespaços e códigos de subespaço. A operação de puncionamento consiste em "remover dimensões"; apesar dessa operação reduzir a distância entre as palavras-código, ela tem a propriedade de—sob certas condições—manter o tamanho do código.

Inicia-se definindo a operação de puncionamento de uma tupla e de uma palavra-código.

Definição. Denotada por $(\cdot)^{\blacktriangledown}$, a operação de **puncionamento de uma tupla** consiste em remover uma dada coordenada da tupla (a última coordenada, por exemplo). Equivalentemente,

$$\begin{aligned} (\cdot)^{\blacktriangledown} : \mathbb{F}_q^m &\longrightarrow \mathbb{F}_q^{m-1} \\ \mathbf{v} = (v_1, \ldots v_m) &\longmapsto \mathbf{v}^{\blacktriangledown} = (v_1, \ldots v_{m-1}). \end{aligned}$$

Além disso, define-se a operação de **puncionamento de um subespaço** por

$$\begin{aligned} (\cdot)^{\blacktriangledown} : \mathcal{P}(\mathbb{F}_q^m) &\longrightarrow \mathcal{P}(\mathbb{F}_q^{m-1}) \\ V &\longmapsto V^{\blacktriangledown} = \{\mathbf{v}^{\blacktriangledown} : \mathbf{v} \in V\}. \end{aligned} \tag{4.10}$$

Pode-se provar que essa definição é coerente no sentido de que $V^{\blacktriangledown}$ é realmente um subespaço de $\mathbb{F}_q^{m-1}$. □

A seguinte propriedade da operação de puncionamento, apresentada sem prova, será útil mais adiante.

Lema 4.5. *Sejam $V, U \in \mathcal{P}(\mathbb{F}_q^m)$. Então*

$$\dim V - 1 \leq \dim V^{\blacktriangledown} \leq \dim V,$$

$$\dim(V \dotplus U) - 1 \leq \dim(V^{\blacktriangledown} \dotplus U^{\blacktriangledown}) \leq \dim(V \dotplus U).$$

Definição. Dado um código $\mathcal{C} \subseteq \mathcal{P}(\mathbb{F}_q^m)$, o correspondente **código puncionado** é definido por

$$\mathcal{C}^{\blacktriangledown} = \{V^{\blacktriangledown} : V \in \mathcal{C}\},$$

subconjunto de $\mathcal{P}(\mathbb{F}_q^{m-1})$. □

Lema 4.6. *Seja um código $\mathcal{C} \subseteq \mathcal{P}(\mathbb{F}_q^m)$ com cardinalidade $|\mathcal{C}|$ e distância mínima $d_{\mathrm{S}}(\mathcal{C}) > 2$. Então o código puncionado $\mathcal{C}^{\blacktriangledown} \subseteq \mathcal{P}(\mathbb{F}_q^{m-1})$ tem cardinalidade $|\mathcal{C}^{\blacktriangledown}| = |\mathcal{C}|$ e distância mínima $d_{\mathrm{S}}(\mathcal{C}) - 2 \leq d_{\mathrm{S}}(\mathcal{C}^{\blacktriangledown}) \leq d_{\mathrm{S}}(\mathcal{C})$.*

Demonstração. Sejam $V, U \in \mathcal{C}$ distintos. Então, pelo Lema 4.5,

$$\begin{aligned} d_{\mathrm{S}}(V^{\blacktriangledown}, U^{\blacktriangledown}) &= 2\dim(V^{\blacktriangledown} \dotplus U^{\blacktriangledown}) - \dim V^{\blacktriangledown} - \dim U^{\blacktriangledown} \\ &\geq 2\dim(V \dotplus U) - 2 - \dim V - \dim U \\ &= d_{\mathrm{S}}(V, U) - 2 \end{aligned}$$

e portanto $d_{\mathrm{S}}(\mathcal{C}^{\blacktriangledown}) \geq d_{\mathrm{S}}(\mathcal{C}) - 2$. Além disso, como $d_{\mathrm{S}}(V, U) > 2$, tem-se que $d_{\mathrm{S}}(V^{\blacktriangledown}, U^{\blacktriangledown}) > 0$, o que faz com que $V^{\blacktriangledown}$ e $U^{\blacktriangledown}$ sejam distintos pois $d_{\mathrm{S}}(\cdot, \cdot)$ é uma métrica. Portanto, $|\mathcal{C}^{\blacktriangledown}| = |\mathcal{C}|$. □

4.4.4 Limitante de Singleton

O limitante superior de Singleton é obtido verificando-se o que acontece com um código se o puncionarmos repetidas vezes. O resultado a seguir segue o método descrito na Seção B.3 do Apêndice B.

Teorema 4.7. *O seguinte limitante sobre o tamanho ótimo de códigos no espaço projetivo é válido:*

$$A_q(m,d) \leq \left|\mathcal{P}(\mathbb{F}_q^{m-\lfloor (d-1)/2 \rfloor})\right|,$$

em que $\left|\mathcal{P}(\mathbb{F}_q^{(\cdot)})\right|$ *é dado por* (A.4).

Demonstração. Seja um código $\mathcal{C} \subseteq \mathcal{P}(\mathbb{F}_q^m)$ qualquer com distância mínima $d_{\mathrm{S}}(\mathcal{C}) = d$. Puncionando o código $\mathcal{C}$ um total de $\lfloor (d-1)/2 \rfloor$ vezes, pelo Lema 4.6, obtém-se um código $\mathcal{C}' = \mathcal{C}^{\blacktriangledown \cdots \blacktriangledown} \subseteq \mathcal{P}(\mathbb{F}_q^{m-\lfloor (d-1)/2 \rfloor})$ com cardinalidade $|\mathcal{C}'| = |\mathcal{C}|$ e distância mínima $d_{\mathrm{S}}(\mathcal{C}') \geq 1$. Portanto,

$$|\mathcal{C}| = |\mathcal{C}'| \leq \left|\mathcal{P}(\mathbb{F}_q^{m-\lfloor (d-1)/2 \rfloor})\right|$$

e o teorema segue. □

4.4.5 Gráficos

Os limitantes superiores de Hamming e Singleton, bem como o limitante inferior de Gilbert-Varshamov estão ilustrados na Figura 4.6 para o espaço projetivo $\mathcal{P}(\mathbb{F}_q^m)$ com $q = 2$ e $m = 20$. No eixo vertical estão os logaritmos dos três limitantes sobre $A_q(m,d)$, normalizados por $\log_q \left|\mathcal{P}(\mathbb{F}_q^m)\right|$.

4.5 Códigos de dimensão constante

Uma classe importante de códigos de subespaço são os chamados *códigos de dimensão constante*, códigos nos quais todas as palavras-código possuem uma mesma dimensão k. São análogos aos *códigos clássicos de peso constante*, isto é, códigos clássicos nos quais todas as palavras-código possuem o mesmo peso de Hamming.

Os códigos de dimensão constante são também conhecidos como *códigos no grassmanniano*, porque o subconjunto

$$\mathcal{P}(\mathbb{F}_q^m, k) = \{V \in \mathcal{P}(\mathbb{F}_q^m) : \dim V = k\},$$

do espaço projetivo $\mathcal{P}(\mathbb{F}_q^m)$ contendo todos os subespaços de uma dada dimensão k é chamado de *grassmanniano* (*cf.* Apêndice A.5).

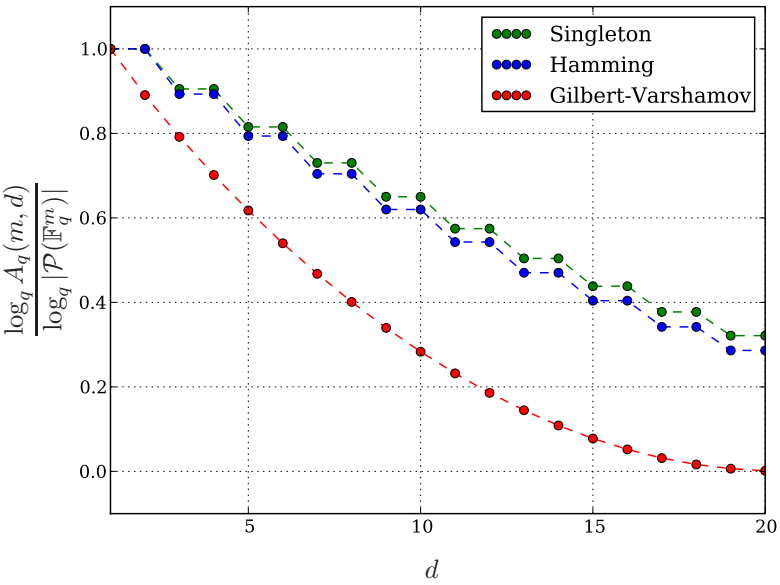


Figura 4.6: Limitantes sobre $A_q(m, d)$ para $q = 2$ e $m = 20$.

O grassmanniano em si é também um espaço métrico com a métrica $d_S(\cdot, \cdot)$, com a propriedade de que se V_1 e V_2 são dois subespaços do grassmanniano, então $d_S(V_1, V_2)$ assume apenas valores pares. Assim, é natural considerar o grafo com vértices $\{V \in \mathcal{P}(\mathbb{F}_q^m, k)\}$ e arestas $\{(V_1, V_2) \in \mathcal{P}(\mathbb{F}_q^m, k)^2 : d_S(V_1, V_2) = 2\}$. Esse grafo é conhecido como *diagrama de Grassmann* e tem a característica de ser regular com relação à distância, de modo que, no grassmanniano, esferas de mesmo raio têm sempre o mesmo volume.

Códigos no grassmanniano já foram considerados mesmo antes do trabalho de Koetter & Kschischang [33], mas em diferentes contextos. Essa referência, além de apontar alguns desses trabalhos anteriores, também deduz limitantes específicos para o grassmanniano, análogos aos de Hamming, Gilbert-Varshamov e Singleton e apresenta construções de códigos de dimensão constante através de procedimentos similares àqueles utilizados no projeto de códigos clássicos Reed-Solomon [27, Chapter 5].

Outras referências que apresentam construções de códigos de dimensão constante são Silva, Kschischang & Koetter [48], Kohnert & Kurz [36], Skachek [51], Gadouleau & Yan [19] e Etzion & Silberstein [11].

4.6 Aplicação em codificação de rede

Esta seção formaliza a conexão entre codificação de subespaço e o problema de controle de erros em codificação de rede não-coerente apresentado na Seção 3.4.

No contexto da codificação de rede externa, para se transmitir um subespaço vetorial $V \in \mathcal{P}(\mathbb{F}_q^m)$, o nó fonte deve injetar no canal matricial uma matriz $\mathbf{X}$ cujas linhas são m-tuplas que geram tal subespaço. Em outras palavras, deve-se ter $\langle \mathbf{X} \rangle = V$. Uma vez que a quantidade de linhas da matriz $\mathbf{X}$ deve ser de pelo menos $\dim V$, a definição a seguir se faz apropriada.

DEFINIÇÃO. A **dimensão máxima** de um código de subespaço $\mathcal{C}$ é definida por

$$\ell(\mathcal{C}) = \max\{\dim V : V \in \mathcal{C}\}.$$

Tem-se $1 \leq \ell(\mathcal{C}) \leq m$. □

Exemplo 4.9. O código $\mathcal{C}$ do Exemplo 4.4 tem $\ell(\mathcal{C}) = 3$, enquanto o código $\mathcal{C}^\perp$ do Exemplo 4.5 tem $\ell(\mathcal{C}^\perp) = 2$. □

A seguinte definição segue naturalmente e considera um código matricial formado por matrizes que geram os subespaços de um dado código de subespaço.

DEFINIÇÃO. Seja $\mathcal{C} \subseteq \mathcal{P}(\mathbb{F}_q^m)$ um código de subespaço e $\mathcal{X} \subseteq \mathbb{F}_q^{h \times m}$ um código matricial, em que $\ell(\mathcal{C}) = h$. O código $\mathcal{X}$ é dito ser **originário** do código $\mathcal{C}$ se $\{\langle \mathbf{X} \rangle : \mathbf{X} \in \mathcal{X}\} = \mathcal{C}$ e $|\mathcal{X}| = |\mathcal{C}|$. □

Observação. A cada código matricial $\mathcal{X}$ corresponde um único código de subespaço $\mathcal{C} = \{\langle \mathbf{X} \rangle : \mathbf{X} \in \mathcal{X}\}$. Por outro lado, podem existir diferentes códigos matriciais $\mathcal{X}$ originários do mesmo código de subespaço $\mathcal{C}$. Do ponto de vista de codificação de rede não-coerente eles são todos "indistinguíveis".

Se $\mathcal{X}$ é um código matricial originário do código de subespaço $\mathcal{C} \subseteq \mathcal{P}(\mathbb{F}_q^m)$ com cardinalidade $|\mathcal{C}|$ e dimensão máxima $\ell(\mathcal{C})$, sua taxa é dada por

$$R(\mathcal{X}) = \frac{\log_q |\mathcal{C}|}{m \cdot \ell(\mathcal{C})},$$

o que mostra que a dimensão máxima do código de subespaço é um parâmetro importante no problema em questão.

Exemplo 4.10. Considere os códigos de subespaço $\mathcal{C}$ e $\mathcal{C}^\perp$ em $\mathcal{P}(\mathbb{F}_q^m)$, com $q = 2$ e $m = 3$, dos Exemplos 4.4 e 4.5, dados por

$$\mathcal{C} = \{S_1, S_2, \ldots, S_7, W\},$$

$$\mathcal{C}^\perp = \{O, S_1^\perp, S_2^\perp, \ldots, S_7^\perp\}.$$

Tem-se $\ell(\mathcal{C}) = 3$ e $\ell(\mathcal{C}^\perp) = 2$. Um possível código matricial $\mathcal{X}_1 \in \mathbb{F}_q^{\ell(\mathcal{C})\times m}$ originário de $\mathcal{C}$ é dado por

$$\mathcal{X}_1 = \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \ldots, \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\},$$

enquanto que um possível código matricial $\mathcal{X}_2 \in \mathbb{F}_q^{\ell(\mathcal{C}^\perp)\times m}$ originário de $\mathcal{C}^\perp$ é dado por

$$\mathcal{X}_2 = \left\{ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \ldots, \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \right\}.$$

Os códigos têm taxas

$$R(\mathcal{X}_1) = \frac{\log_2 8}{3 \cdot 3} = \frac{1}{3} \qquad \text{e} \qquad R(\mathcal{X}_2) = \frac{\log_2 8}{3 \cdot 2} = \frac{1}{2},$$

respectivamente. □

A seguir, apresenta-se um resultado que relaciona a distância mínima do código de subespaço $\mathcal{C}$ com a capacidade de correção de erros de um código matricial originário em termos dos parâmetros τ e ρ do modelo da Seção 3.4. Resultados nessa linha foram obtidos no trabalho original [33] e, posteriormente, em [47, 48], sendo esse último aquele no qual os próximos lema e teorema são baseados.

Lema 4.8. *Se* $\mathbf{X}, \mathbf{Y} \in \mathbb{F}_q^{h\times m}$ *e* $\mathbf{A} \in \mathbb{F}_q^{h\times h}$ *são matrizes quaisquer então*

$$d_{\mathrm{S}}(\langle \mathbf{X} \rangle, \langle \mathbf{A}\cdot\mathbf{X} \rangle) \leq \operatorname{rankdef} \mathbf{A},$$

$$d_{\mathrm{S}}(\langle \mathbf{X} \rangle, \langle \mathbf{Y} \rangle) \leq 2\operatorname{rank}(\mathbf{Y}-\mathbf{X}).$$

Demonstração. Uma vez que

$$\dim(\langle \mathbf{X} \rangle \dotplus \langle \mathbf{Y} \rangle) = \operatorname{rank} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix},$$

a segunda linha de (4.2) fornece

$$d_{\mathrm{S}}(\langle \mathbf{X} \rangle, \langle \mathbf{Y} \rangle) = 2\operatorname{rank} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} - \operatorname{rank}\mathbf{X} - \operatorname{rank}\mathbf{Y}.$$

Desse modo, a primeira proposição segue de

$$\begin{aligned} d_{\mathrm{S}}(\langle \mathbf{X} \rangle, \langle \mathbf{A}\cdot\mathbf{X} \rangle) &= 2\operatorname{rank} \begin{bmatrix} \mathbf{X} \\ \mathbf{A}\cdot\mathbf{X} \end{bmatrix} - \operatorname{rank}\mathbf{X} - \operatorname{rank}(\mathbf{A}\cdot\mathbf{X}) \\ &= 2\operatorname{rank}\mathbf{X} - \operatorname{rank}\mathbf{X} - \operatorname{rank}(\mathbf{A}\cdot\mathbf{X}) \\ &= \operatorname{rank}\mathbf{X} - \operatorname{rank}(\mathbf{A}\cdot\mathbf{X}) \\ &\leq h - \operatorname{rank}\mathbf{A} \\ &= \operatorname{rankdef}\mathbf{A}, \end{aligned}$$

em que a desigualdade segue da inequação de Sylvester. Já a segunda proposição segue de

$$\begin{aligned} d_{\mathrm{S}}(\langle \mathbf{X} \rangle, \langle \mathbf{Y} \rangle) &= 2\operatorname{rank} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} - \operatorname{rank}\mathbf{X} - \operatorname{rank}\mathbf{Y} \\ &\leq 2\operatorname{rank}(\mathbf{Y}-\mathbf{X}), \end{aligned}$$

em que a desigualdade segue da soma de

$$\operatorname{rank} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \overset{(a)}{=} \operatorname{rank} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y}-\mathbf{X} \end{bmatrix} \overset{(b)}{\leq} \operatorname{rank}(\mathbf{Y}-\mathbf{X}) + \operatorname{rank}\mathbf{X},$$

$$\operatorname{rank} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \overset{(a)}{=} \operatorname{rank} \begin{bmatrix} \mathbf{Y}-\mathbf{X} \\ \mathbf{Y} \end{bmatrix} \overset{(b)}{\leq} \operatorname{rank}(\mathbf{Y}-\mathbf{X}) + \operatorname{rank}\mathbf{Y},$$

em que (a) segue do fato de que operações elementares de linha não

alteram o posto de uma matriz e (b) é equivalente a $\dim(\langle \mathbf{A} \rangle \dotplus \langle \mathbf{B} \rangle) \leq \dim \langle \mathbf{A} \rangle + \dim \langle \mathbf{B} \rangle$, o que é assegurado por (A.1). □

Teorema 4.9. *Seja $\mathcal{C}$ um código de subespaço com distância de subespaço mínima $d_{\mathrm{S}}(\mathcal{C}) = d$ e $\mathcal{X}$ um código matricial originário de $\mathcal{C}$. Se $d > 2(\rho + 2\tau)$ então $\mathcal{X}$ é bem-sucedido no canal matricial não-coerente com parâmetros ρ e τ.*

Demonstração. Seja $V \in \mathcal{C}$ e $\mathbf{X} \in \mathcal{X}$ tal que $\langle \mathbf{X} \rangle = V$. Seja $\mathbf{Y} = \mathbf{G} \cdot \mathbf{X} + \mathbf{H} \cdot \mathbf{Z}$ e $U = \langle \mathbf{Y} \rangle$. Tem-se

$$\begin{aligned} d_{\mathrm{S}}(V, U) &= d_{\mathrm{S}}(\langle \mathbf{X} \rangle, \langle \mathbf{Y} \rangle) \\ &\overset{(a)}{\leq} d_{\mathrm{S}}(\langle \mathbf{X} \rangle, \langle \mathbf{G} \cdot \mathbf{X} \rangle) + d_{\mathrm{S}}(\langle \mathbf{G} \cdot \mathbf{X} \rangle, \langle \mathbf{Y} \rangle) \\ &\overset{(b)}{\leq} \operatorname{rankdef} \mathbf{G} + 2 \operatorname{rank}(\mathbf{H} \cdot \mathbf{Z}) \\ &\overset{(c)}{\leq} \rho + 2\tau \\ &\overset{(d)}{\leq} \lfloor (d-1)/2 \rfloor, \end{aligned}$$

em que (a) segue da desigualdade triangular para $d_{\mathrm{S}}(\cdot, \cdot)$; (b) segue do Lema 4.8 e (c) segue da definição dos parâmetros ρ e τ e (d) segue da hipótese do teorema. Assim, U está dentro da esfera de centro V e raio $\lfloor (d-1)/2 \rfloor$. Como, por hipótese, não existe nenhuma outra palavra-código dentro dessa esfera, um decodificador de mínima distância de subespaço produz V dado U ou, equivalentemente, $\mathbf{X}$ dado $\mathbf{Y}$. □

Observação. O Teorema 4.9 apresenta apenas uma condição *suficiente* para que um código matricial $\mathcal{X}$ originário do código de subespaço $\mathcal{C}$ seja bem-sucedido no canal matricial não-coerente. Silva & Kschischang discutem o problema detalhadamente em [47], apresentando condições *necessárias e suficientes* para que $\mathcal{X}$ seja bem-sucedido e concluem que a **distância de injeção**, definida por

$$\begin{aligned} d_{\mathrm{I}}(V, U) &= \max\{\dim V, \dim U\} - \dim(U \cap V) \\ &= \dim(U \dotplus V) - \min\{\dim V, \dim U\} \\ &= \tfrac{1}{2} d_{\mathrm{S}}(V, U) + \tfrac{1}{2} |\dim V - \dim U| \end{aligned}$$

é a métrica adequada ao problema em questão. É provado o seguinte resultado (contraste com o Teorema 4.9):

> Seja $\mathcal{C}$ um código de subespaço com distância de *injeção* mínima $d_{\mathrm{I}}(\mathcal{C}) = d$ e $\mathcal{X}$ um código matricial originário de $\mathcal{C}$. Então $d > \rho + 2\tau$ *se e somente se* $\mathcal{X}$ é bem-sucedido no canal matricial não-coerente com parâmetros ρ e τ.

Recentemente, códigos de subespaço utilizando a métrica de injeção foram apresentados por Khaleghi & Kschischang em [31].

CAPÍTULO 5

Codificação de subespaço *multishot*

Os códigos estudados no Capítulo 4 utilizam o canal de subespaço apenas uma vez. Neste capítulo é explorada a ideia de utilizar o canal várias vezes, codificando a informação em uma *sequência* de subespaços a ser transmitida e não apenas em um único subespaço.

Com efeito, um dos problemas fundamentais no contexto da codificação de subespaço é a obtenção de códigos com boas taxas e boas capacidades de correção de erros. Para atingir ambos os objetivos simultaneamente pode ser inevitável o aumento do tamanho do corpo finito, q, ou do comprimento do vetor, m. Este capítulo apresenta uma terceira alternativa: o aumento do número de usos do canal de subespaço, n, e define os chamados *códigos de subespaço multishot*.

Com isso em mente, há dois motivos que justificam o estudo de códigos de subespaço *multishot*. Primeiro, a aplicação sob consideração pode ter seus parâmetros imutáveis, no sentido de que pode não ser possível o ajuste do tamanho q do corpo ou do comprimento m do vetor. E segundo, mesmo que tais parâmetros sejam modificáveis, questões de complexidade podem ser determinantes—por exemplo, códigos *one-shot* em $\mathcal{P}(\mathbb{F}_q^{mn})$ são consideravelmente mais complexos que códigos *n-shot* sobre $\mathcal{P}(\mathbb{F}_q^m)$.

O presente capítulo, considerado a contribuição deste trabalho, inicia com a construção de um espaço métrico apropriado ao problema e a definição de códigos corretores de erro sobre tal espaço, os *códigos de subespaço multishot*. Uma motivação para o uso repetido do canal de subespaço é apresentada em seguida. É mostrado, então, que códigos de subespaço *multishot* podem ser vistos como uma classe especial de códigos de subespaço *one-shot*. Após isso, são obtidos limitantes sobre o tamanho ótimo dos códigos. É também proposta uma técnica para o projeto de códigos de subespaço *multishot* baseada em *codificação multinível*. O capítulo finaliza com a aplicação da codificação de subespaço *multishot* ao controle de erros em codificação de rede não-coerente.

Parte dos resultados deste capítulo pode ser encontrada em [42], artigo publicado por Nóbrega & Uchôa-Filho. Deseja-se agradecer a Frank Kschischang e Danilo Silva pelas críticas e sugestões que, por falta de tempo e espaço, não puderam ser incorporadas ao artigo mas que, agora, estão presentes neste capítulo.

5.1 Definições

Seja $\mathcal{P}(\mathbb{F}_q^m)$ um dado espaço projetivo e n um inteiro positivo. Esta seção apresenta definições para a *codificação de subespaço multishot*, no qual o canal de subespaço definido por $\mathcal{P}(\mathbb{F}_q^m)$ é utilizado n vezes consecutivamente. Permanece-se com a abordagem de erros adversária, de modo que as probabilidades de transição do canal continuam irrelevantes.

São considerados *códigos de bloco* sobre o espaço projetivo, ou seja, aqueles nos quais as palavras-código consistem em um bloco (tupla) de comprimento n com componentes em $\mathcal{P}(\mathbb{F}_q^m)$. As definições a seguir criam um espaço métrico que reflete essa proposta.

DEFINIÇÃO. A **n-ésima extensão do espaço projetivo** $\mathcal{P}(\mathbb{F}_q^m)$ é definida por $\mathcal{P}(\mathbb{F}_q^m)^n$, isto é, a n-ésima potência cartesiana do espaço projetivo. Dessa forma, elementos de $\mathcal{P}(\mathbb{F}_q^m)^n$ são n-tuplas tendo como componentes subespaços do espaço projetivo original $\mathcal{P}(\mathbb{F}_q^m)$. □

A seguinte definição de *distância* entre pontos do espaço projetivo estendido $\mathcal{P}(\mathbb{F}_q^m)^n$ simplesmente acumula os *pesos* dos erros ocorridos em cada transmissão.

DEFINIÇÃO. A **distância (de subespaço estendida)** entre dois elementos $\mathbf{V} = (V_1, \ldots, V_n)$ e $\mathbf{U} = (U_1, \ldots, U_n)$ do espaço projetivo estendido $\mathcal{P}(\mathbb{F}_q^m)^n$ é definida como

$$d_{\mathrm{S}}(\mathbf{V}, \mathbf{U}) = \sum_{i=1}^{n} d_{\mathrm{S}}(V_i, U_i), \tag{5.1}$$

em que $d_{\mathrm{S}}(\cdot, \cdot)$ no lado direito da equação é dada por (4.1). Tem-se $1 \leq d_{\mathrm{S}}(\mathbf{V}, \mathbf{U}) \leq mn$. □

Assim sendo, no caso *multishot*, transmite-se uma n-tupla de subespaços, $\mathbf{V} = (V_1, \ldots V_n)$, e recebe-se outra n-tupla de subespaços, $\mathbf{U} = (U_1, \ldots, U_n)$. Na ausência de erros, $\mathbf{V} = \mathbf{U}$. Caso contrário, é dito que um erro de **peso total** $d_{\mathrm{S}}(\mathbf{U}, \mathbf{V})$ ocorreu.

Observação. Essa medida desconsidera como os pesos dos erros se "distribuem" ao longo das n transmissões. Por exemplo, dois erros "simples" de subespaço ocorrendo em transmissões diferentes são equivalentes a um erro "duplo" de subespaço ocorrendo em uma única transmissão, pois ambos os casos conduzem a um peso total de 2.

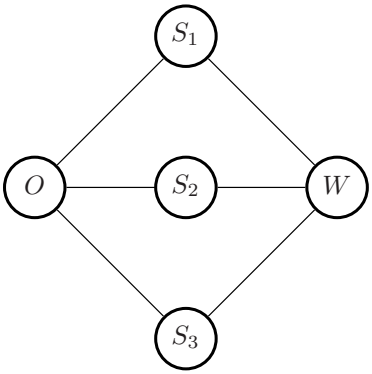
Pode-se provar que a distância definida por (5.1) é uma métrica em $\mathcal{P}(\mathbb{F}_q^m)^n$. Assim, é possível definir códigos sobre esse espaço métrico.

DEFINIÇÃO. Um **código de subespaço de bloco** é definido como um subconjunto não-vazio $\mathcal{C} \subseteq \mathcal{P}(\mathbb{F}_q^m)^n$; é também denominado de **código (de subespaço)** n-***shot*** ou **código (de subespaço)** ***multishot***. □

Parâmetros como o tamanho do código e a distância mínima do código também são válidos e são definidos da maneira usual.

5.2 Motivação

Suponha que se queira um código de subespaço *multishot* utilizando o espaço projetivo $\mathcal{P}(\mathbb{F}_2^2)$ (cujo diagrama de Hasse está mostrado na Figura 5.1) duas vezes consecutivas. Suponha também que se deseje detectar um único erro ocorrendo em qualquer uma das $n = 2$ transmissões. Assim sendo, é suficiente encontrar um código 2-*shot* com distância mínima $d = 2$.



$$O = \{00\}$$
$$S_1 = \{00, 01\}$$
$$S_2 = \{00, 10\}$$
$$S_3 = \{00, 11\}$$
$$W = \{00, 01, 10, 11\}$$

Figura 5.1: Diagrama de Hasse de $\mathcal{P}(\mathbb{F}_2^2)$.

Uma primeira abordagem é simplesmente estender o melhor código *one-shot* em $\mathcal{P}(\mathbb{F}_2^2)$ com distância mínima $d = 2$, que é

$$\mathcal{C}_1' = \{S_1, S_2, S_3\}.$$

Fazendo isto obtém-se o código

$$\begin{aligned} \mathcal{C}_1 &= \mathcal{C}_1' \times \mathcal{C}_1' \\ &= \{S_1S_1, S_1S_2, S_1S_3, S_2S_1, S_2S_2, S_2S_3, S_3S_1, S_3S_2, S_3S_3\}, \end{aligned}$$

com $|\mathcal{C}_1| = 9$.

É possível encontrar código melhor? Como tentativa, considere o espaço projetivo $\mathcal{P}(\mathbb{F}_q^m)$ como sendo o alfabeto de um *código clássico*. De acordo, escolha qualquer mapeamento bijetivo entre $\mathcal{P}(\mathbb{F}_q^m) = \{O, S_1, S_2, S_3, W\}$ e $\mathbb{Z}_5 = \{0, 1, 2, 3, 4\}$, por exemplo, $O \mapsto 0$, $S_1 \mapsto 1$, $S_2 \mapsto 2$, $S_3 \mapsto 3$ e $W \mapsto 4$. Um código clássico ótimo de comprimento 2 sobre $\mathbb{Z}_5$ com distância mínima de Hamming 2 é o *código de repetição*, dado por

$$\mathcal{H}_2 = \{00, 11, 22, 33, 44\},$$

que é mapeado de volta para

$$\mathcal{C}_2 = \{OO, S_1S_1, S_2S_2, S_3S_3, WW\},$$

com $|\mathcal{C}_2| = 5$, menor que $|\mathcal{C}_1|$.

Essa segunda abordagem não foi bem-sucedida porque desconsiderou a estrutura de subespaço por trás de $\mathcal{P}(\mathbb{F}_2^2)$ e fez uso apenas de co-

dificação clássica. Se se deseja atingir melhores resultados, é necessário projetar códigos no espaço métrico $\mathcal{P}(\mathbb{F}_2^2)^2$ (mostrado na Figura 5.2), o que leva em conta tanto a estrutura de subespaço quanto a evolução temporal. De fato, o melhor código com distância mínima 2 para esse caso é dado por

$$\mathcal{C}_3 = \{OO, S_1S_1, S_1S_2, S_1S_3, OW, S_2S_1, S_2S_2, S_2S_3, WO, S_3S_1, \\ S_3S_2, S_3S_3, WW\},$$

com $|\mathcal{C}_3| = 13$. Esse código pode ser obtido por meio da construção multinível apresentada na Seção 5.5.

Observação. Deve-se tomar cuidado ao se comparar a capacidade de correção/deteção de erros de dois códigos *multishot*. Por exemplo, considere os códigos $\mathcal{C}_1$ e $\mathcal{C}_3$ definidos anteriormente. Como ambos têm distância mínima $d = 2$, ambos são capazes de detectar qualquer padrão de erro de peso total 1 e podem, portanto, ser rotulados como "códigos detectores de 1 erro" no canal de subespaço de 2 *shots*. No entanto, o código $\mathcal{C}_1$ tem a capacidade de detectar 2 erros, desde que cada erro ocorra em uma transmissão diferente. Apesar disso, $\mathcal{C}_1$ não pode ser denominado de "código detector de 2 erros", visto que não pode detectar *todos* os padrões de erro de peso total 2 ou menos (por exemplo, um erro duplo ocorrendo em qualquer transmissão).

5.3 Relação com códigos de subespaço *one-shot*

Obviamente, códigos de subespaço *one-shot* são apenas um caso especial de códigos de subespaço *n-shot*—quando $n = 1$. Nesta seção é mostrado como o contrário também pode ser interpretado como verdadeiro sob certa ótica.

No caso da codificação clássica, a extensão $(\mathbb{Z}_q^m)^n$ é equivalente a $\mathbb{Z}_q^{mn}$, no sentido em que existe uma *isometria* entre esses dois espaços métricos (isto é, uma bijeção que preserva a distância entre os pontos). Esse fato, no entanto, não ocorre na codificação de subespaço; contudo, a n-ésima extensão de um espaço projetivo, $\mathcal{P}(\mathbb{F}_q^m)^n$, ainda pode ser vista como um "subconjunto" do espaço projetivo maior $\mathcal{P}(\mathbb{F}_q^{mn})$.

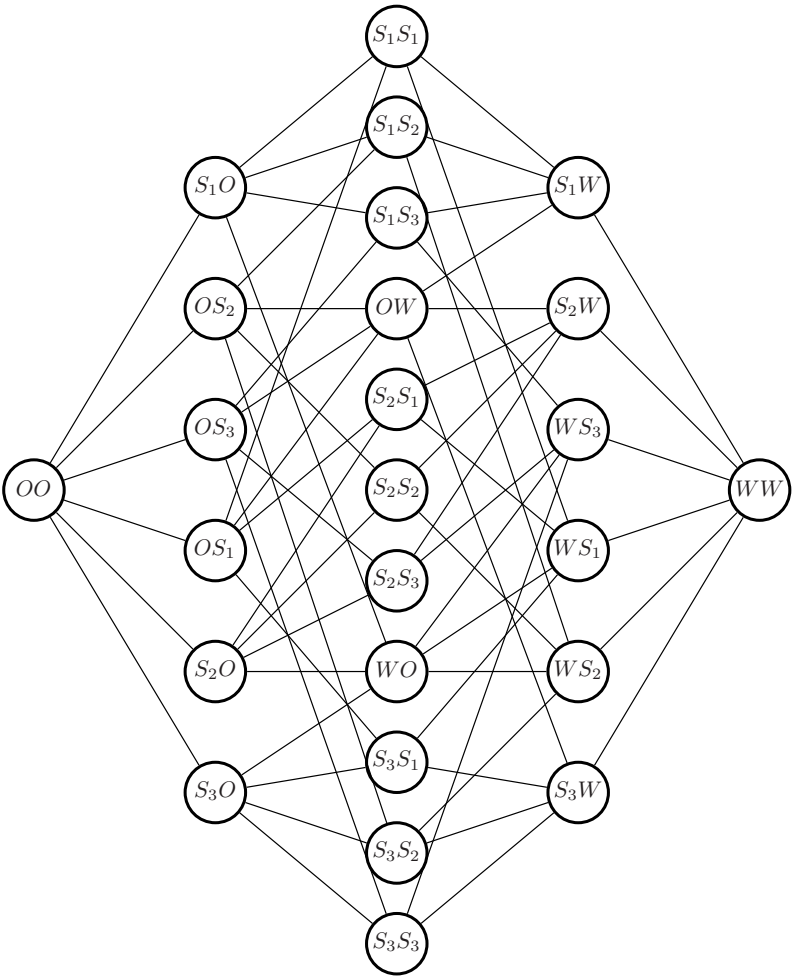


Figura 5.2: Diagrama do espaço estendido $\mathcal{P}(\mathbb{F}_2^2)^2$.

A Figura 5.3 compara os tamanhos de $\mathcal{P}(\mathbb{F}_q^{mn})$ e $\mathcal{P}(\mathbb{F}_q^m)^n$.

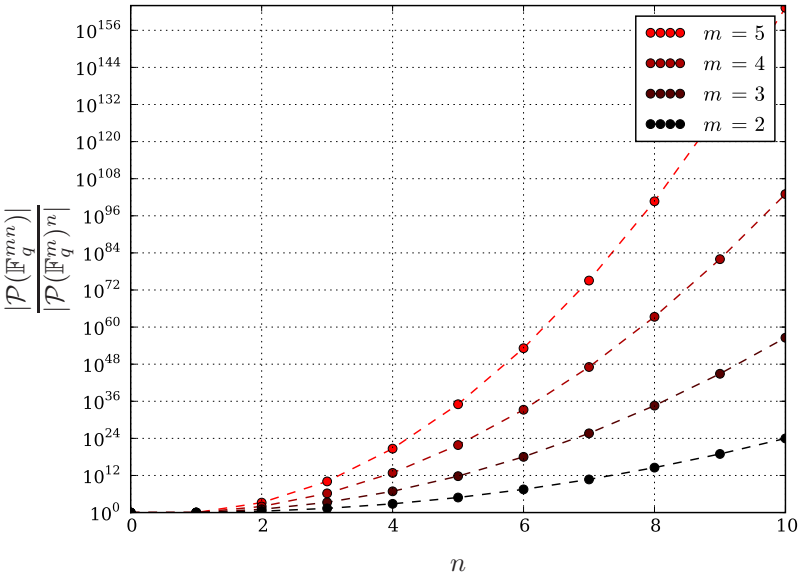


Figura 5.3: Comparação entre os tamanhos de $\mathcal{P}(\mathbb{F}_q^{mn})$ e $\mathcal{P}(\mathbb{F}_q^m)^n$ para $q = 2$.

Para ver como códigos de subespaço *multishot* em $\mathcal{P}(\mathbb{F}_q^m)^n$ formam uma classe especial de códigos *one-shot* em $\mathcal{P}(\mathbb{F}_q^{mn})$, considere o mapeamento injetivo

$$\begin{aligned} f : \mathcal{P}(\mathbb{F}_q^m)^n &\longrightarrow \mathcal{P}(\mathbb{F}_q^{mn}) \\ \mathbf{V} &\longmapsto V_1 \times \cdots \times V_n, \end{aligned} \tag{5.2}$$

em que $\mathbf{V} = (V_1, \ldots, V_n) \in \mathcal{P}(\mathbb{F}_q^m)^n$.

Observação. A rigor, o produto cartesiano $V_1 \times \cdots \times V_n$ ainda pertence a $\mathcal{P}(\mathbb{F}_q^m)^n$ mas aqui é feito o abuso de notação

$$\begin{aligned} ((v_{1,1}, \ldots, v_{1,m}), (v_{2,1}, \ldots, v_{2,m}), \ldots, (v_{n,1}, \ldots, v_{n,m})) = \\ (v_{1,1}, \ldots, v_{1,m}, v_{2,1}, \ldots, v_{2,m}, \ldots, v_{n,1}, \ldots, v_{n,m}), \end{aligned}$$

que associa uma tupla de vetores em $(\mathbb{F}_q^m)^n$ a um vetor em $\mathbb{F}_q^{mn}$.

Uma maneira mais explícita de expressar o mapeamento (5.2) se dá como segue. Para cada $i \in \{1, \ldots, n\}$, seja $\{\mathbf{b}_{i,1}, \ldots, \mathbf{b}_{i,m}\}$ um conjunto de vetores de $\mathbb{F}_q^m$ que geram o subespaço V_i, isto é, $V_i = \langle \mathbf{b}_{i,1}, \ldots, \mathbf{b}_{i,m} \rangle$. O mapeamento f é, então, dado por

$$\begin{aligned} f(\mathbf{V}) \quad = \quad & \langle (\mathbf{b}_{1,1}\,\mathbf{0}\,\cdots\,\mathbf{0}), (\mathbf{b}_{1,2}\,\mathbf{0}\,\cdots\,\mathbf{0}), \ldots, (\mathbf{b}_{1,m}\,\mathbf{0}\,\cdots\,\mathbf{0}), \\ & (\mathbf{0}\,\mathbf{b}_{2,1}\,\cdots\,\mathbf{0}), (\mathbf{0}\,\mathbf{b}_{2,2}\,\cdots\,\mathbf{0}), \ldots, (\mathbf{0}\,\mathbf{b}_{2,m}\,\cdots\,\mathbf{0}), \\ & \qquad\qquad \vdots \\ & (\mathbf{0}\,\cdots\,\mathbf{0}\,\mathbf{b}_{n,1}), (\mathbf{0}\,\cdots\,\mathbf{0}\,\mathbf{b}_{n,2}), \ldots, (\mathbf{0}\,\cdots\,\mathbf{0}\,\mathbf{b}_{n,m}) \rangle. \end{aligned} \tag{5.3}$$

Desse modo, a dimensão de $f(\mathbf{V}) \in \mathcal{P}(\mathbb{F}_q^{mn})$ é

$$\dim f(\mathbf{V}) = \sum_{i=1}^{n} \dim V_i. \tag{5.4}$$

Exemplo 5.1. Considere o espaço projetivo estendido $\mathcal{P}(\mathbb{F}_2^2)^2$ cujo diagrama está mostrado na Figura 5.2. Este exemplo ilustra como $\mathcal{P}(\mathbb{F}_2^2)^2$ pode ser visto como um subconjunto de $\mathcal{P}(\mathbb{F}_2^4)$. Seja $\mathbf{V} = (S_1, W)$. Então, pode-se ter como vetores geradores

$$\begin{aligned} \mathbf{b}_{1,1} &= 01, \qquad \mathbf{b}_{1,2} = 00, \\ \mathbf{b}_{2,1} &= 01, \qquad \mathbf{b}_{2,2} = 10. \end{aligned}$$

Assim, utilizando (5.3), o elemento $\mathbf{V}$ é mapeado em

$$\begin{aligned} f(\mathbf{V}) \quad &= \quad \langle 0100, 0000, 0001, 0010 \rangle \\ &= \quad \{0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111\}. \end{aligned}$$

Analogamente, $\mathbf{U} = (S_3, S_3)$ é mapeado em

$$f(\mathbf{U}) = \{0000, 0011, 1100, 1111\}.$$

Note, no entanto, que o subespaço $T \in \mathcal{P}(\mathbb{F}_2^4) = \{0000, 1111\}$ não pode ser expresso da forma apresentada em (5.3). □

A seguir, é mostrado que as distâncias são preservadas sob a aplicação de f. A injetividade de f segue imediatamente, visto que todo mapeamento que preserva distâncias é automaticamente injetivo.

Teorema 5.1. *Se* $\mathbf{V}, \mathbf{U} \in \mathcal{P}(\mathbb{F}_q^m)^n$ *então* $d_{\mathrm{S}}(\mathbf{V}, \mathbf{U}) = d_{\mathrm{S}}(f(\mathbf{V}), f(\mathbf{U}))$.

Demonstração. Sejam $\mathbf{V} = (V_1, \ldots, V_n)$ e $\mathbf{U} = (U_1, \ldots, U_n)$. É preciso mostrar que

$$\sum_{i=1}^{n} d_{\mathrm{S}}(V_i, U_i) = d_{\mathrm{S}}(V_1 \times \cdots \times V_n, U_1 \times \cdots \times U_n),$$

o que, de acordo com (4.1), equivale a

$$\begin{aligned}\sum_{i=1}^{n} \dim(V_i \dotplus U_i) - \sum_{i=1}^{n} \dim(V_i \cap U_i) \\ = \dim\left[(V_1 \times \cdots \times V_n) \dotplus (U_1 \times \cdots \times U_n)\right] \\ - \dim\left[(V_1 \times \cdots \times V_n) \cap (U_1 \times \cdots \times U_n)\right].\end{aligned}$$

Tem-se

$$\begin{aligned}&(V_1 \times \cdots \times V_n) \dotplus (U_1 \times \cdots \times U_n) \\ &\quad = \{(\mathbf{v}_1, \ldots, \mathbf{v}_n) + (\mathbf{u}_1, \ldots, \mathbf{u}_n) : \mathbf{v}_i \in V_i, \mathbf{u}_i \in U_i\} \\ &\quad = \{(\mathbf{v}_1 + \mathbf{u}_1, \ldots, \mathbf{v}_n + \mathbf{u}_n) : \mathbf{v}_i \in V_i, \mathbf{u}_i \in U_i\} \\ &\quad = \{(\mathbf{z}_1, \ldots, \mathbf{z}_n) : \mathbf{z}_i \in V_i \dotplus U_i\} \\ &\quad = (V_1 \dotplus U_1) \times \cdots \times (V_n \dotplus U_n),\end{aligned}$$

$$\begin{aligned}&(V_1 \times \cdots \times V_n) \cap (U_1 \times \cdots \times U_n) \\ &\quad = \{(\mathbf{z}_1, \ldots, \mathbf{z}_n) : \mathbf{z}_i \in V_i, \mathbf{z}_i \in U_i\} \\ &\quad = \{(\mathbf{z}_1, \ldots, \mathbf{z}_n) : \mathbf{z}_i \in V_i \cap U_i\} \\ &\quad = (V_1 \cap U_1) \times \cdots \times (V_n \cap U_n),\end{aligned}$$

e o resultado segue de (5.4). □

O Teorema 5.1 afirma que a cada código *n-shot* $\mathcal{C} \subseteq \mathcal{P}(\mathbb{F}_q^m)^n$ corresponde um código *one-shot* $f(\mathcal{C}) \subseteq \mathcal{P}(\mathbb{F}_q^{mn})$ de mesma distância mínima e mesma cardinalidade. Isso também sugere a construção de códigos *multishot* em $\mathcal{P}(\mathbb{F}_q^m)^n$ baseados em códigos *one-shot* em $\mathcal{P}(\mathbb{F}_q^{mn})$. Com efeito, se $\mathcal{C} \subseteq \mathcal{P}(\mathbb{F}_q^{mn})$ é um código com distância mínima d, basta eliminar as palavras-código que não estão em $f(\mathcal{P}(\mathbb{F}_q^m)^n)$ para se obter o código $\mathcal{C}' \subseteq \mathcal{P}(\mathbb{F}_q^{mn})$. Então, $f^{-1}(\mathcal{C}') \subseteq \mathcal{P}(\mathbb{F}_q^m)^n$ é um código *n-shot* com distância mínima de pelo menos d, mas com menor número de palavras-código. O código $\mathcal{C}_3 \subseteq \mathcal{P}(\mathbb{F}_2^2)^2$ da Seção 5.2 pode ser obtido do código $\{V \in \mathcal{P}(\mathbb{F}_2^4) : \dim V \in \{0, 2, 4\}\}$ dessa maneira.

5.4 Limitantes

Seja

$$A_q^n(m,d) = \max\{|\mathcal{C}| : \mathcal{C} \subseteq \mathcal{P}(\mathbb{F}_q^m)^n, d_{\mathrm{S}}(\mathcal{C}) = d\}$$

o tamanho do maior código *n-shot* sobre $\mathcal{P}(\mathbb{F}_q^m)$ com distância mínima igual a d. Esta seção busca obter limitantes inferiores e superiores sobre essa quantidade.

Evidentemente, todo limitante inferior para códigos de bloco clássicos $\left|\mathcal{P}(\mathbb{F}_q^m)\right|$-ários de comprimento n são limitantes inferiores para $A_q^n(m,d)$, fato esse que decorre da discussão na Seção 5.2. Além disso, todo limitante superior para códigos de subespaço *one-shot* em $\mathcal{P}(\mathbb{F}_q^{mn})$ é um limitante superior para $A_q^n(m,d)$, de acordo com a Seção 5.3. Assim,

$$\bar{A}_{\left|\mathcal{P}(\mathbb{F}_q^m)\right|}(n,d) \leq A_q^n(m,d) \leq A_q(mn,d),$$

em que $\bar{A}_{q'}(n,d)$ é o tamanho do maior código de bloco clássico de comprimento n sobre $\mathbb{Z}_{q'}$ com mínima distância de Hamming d e $A_q(m',d) = A_q^1(m',d)$, definido em 4.3, é o tamanho do maior código de subespaço *one-shot* em $\mathcal{P}(\mathbb{F}_q^{m'})$ com mínima distância de subespaço d.

5.4.1 Esferas no espaço projetivo estendido

Seguindo os mesmos passos efetuados no Capítulo 4, esta seção estuda esferas no espaço métrico $\mathcal{P}(\mathbb{F}_q^m)^n$ para que, na seção seguinte, sejam obtidos os limitantes esféricos. De acordo com o Apêndice B, tem-se as seguintes definições.

Definição. Uma **esfera** centrada em $\mathbf{V} = (V_1, \ldots, V_n) \in \mathcal{P}(\mathbb{F}_q^m)^n$ de raio r é definida por

$$\mathcal{B}_{(q,m,n)}(\mathbf{V}, r) = \{\mathbf{V}' \in \mathcal{P}(\mathbb{F}_q^m)^n : d_{\mathrm{S}}(\mathbf{V}', \mathbf{V}) \leq r\}$$

e o **volume** dessa esfera é definido por

$$\mathrm{Vol}_{(q,m,n)}(\mathbf{V}, r) = \left|\mathcal{B}_{(q,m,n)}(\mathbf{V}, r)\right|,$$

como usual. □

É obtido primeiro o volume da *casca* de centro $\mathbf{V}$ e raio j, definida como o conjunto de todos os pontos do espaço métrico que distam de $\mathbf{V}$ de exatamente j.

Lema 5.2. *O volume da casca de centro* $\mathbf{V} = (V_1, \ldots, V_n) \in \mathcal{P}(\mathbb{F}_q^m)^n$ *e raio* $j \in \{0, \ldots, mn\}$ *é dado por*

$$\mathrm{Vol}^{\mathrm{Casca}}_{(q,m,n)}(\mathbf{V}, j) = \sum_{\substack{\mathbf{j} \in \{0,\ldots,m\}^n : \\ j_1 + \cdots + j_n = j}} \prod_{i=1}^{n} \mathrm{Vol}^{\mathrm{Casca}}_{(q,m)}(k_i, j_i), \tag{5.5}$$

em que $\mathbf{k} = (k_1, \ldots, k_n) = (\dim V_1, \ldots, \dim V_n)$ *e* $\mathrm{Vol}^{\mathrm{Casca}}_{(q,m)}(\cdot, \cdot)$ *é dado por* (4.4).

A demonstração é omitida.

Exemplo 5.2. Seja $q = 2$, $m = 4$, $n = 3$. Deseja-se o volume da casca centrada em um dado $\mathbf{V}$ e de raio $j = 6$. Para tanto, deve-se considerar as n-tuplas $\mathbf{j} \in \{0, 1, 2, 3, 4\}^3$ cujas somas dos elementos sejam iguais a $j = 6$. Tais tuplas são

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

e suas respectivas permutações (19 ao todo). □

O seguinte teorema segue imediatamente do Lema 5.2.

Teorema 5.3. *O volume da esfera de centro* $\mathbf{V} \in \mathcal{P}(\mathbb{F}_q^m)^n$ *raio* $r \in \{0, \ldots, mn\}$ *é dado por*

$$\mathrm{Vol}_{(q,m,n)}(\mathbf{V}, r) = \sum_{j=0}^{r} \mathrm{Vol}^{\mathrm{Casca}}_{(q,m,n)}(\mathbf{V}, j),$$

em que $\mathrm{Vol}^{\mathrm{Casca}}_{(q,m,n)}(\cdot, \cdot)$ *é dado por* (5.5).

Analogamente ao caso *one-shot*, o volume de uma esfera centrada em $\mathbf{V} = (V_1, \ldots, V_n)$ depende apenas de $\mathbf{k} = (\dim V_1, \ldots, \dim V_n)$, de modo que a notação $\mathrm{Vol}_{(q,m,n)}(\mathbf{k}, r)$ é utilizada.

Dada uma tupla $\mathbf{k} = (k_1, \ldots, k_n)$, há um total de

$$\mathrm{Freq}_{(q,m,n)}(\mathbf{k}) = \binom{m}{k_1}_q \cdots \binom{m}{k_n}_q$$

pontos $\mathbf{V} \in \mathcal{P}(\mathbb{F}_q^m)^n$ tais que $\mathbf{k} = (\dim V_1, \ldots, \dim V_n)$. Portanto, o volume médio de uma esfera de raio r em $\mathcal{P}(\mathbb{F}_q^m)^n$ é

$$\begin{aligned} \mathrm{Vol}^{\mathrm{med}}_{(q,m,n)}(r) &= \frac{1}{\left|\mathcal{P}(\mathbb{F}_q^m)^n\right|} \sum_{\mathbf{V}\in\mathcal{P}(\mathbb{F}_q^m)^n} \mathrm{Vol}_{(q,m,n)}(\mathbf{V}, r) \qquad (5.6) \\ &= \frac{1}{\left|\mathcal{P}(\mathbb{F}_q^m)\right|^n} \sum_{\mathbf{k}\in\{0,\ldots,m\}^n} \mathrm{Freq}_{(q,m,n)}(\mathbf{k}) \mathrm{Vol}_{(q,m,n)}(\mathbf{k}, r). \end{aligned}$$

Em adição, pode-se mostrar que os volumes mínimo e máximo são

$$\begin{aligned} \mathrm{Vol}^{\mathrm{min}}_{(q,m,n)}(r) &= \mathrm{Vol}_{(q,m,n)}\left((\lfloor m/2 \rfloor, \ldots, \lfloor m/2 \rfloor), r\right), \qquad (5.7) \\ \mathrm{Vol}^{\mathrm{max}}_{(q,m,n)}(r) &= \mathrm{Vol}_{(q,m,n)}\left((0, \ldots, 0), r\right), \qquad (5.8) \end{aligned}$$

resultado análogo ao caso *one-shot*.

Por fim apresentam-se duas propriedades de $\mathrm{Vol}_{(q,m,n)}(\mathbf{k}, r)$ visto como uma função de $\mathbf{k}$:

$$\mathrm{Vol}_{(q,m,n)}\left((\ldots, k_i, \ldots), r\right) = \mathrm{Vol}_{(q,m,n)}\left((\ldots, m - k_i, \ldots), r\right),$$

$$\mathrm{Vol}_{(q,m,n)}\left(\mathbf{k}, r\right) = \mathrm{Vol}_{(q,m,n)}\left(\mathbf{k}', r\right),$$

em que, nessa última, $\mathbf{k}'$ é qualquer permutação de $\mathbf{k}$.

A ideia é adotar essas propriedades para simplificar o cálculo do volume em (5.6). Naquela equação, ao passar da primeira para a segunda igualdade, o número de parcelas da somatória foi reduzido de $\left|\mathcal{P}(\mathbb{F}_q^m)\right|^n$ (número de pontos $\mathbf{V}$ no espaço métrico) para $(m+1)^n$ (número de n-tuplas $\mathbf{k}$ com elementos em $\{0, \ldots, m\}$). As duas propriedades recém apresentadas permitem reduzir ainda mais o número de parcelas, como ilustram os próximos exemplos.

Exemplo 5.3. Sejam os parâmetros $q = 2$, $m = 2$, $n = 2$. A tabela a seguir mostra o volume de esferas em $\mathcal{P}(\mathbb{F}_q^m)^n$ em função do centro e do raio da esfera.

$\mathbf{k}$	Freq($\mathbf{k}$)	Centros da esfera	Vol($\mathbf{k}, r$)				
			0	1	2	3	4
$(0,0)$	1	OO	1	7	18	24	25
$(0,1)$	3	OV_1, OV_2, OV_3	1	6	15	23	25
$(0,2)$	1	OW	1	7	18	24	25
$(1,0)$	3	V_1O, V_2O, V_3O	1	6	15	23	25
$(1,1)$	9	$V_1V_1, V_1V_2, \ldots, V_3V_3$	1	5	13	21	25
$(1,2)$	3	V_1W, V_2W, V_3W	1	6	15	23	25
$(2,0)$	1	WO	1	7	18	24	25
$(2,1)$	3	WV_1, WV_2, WV_3	1	6	15	23	25
$(2,2)$	1	WW	1	7	18	24	25

Há um total de $\left|\mathcal{P}(\mathbb{F}_q^m)\right|^n = 5^2 = 25$ pontos no espaço métrico. Esse número é reduzido a $(m+1)^n = 3^2 = 9$, o número de pares ordenados com elementos em $\{0,1,2\}$. Usando as propriedades é possível reduzir esse número para 3. De fato, para $r = 1$, por exemplo, existem

(*i*) 9 pontos com volume 5 (mínimo),

(*ii*) 12 pontos com volume 6 e

(*iii*) 4 pontos com volume 7 (máximo).

O volume médio é, portanto, 5,8. □

Exemplo 5.4. A Figura 5.4 mostra o perfil de volumes em $\mathcal{P}(\mathbb{F}_q^m)^n$ para o caso $q = 2$, $m = 4$ e $n = 3$. O raio foi fixado em $r = 6$. Os valores de $\mathbf{k}$ estão em ordem lexicográfica. Apenas as tuplas $\mathbf{k}$ com $k_3 = 0$ são mostradas no eixo horizontal, por falta de espaço.

O volumes mínimo, médio e máximo de uma esfera com esses parâmetros são, respectivamente,

$$\begin{aligned} \mathrm{Vol}^{\min}_{(2,4,3)}(6) &= 49371, \\ \mathrm{Vol}^{\mathrm{med}}_{(2,4,3)}(6) &= 62232{,}85 \text{ e} \\ \mathrm{Vol}^{\max}_{(2,4,3)}(6) &= 196224. \end{aligned}$$

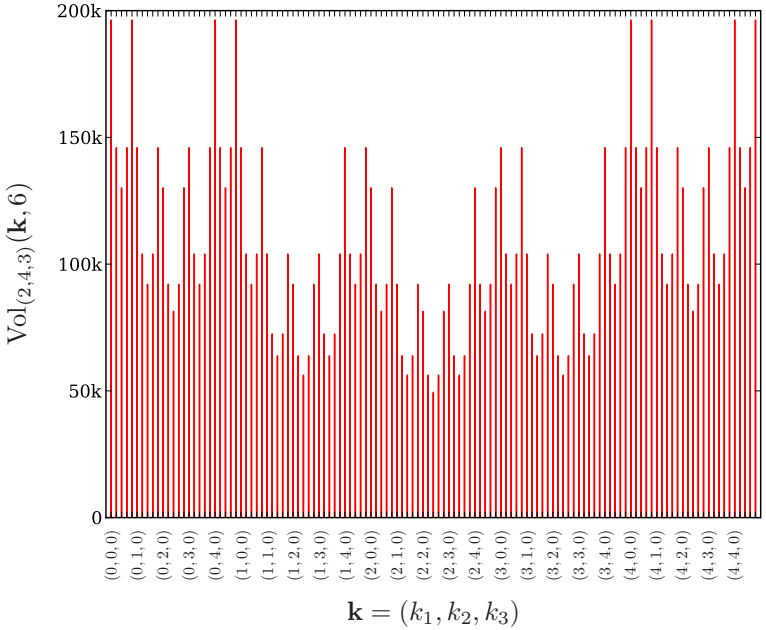


Figura 5.4: Perfil de volumes para $\mathcal{P}(\mathbb{F}_2^4)^3$ e $r = 6$.

A figura mostra que o volume é mínimo quando $\mathbf{k} = (2, 2, 2)$ e máximo quando $\mathbf{k} = (0, 0, 0)$, como esperado. Além disso, usando as propriedades, é possível reduzir o número de casos de 125 para 10. □

5.4.2 Limitantes de Hamming e Gilbert-Varshamov

Aplicando os resultados da Seção B.2 do Apêndice B, obtém-se o limitante inferior de Hamming e o limitante superior de Gilbert-Varshamov.

Teorema 5.4. *Os seguintes limitantes sobre o tamanho ótimo de códigos no espaço projetivo estendido são válidos:*

$$\frac{\left|\mathcal{P}(\mathbb{F}_q^m)\right|^n}{\mathrm{Vol}^{\mathrm{med}}_{(q,m,n)}(d-1)} \leq A_q^n(m,d) \leq \frac{\left|\mathcal{P}(\mathbb{F}_q^m)\right|^n}{\mathrm{Vol}^{\mathrm{min}}_{(q,m,n)}(\lfloor (d-1)/2 \rfloor)},$$

em que $\mathrm{Vol}^{\mathrm{med}}_{(q,m,n)}(\cdot)$ *é dado por* (5.6) *e* $\mathrm{Vol}^{\mathrm{min}}_{(q,m,n)}(\cdot)$ *é dado por* (5.7).

5.4.3 Puncionamento

Para o espaço projetivo estendido $\mathcal{P}(\mathbb{F}_q^m)^n$, duas operações de puncionamento são naturais. A primeira aplica o puncionamento definido na Subseção 4.4.3 do capítulo anterior a cada componente da tupla $\mathbf{V} = (V_1, \ldots, V_n) \in \mathcal{P}(\mathbb{F}_q^m)^n$, ou seja,

$$\begin{aligned}(\cdot)^{\blacktriangledown_1} : \mathcal{P}(\mathbb{F}_q^m)^n &\longrightarrow \mathcal{P}(\mathbb{F}_q^{m-1})^n \\ \mathbf{V} &\longmapsto (V_1^{\blacktriangledown}, \ldots, V_n^{\blacktriangledown}),\end{aligned}$$

em que $(\cdot)^{\blacktriangledown}$ ao lado direito do símbolo de mapeamento é dado por (4.10).

A segunda operação de puncionamento no espaço projetivo estendido consiste em remover um elemento (isto é, um subespaço) da tupla $\mathbf{V} = (V_1, \ldots V_n)$—por exemplo, o último elemento. Em outras palavras,

$$\begin{aligned}(\cdot)^{\blacktriangledown_2} : \mathcal{P}(\mathbb{F}_q^m)^n &\longrightarrow \mathcal{P}(\mathbb{F}_q^m)^{n-1} \\ \mathbf{V} &\longmapsto (V_1, \ldots, V_{n-1}).\end{aligned}$$

Naturalmente, se $\mathcal{C} \subseteq \mathcal{P}(\mathbb{F}_q^m)^n$ é um código de subespaço estendido, os correspondentes códigos puncionados $\mathcal{C}^{\blacktriangledown_1}$ ou $\mathcal{C}^{\blacktriangledown_2}$ são obtidos aplicando o respectivo puncionamento a cada palavra-código de $\mathcal{C}$.

O seguinte lema é útil para o limitante de Singleton que será obtido na seção seguinte.

Lema 5.5. *Seja um código $\mathcal{C} \subseteq \mathcal{P}(\mathbb{F}_q^m)^n$.*

(i) Se $d_{\mathrm{S}}(\mathcal{C}) > 2n$ então $\mathcal{C}^{\blacktriangledown_1} \subseteq \mathcal{P}(\mathbb{F}_q^{m-1})^n$ tem cardinalidade $|\mathcal{C}^{\blacktriangledown_1}| = |\mathcal{C}|$ e distância mínima $d_{\mathrm{S}}(\mathcal{C}^{\blacktriangledown_1}) \geq d_{\mathrm{S}}(\mathcal{C}) - 2n$.

(ii) Se $d_{\mathrm{S}}(\mathcal{C}) > m$ então $\mathcal{C}^{\blacktriangledown_2} \subseteq \mathcal{P}(\mathbb{F}_q^m)^{n-1}$ tem cardinalidade $|\mathcal{C}^{\blacktriangledown_2}| = |\mathcal{C}|$ e distância mínima $d_{\mathrm{S}}(\mathcal{C}^{\blacktriangledown_2}) \geq d_{\mathrm{S}}(\mathcal{C}) - m$.

Demonstração. Sejam $\mathbf{V}, \mathbf{U} \in \mathcal{P}(\mathbb{F}_q^m)^n$ distintos. Então,

$$\begin{aligned}d_{\mathrm{S}}(\mathbf{V}^{\blacktriangledown_1}, \mathbf{U}^{\blacktriangledown_1}) &= \sum_{i=1}^{n} d_{\mathrm{S}}(V_i^{\blacktriangledown}, U_i^{\blacktriangledown}) \\ &\geq \sum_{i=1}^{n} [d_{\mathrm{S}}(V_i, U_i) - 2] \\ &= d_{\mathrm{S}}(\mathbf{V}, \mathbf{U}) - 2n,\end{aligned}$$

em que a desigualdade segue da demonstração do Lema 4.6. Utilizando argumentos similares aos dessa mesma demonstração, a primeira afirmação segue. Além disso,

$$\begin{aligned} d_{\mathrm{S}}(\mathbf{V},\mathbf{U}) &= \sum_{i=1}^{n} d_{\mathrm{S}}(V_i,U_i) \\ &= \sum_{i=1}^{n-1} d_{\mathrm{S}}(V_i,U_i) + d_{\mathrm{S}}(V_n,U_n) \\ &\leq d_{\mathrm{S}}(\mathbf{V}^{\blacktriangledown_2},\mathbf{U}^{\blacktriangledown_2}) + m. \end{aligned}$$

e a segunda afirmação segue analogamente. □

5.4.4 Limitantes de Singleton

A seguir são apresentados dois limitantes superiores análogos ao limitante de Singleton.

Teorema 5.6. *Os seguintes limitantes sobre o tamanho ótimo de códigos no espaço projetivo estendido são válidos:*

$$A_q^n(m,d) \leq \left|\mathcal{P}(\mathbb{F}_q^{m-\left\lfloor\frac{d-1}{2n}\right\rfloor})\right|^n,$$

$$A_q^n(m,d) \leq \left|\mathcal{P}(\mathbb{F}_q^m)\right|^{n-\left\lfloor\frac{d-1}{m}\right\rfloor}.$$

Demonstração. Seja um código $\mathcal{C} \subseteq \mathcal{P}(\mathbb{F}_q^m)^n$ qualquer com distância mínima $d_{\mathrm{S}}(\mathcal{C}) = d$. Aplicando o puncionamento $(\cdot)^{\blacktriangledown_1}$ ao código $\mathcal{C}$ um total de $\lfloor (d-1)/(2n) \rfloor$ vezes, pelo Lema 5.5, obtém-se um código $\mathcal{C}' = \mathcal{C}^{\blacktriangledown_1 \cdots \blacktriangledown_1} \subseteq \mathcal{P}(\mathbb{F}_q^{m-\left\lfloor\frac{d-1}{2n}\right\rfloor})^n$ com cardinalidade $|\mathcal{C}'| = |\mathcal{C}|$ e distância mínima $d_{\mathrm{S}}(\mathcal{C}') \geq 1$. Similarmente, aplicando o puncionamento $(\cdot)^{\blacktriangledown_2}$ ao código $\mathcal{C}$ um total de $\lfloor (d-1)/m \rfloor$ vezes, pelo Lema 5.5, obtém-se um código $\mathcal{C}'' = \mathcal{C}^{\blacktriangledown_2 \cdots \blacktriangledown_2} \subseteq \mathcal{P}(\mathbb{F}_q^m)^{n-\left\lfloor\frac{d-1}{m}\right\rfloor}$ com cardinalidade $|\mathcal{C}''| = |\mathcal{C}|$ e distância mínima $d_{\mathrm{S}}(\mathcal{C}'') \geq 1$. □

Exemplo 5.5. A Figura 5.5 mostra os dois limitantes de Singleton obtidos para os parâmetros $q = 2$, $m = 40$ e $n = 5$.O eixo vertical está normalizado por $\log_q |\mathcal{P}(\mathbb{F}_q^m)^n|$. Nota-se que, dependendo da distância

mínima d desejada, um caso pode fornecer um limitante mais forte (menor) que o outro. □

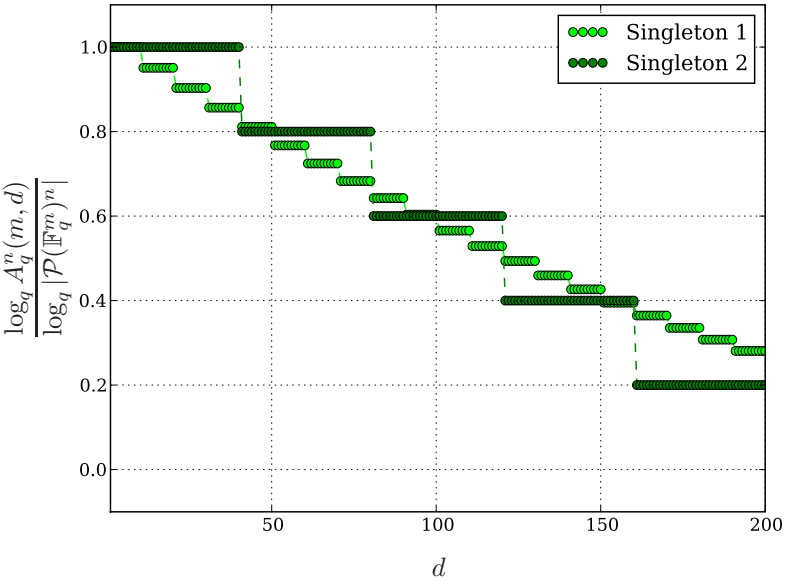


Figura 5.5: Limitantes de Singleton sobre $A_q^n(m,d)$ para $q=2$, $m=40$ e $n=5$.

5.4.5 Gráficos

Os limitantes recém deduzidos estão ilustrados na Figura 5.6 para o espaço projetivo estendido $\mathcal{P}(\mathbb{F}_q^m)^n$ com $q=2$, $m=5$ e $n=4$. O eixo vertical está normalizado por $\log_q \left|\mathcal{P}(\mathbb{F}_q^m)^n\right|$. O limitante de Singleton apresentado consiste no mínimo entre os dois limitantes obtidos anteriormente.

5.5 Construção multinível

Esta seção apresenta um método de construção de códigos de subespaço *multishot* baseado na chamada *modulação codificada de bloco*. Este

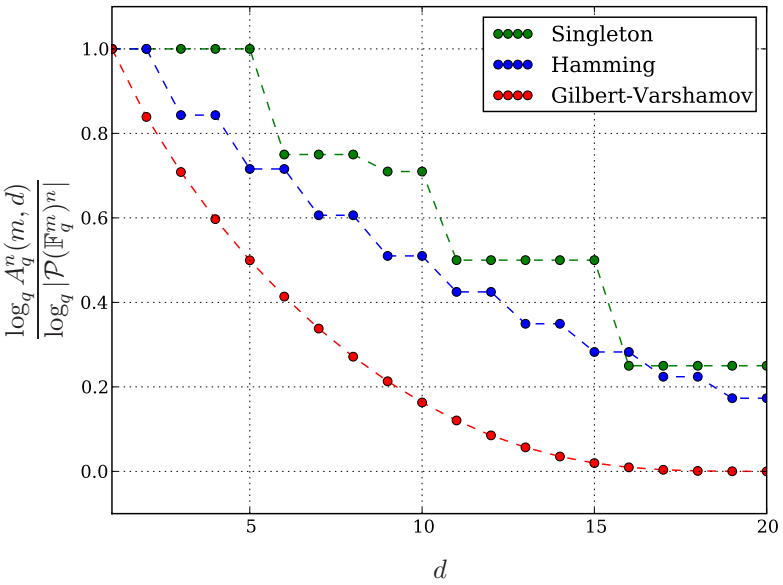


Figura 5.6: Limitantes sobre $A_q^n(m,d)$ para $q=2$ e $m=5$ e $n=4$.

esquema, proposto pela primeira vez por Imai e Hirakawa em [28] foi bastante popular nas décadas de 80 e 90 para o projeto de códigos para canais de comunicação contínuos (como o canal gaussiano). Aqui, aplica-se a teoria ao canal de subespaço. A abordagem descrita nesta dissertação segue o trabalho de Calderbank [4], no qual várias outras referências nesse assunto são listadas.

5.5.1 Particionamentos multiníveis

A seguir são descritos alguns conceitos fundamentais no procedimento de construção a ser proposto.

Particionamento multinível. Dado um conjunto $\mathcal{S}$, um *particionamento L-nível* de $\mathcal{S}$ é definido como uma sequência de $L+1$ partições[*] $\Gamma_0, \ldots, \Gamma_L$ tais que

[*] *cf.* Apêndice A.7.

(*i*) a partição do nível 0 consiste no conjunto $\mathcal{S}$ completo, isto é, $\Gamma_0 = \{\mathcal{S}\}$,

(*ii*) a partição Γ_l é um refinamento da partição Γ_{l-1}, para $l = 1, \ldots, L$ e

(*iii*) a partição do nível L consiste nos subconjuntos unitários de $\mathcal{S}$, isto é, $\Gamma_L = \{\{s\} : s \in \mathcal{S}\}$.

O particionamento em questão é também chamado de *particionamento multinível*.

Representação em árvore. Pode-se construir uma *árvore* representando um particionamento multinível da seguinte maneira. A árvore possui $L+1$ níveis, numerados de 0 a L. O nível l possui como nós os elementos da partição Γ_l. Assim, o nó *raiz* da árvore (nível 0) é o único elemento de Γ_0, isto é, o conjunto $\mathcal{S}$. Um nó $\mathcal{Y} \in \Gamma_l$ (nível l) é *filho* do único nó $\mathcal{X} \in \Gamma_{l-1}$ (nível $l-1$) tal que $\mathcal{Y} \subseteq \mathcal{X}$. Equivalentemente, um nó $\mathcal{Y} \in \Gamma_l$ (nível l) é *pai* de todo nó $\mathcal{Z} \in \Gamma_{l+1}$ (nível $l+1$) tal que $\mathcal{Z} \subseteq \mathcal{Y}$.

Aninhamento. Como será explicado posteriormente, a construção multinível de códigos de subespaço *multishot* exige "aninhamento" consecutivo até um certo nível. O nível $l \geq 1$ é dito estar *aninhado* se cada nó do nível $l-1$ possui o mesmo número p_l de nós filhos. (As cardinalidades dos elementos do nível l podem ser diferentes.)

Caminhos. Suponha que o nível l esteja aninhado, para cada $l \in \{1, \ldots, L'\}$, para algum $L' \geq 1$. Construída a árvore de um dado particionamento multinível, é possível rotular os ramos que partem de um dado nó com os inteiros do conjunto $\{0, \ldots, p_l - 1\}$, em que p_l é o número de filhos de um nó do nível $l-1$. Feito isto, cada elemento $\mathcal{X} \in \Gamma_{l'}$ do nível L' pode ser identificado por um caminho $\mathbf{a}(\mathcal{X}) = (a_1, \ldots, a_{L'})$ na árvore, em que $a_l \in \{0, \ldots, p_l - 1\}$.

Exemplo 5.6. Seja $\mathcal{S}$ o espaço projetivo $\mathcal{P}(\mathbb{F}_2^3)$ mostrado na Figura 4.1 do Capítulo 4. Considere o particionamento binível $\Gamma_0, \Gamma_1, \Gamma_2$ de $\mathcal{S}$ dado por

$$\begin{aligned}
\Gamma_0 &= \{\{O, S_1, S_2, \cdots, S_7, S_1^\perp, S_2^\perp, \cdots, S_7^\perp, W\}\}, \\
\Gamma_1 &= \{\{O, S_1^\perp, S_2^\perp, \cdots, S_7^\perp\}, \{S_1, S_2, \cdots, S_7, W\}\}, \\
\Gamma_2 &= \{\{O\}, \{S_1\}, \{S_2\}, \cdots, \{S_7\}, \{S_1^\perp\}, \{S_2^\perp\}, \cdots, \{S_7^\perp\}, \{W\}\}.
\end{aligned}$$

A representação em árvore desse particionamento está mostrada na Figura 5.7. O nível $l = 1$ está aninhado (e sempre estará, qualquer que

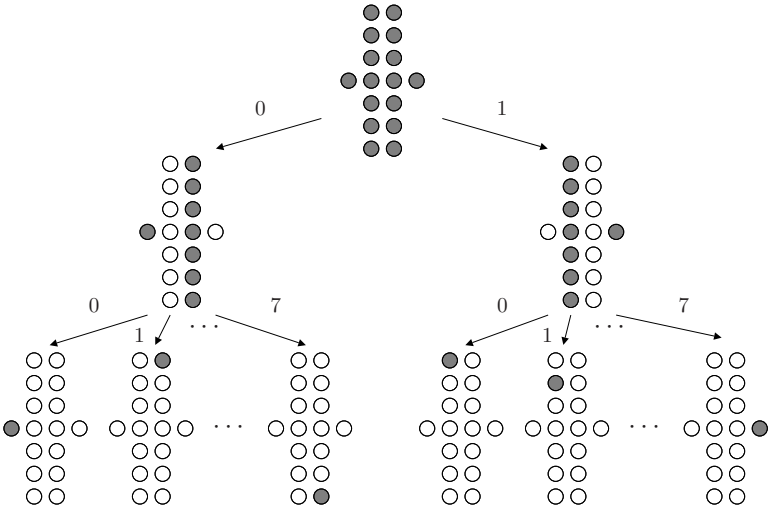


Figura 5.7: Representação em árvore de um particionamento multinível de $\mathcal{P}(\mathbb{F}_2^3)$.

seja o particionamento multinível, pois o nó raiz da árvore é sempre um só), com $p_1 = 2$. O nível $l = 2$ também está aninhado, visto que os dois nós do nível 1 têm ambos $p_2 = 8$ filhos cada. Por fim, os caminhos

$$\begin{aligned}
\mathbf{a}(\{O\}) &= (0,0),\\
\mathbf{a}(\{S_1^{\perp}\}) &= (0,1)\\
\mathbf{a}(\{S_2^{\perp}\}) &= (0,2),\\
&\vdots\\
\mathbf{a}(\{S_7^{\perp}\}) &= (0,7),\\
\mathbf{a}(\{S_1\}) &= (1,0),\\
\mathbf{a}(\{S_2\}) &= (1,1),\\
&\vdots\\
\mathbf{a}(\{S_7\}) &= (1,6),\\
\mathbf{a}(\{W\}) &= (1,7)
\end{aligned}$$

identificam os elementos de Γ_2, enquanto os caminhos

$$\begin{aligned}\mathbf{a}(\{O, S_1, S_2, \cdots, S_7\}) &= (0),\\ \mathbf{a}(\{S_1^\perp, S_2^\perp, \cdots, S_7^\perp, W\}) &= (1)\end{aligned}$$

identificam os elementos de Γ_1. □

5.5.2 Procedimento de construção

Descreve-se agora o procedimento de construção de códigos de subespaço *multishot* em $\mathcal{P}(\mathbb{F}_q^m)^n$, com q, m e n fornecidos. Tem-se como objetivo um código $\mathcal{C} \subseteq \mathcal{P}(\mathbb{F}_q^m)^n$ com uma dada distância mínima $d_{\mathrm{S}}(\mathcal{C}) = d$.

O procedimento tem como entrada um particionamento L-nível de um subconjunto $\mathcal{S} \subseteq \mathcal{P}(\mathbb{F}_q^m)$, possivelmente o próprio $\mathcal{P}(\mathbb{F}_q^m)$. Seja $\Gamma_0, \ldots, \Gamma_L$ tal particionamento.

Distâncias *intrasubset*. A *distância (de subespaço) intrasubset* do nível l é definida por

$$d_{\mathrm{S}}^{(l)} = \min\{d_{\mathrm{S}}(\mathcal{X}) : \mathcal{X} \in \Gamma_l\},$$

para $l = 0, \ldots, L$. Tem-se sempre $d_{\mathrm{S}}^{(0)} = d_{\mathrm{S}}(\mathcal{S})$ e $d_{\mathrm{S}}^{(L)} = \infty$.

Exemplo 5.7. Tem-se

$$d_{\mathrm{S}}^{(0)} = 1, \qquad d_{\mathrm{S}}^{(1)} = 2, \qquad d_{\mathrm{S}}^{(2)} = \infty$$

no particionamento binível do Exemplo 5.6. □

Códigos componentes. Seja L' o menor nível tal que $d_{\mathrm{S}}^{(l)} \geq d$ para todo l tal que $L' \leq l \leq L$. A construção multinível exige aninhamento de todo o nível l tal que $1 \leq l \leq L'$. Tais níveis devem ser "protegidos" por códigos clássicos de comprimento n sobre $\mathbb{Z}_{p^l}$, chamados de *códigos componentes* e denotados por $\mathcal{H}_l$, com distâncias (de Hamming) mínimas $d_{\mathrm{H}}^{(l)} = d_{\mathrm{H}}(\mathcal{H}_l)$. Segue da teoria multinível (*cf.* [4]) que a distância mínima do código de subespaço *multishot* projetado é limitada inferiormente por

$$d_{\mathrm{S}}(\mathcal{C}) \geq \min\{d_{\mathrm{S}}^{(l-1)} d_{\mathrm{H}}^{(l)} : 1 \leq l \leq L'\}. \tag{5.9}$$

Observe que talvez seja necessário o reajuste do valor de n, pois códigos componentes $\mathcal{H}_l$ satisfazendo (5.9) podem não existir.

Exemplo 5.8. Continuando o exemplo corrente, suponha que se deseje um código de subespaço *multishot* com distância mínima $d = 3$. Então, o menor nível L' tal que $d_{\mathrm{S}}^{(l)} \geq d$ para todo l tal que $L' \leq l \leq L$ é $L' = 2$. A exigência de aninhamento é satisfeita.

É necessário encontrar código componentes $\mathcal{H}_1$ binário (pois $p_1 = 2$) e $\mathcal{H}_2$ octal (pois $p_2 = 8$) tais que $d_{\mathrm{H}}(\mathcal{H}_1) \geq 3$ e $d_{\mathrm{H}}(\mathcal{H}_2) \geq \frac{3}{2}$, essa última exigência decorrendo de (5.9). O código de repetição

$$\mathcal{H}_1 = \{000, 111\}$$

juntamente com o código de paridade de 64 palavras-código sobre $\mathbb{Z}_8$

$$\mathcal{H}_2 = \{000, 017, 026, \ldots, 772\},$$

com $n = 3$, são suficientes. □

Determinação das palavras-código. Finalmente, a determinação das palavras-código do código *multishot* $\mathcal{C} \subseteq \mathcal{P}(\mathbb{F}_q^m)^n$ projetado se dá como segue.

(*i*) Formam-se todas as possíveis matrizes, cada uma com L' linhas e n colunas, contendo as possíveis palavras-código de $\mathcal{H}_l$ na l-ésima linha:

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \cdots \\ a_{L',1} & \cdots & a_{L',n} \end{bmatrix}.$$

O conjunto de todas essas matrizes é denotado por $\mathcal{A}$ e tem cardinalidade $|\mathcal{A}| = |\mathcal{H}_1| \cdots |\mathcal{H}_{L'}|$.

(*ii*) Cada coluna $\mathbf{A}[j] = (a_{1,j}, \ldots, a_{L',j})$ de uma dada matriz $\mathbf{A} \in \mathcal{A}$ representa um caminho na árvore. O elemento de $\Gamma_{L'}$ associado ao caminho $\mathbf{A}[j]$ é representado por $\Gamma_{L'}(\mathbf{A}[j])$.

(*iii*) Cada matriz $\mathbf{A} \in \mathcal{A}$ resulta em várias palavras-código. As possíveis palavras-código associadas à matriz $\mathbf{A}$ têm como j-ésima coordenada um subespaço de $\Gamma_{L'}(\mathbf{A}[j])$.

Tem-se, portanto, um total de

$$|\mathcal{C}| = \sum_{\mathbf{A}\in\mathcal{A}} \prod_{j=1}^{n} |\Gamma_{L'}(\mathbf{A}[j])|$$

palavras-código no código construído.

Exemplo 5.9. Continuando o Exemplo 5.8, as possíveis matrizes são

$$\begin{aligned}\mathcal{A} \;=\; \Bigg\{ &\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 7 \end{bmatrix}, \ldots, \begin{bmatrix} 0 & 0 & 0 \\ 7 & 7 & 2 \end{bmatrix}, \\ &\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 7 \end{bmatrix}, \ldots, \begin{bmatrix} 1 & 1 & 1 \\ 7 & 7 & 2 \end{bmatrix} \Bigg\}\end{aligned}$$

e são $|\mathcal{A}| = |\mathcal{H}_1| \cdot |\mathcal{H}_2| = 128$ ao todo. Nesse caso em particular, cada matriz de $\mathcal{A}$ dá origem a uma única palavra-código. Por exemplo, a matriz

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 7 \end{bmatrix}$$

dá origem à palavra-código

$$(S_1, S_2, W),$$

especificada pelos três caminhos, $(1,0)$, $(1,1)$ e $(1,7)$, determinados pelas colunas de $\mathbf{A}$ (*cf.* Exemplo 5.6). Assim, $|\mathcal{C}| = 128$. □

5.5.3 Mais exemplos

O exemplo a seguir mostra como o particionamento pode influenciar no projeto do código.

Exemplo 5.10. Considere o particionamento binível alternativo de $\mathcal{P}(\mathbb{F}_2^3)$ em que Γ_0 e Γ_2 são definidos da maneira obrigatória e

$$\begin{aligned}\Gamma_1 = \{\{O, W\}, \{S_1, S_1^\perp\}, \{S_2, S_2^\perp\}, \{S_4, S_4^\perp\}, \{S_7, S_7^\perp\}, \{S_3, S_5^\perp\}, \\ \{S_5, S_6^\perp\}, \{S_6, S_3^\perp\}\}.\end{aligned}$$

A Figura 4.1 do Capítulo 4 mostra que $d_{\mathrm{S}}(\mathcal{X}) = 3$ para todo $\mathcal{X} \in \Gamma_1$,

de modo que

$$d_{\mathrm{S}}^{(0)} = 1, \qquad d_{\mathrm{S}}^{(1)} = 3, \qquad d_{\mathrm{S}}^{(2)} = \infty.$$

A Figura 5.8 mostra o particionamento em questão. Assim, se se deseja

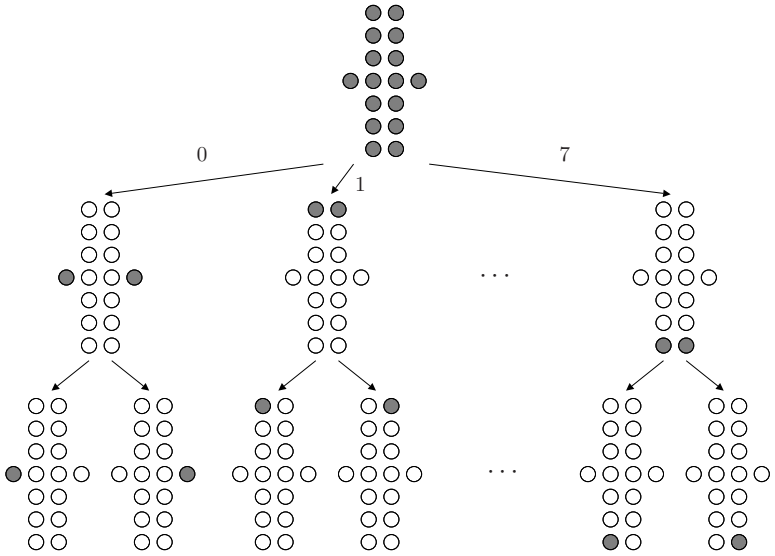


Figura 5.8: Representação em árvore de um particionamento alternativo de $\mathcal{P}(\mathbb{F}_2^3)$.

um código *multishot* com distância mínima $d = 3$, tem-se $L' = 1$. A exigência de aninhamento é trivialmente satisfeita, com $p_1 = 8$. É necessário, portanto, apenas um único código componente, $\mathcal{H}_1$, sendo esse octal e de distância mínima $d_{\mathrm{H}}(\mathcal{H}_1) \geq 3$. O código de repetição

$$\mathcal{H}_1 = \{000, 111, \ldots, 777\}$$

com comprimento $n = 3$ é suficiente.

As possíveis matrizes são oito:

$$\mathcal{A} = \left\{ \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}, \ldots, \begin{bmatrix} 7 & 7 & 7 \end{bmatrix} \right\}.$$

Nesse caso, cada matriz de $\mathcal{A}$ dá origem a 8 palavras-código. Por exemplo, a matriz

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$$

dá origem às palavras-código

$$\{O,W\} \times \{O,W\} \times \{O,W\} = \\ \{OOO, OOW, OWO, OWW, WOO, WOW, WWO, WWW\},$$

especificadas pelos três caminhos, (0), (0) e (0), determinados pelas colunas de $\mathbf{A}$. Assim, $|\mathcal{C}| = 8 \cdot 8 = 64$. □

O próximo exemplo ilustra como a escolha dos códigos componentes pode afetar o tamanho do código construído.

Exemplo 5.11. Este exemplo obtém o código $\mathcal{C}_3$ apresentado na Seção 5.2. A Figura 5.9 mostra um particionamento binível de $\mathcal{P}(\mathbb{F}_2^2)$.

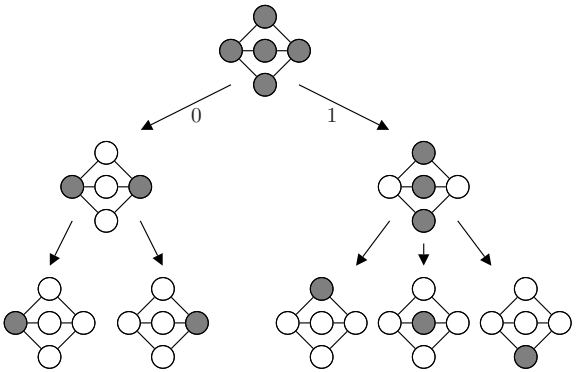


Figura 5.9: Representação em árvore de um particionamento de $\mathcal{P}(\mathbb{F}_2^2)$.

Para $d = 2$, tem-se $L' = 1$, de modo que apenas um código componente binário com distância mínima de Hamming igual a 2 é necessário. Se $n = 2$, os únicos códigos que satisfazem os requisitos são o código de repetição $\{00, 11\}$ e seu *coset* $\{01, 10\}$. O primeiro dá origem a

$$\begin{aligned} \mathcal{C}_3 &= \{O,W\} \times \{O,W\} \cup \\ &\quad \{S_1,S_2,S_3\} \times \{S_1,S_2,S_3\} \\ &= \{OO, OW, WO, WW, \\ &\quad S_1S_1, S_1S_2, \ldots, S_3S_2, S_3S_3\}, \end{aligned}$$

com $|\mathcal{C}_3| = 2 \cdot 2 + 3 \cdot 3 = 13$ enquanto o segundo dá origem a

$$\begin{aligned}\mathcal{C}_3' &= \{O, W\} \times \{S_1, S_2, S_3\} \cup \\ &\quad \{S_1, S_2, S_3\} \times \{O, W\} \\ &= \{OS_1, OS_2, OS_3, WS_1, WS_2, WS_3, \\ &\quad S_1O, S_2O, S_3O, S_1W, S_2W, S_3W\},\end{aligned}$$

com $|\mathcal{C}_3'| = 2 \cdot 3 + 3 \cdot 2 = 12$. □

5.5.4 Famílias de códigos

Finalizando esta seção apresentam-se duas famílias de códigos, construídas através de particionamentos multinível.

Códigos com distância mínima 2. Sejam q uma potência de primo, m um número ímpar e n um inteiro positivo qualquer. Considere o particionamento binível mostrado na Figura 5.10, na qual o nível $l = 2$ foi omitido por conveniência. O subconjunto da esquerda consiste nos su-

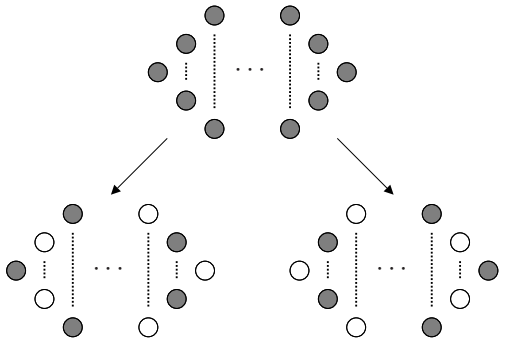


Figura 5.10: Particionamento binível de $\mathcal{P}(\mathbb{F}_q^m)$.

bespaços de $\mathcal{P}(\mathbb{F}_q^m)$ de dimensão par, enquanto o subconjunto da direita nos de dimensão ímpar. Defina também o código componente binário de comprimento n dado por

$$\mathcal{H}_1 = \{b_1 \cdots b_n \in \mathbb{Z}_2^n : b_1 + \cdots + b_n = 0\},$$

isto é, um código de paridade. Tem-se $|\mathcal{H}_1| = 2^{n-1}$ e $d_{\mathrm{H}}(\mathcal{H}_1) = 2$. O código *multishot* $\mathcal{C} \subseteq \mathcal{P}(\mathbb{F}_q^m)^n$ construído a partir desse particionamento

e desse código componente tem distância mínima

$$d_{\mathrm{S}}(\mathcal{C}) = 2$$

e taxa

$$R(\mathcal{C}) = \log_2 \left|\mathcal{P}(\mathbb{F}_q^m)\right| - \frac{1}{n},$$

em bits por uso do canal de subespaço. Essa construção, com alguns ajustes, também se aplica ao caso m par. O código do Exemplo 5.11 se encaixa nessa família.

Códigos com $n = m$ e distância mínima m. Sejam $q = 2$ e $n = m$ um número ímpar. Considere o particionamento binível de $\mathcal{P}(\mathbb{F}_q^m)$ mostrado na Figura 5.11, em que Γ_0 e Γ_2 são definidos da maneira obrigatória e a partição Γ_1 é constituída de $p_1 = \frac{1}{2}\left|\mathcal{P}(\mathbb{F}_q^m)\right|$ subconjuntos (cada

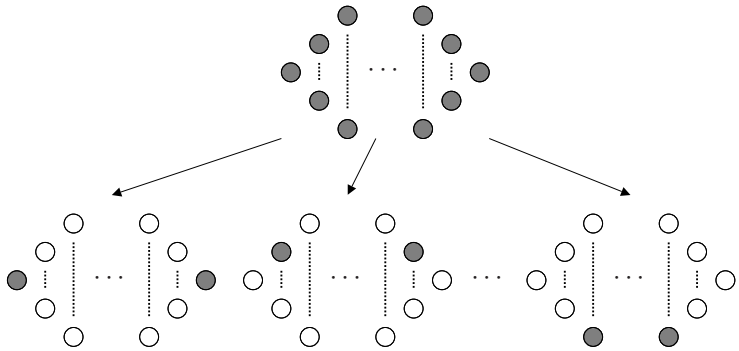


Figura 5.11: Particionamento binível alternativo de $\mathcal{P}(\mathbb{F}_q^m)$.

um com 2 elementos) com a seguinte propriedade: se $\{V, U\} \in \Gamma_1$ então $d_{\mathrm{S}}(V, U) = m$. A existência desse particionamento é garantida por [12, Theorem 13]. Defina também o código componente p_1-ário como sendo o código de repetição, de modo que $|\mathcal{H}_1| = p_1$ e $d_{\mathrm{H}}(\mathcal{H}_1) = m$. O código *multishot* $\mathcal{C} \subseteq \mathcal{P}(\mathbb{F}_q^m)^n$ construído a partir desse particionamento e desse código componente tem distância mínima

$$d_{\mathrm{S}}(\mathcal{C}) = m$$

e taxa

$$R(\mathcal{C}) = \left(1 - \frac{1}{m}\right) + \frac{1}{m} \log_2 \left|\mathcal{P}(\mathbb{F}_2^m)\right|,$$

em bits por uso do canal de subespaço. Nesse caso, uma construção análoga para m par não existe (*cf.* [12, Theorem 13]). O código do Exemplo 5.10 se encaixa nessa família.

5.6 Aplicação em codificação de rede

Códigos de subespaço *multishot* dão origem a códigos matriciais *multishot* que podem ser utilizados no controle de erros em codificação de rede não-coerente. Seja $\mathcal{C} \in \mathcal{P}(\mathbb{F}_q^m)^n$ um código de subespaço *multishot*. O código $\mathcal{C}$ pode ser visto como a composição de n subcódigos *one-shot* dados por

$$\mathcal{C}_i = \{V_i : \mathbf{V} = (V_1, \ldots, V_n) \in \mathcal{C}\},$$

de modo que $\mathcal{C} = \bar{\mathcal{C}}_1 \times \cdots \times \bar{\mathcal{C}}_n$. Seja $\bar{\mathcal{X}}_i$ um código matricial originário de $\bar{\mathcal{C}}_i$, para $i = 1, \ldots, n$. Então, o código matricial *multishot* formado por $\mathcal{X} = \bar{\mathcal{X}}_1 \times \cdots \times \bar{\mathcal{X}}_n$ tem taxa

$$R(\mathcal{X}) = \frac{\log_q |\mathcal{C}|}{m \cdot \sum_{i=1}^{n} \ell(\bar{\mathcal{C}}_i)},$$

em que $\ell(\bar{\mathcal{C}}_i)$ é a dimensão máxima do i-ésimo subcódigo de subespaço.

Exemplo 5.12. Considere o canal de subespaço sobre $\mathcal{P}(\mathbb{F}_2^2)$, mostrado na Figura 5.1. Esse exemplo ilustra como a presença de subespaços de dimensão elevada pode prejudicar a taxa do código matricial obtido. Para tanto, são comparados os desempenhos do código *one-shot*

$$\mathcal{C}_1 = \{S_1, S_2, S_3\},$$

e do código 3-*shot*

$$\begin{aligned} \mathcal{C}_2 \quad = \quad & \{O, W\} \times \{O, W\} \times \{S_1, S_2, S_3\} \cup \\ & \{O, W\} \times \{S_1, S_2, S_3\} \times \{O, W\} \cup \\ & \{S_1, S_2, S_3\} \times \{O, W\} \times \{O, W\} \cup \\ & \{S_1, S_2, S_3\} \times \{S_1, S_2, S_3\} \times \{S_1, S_2, S_3\}, \end{aligned}$$

com $|\mathcal{C}_2| = 2 \cdot 2 \cdot 3 + 2 \cdot 3 \cdot 2 + 3 \cdot 2 \cdot 2 + 3 \cdot 3 \cdot 3 = 63$, obtido da construção multinível com o particionamento binível da Figura 5.9 e com o código

componente $\mathcal{H}_1 = \{001, 010, 100, 111\}$. Ambos os códigos têm distância mínima 2. O código *one-shot* $\mathcal{C}_1$ dá origem a um código matricial $\mathcal{X}_1$ de taxa

$$R(\mathcal{X}_1) = \frac{\log_q |\mathcal{C}_1|}{m \cdot \ell(\mathcal{C}_1)} = \frac{\log_2 3}{2 \cdot 1} = 0{,}792,$$

enquanto o código *multishot* $\mathcal{C}_2$ dá origem a um código matricial $\mathcal{X}_2$ de taxa

$$R(\mathcal{X}_2) = \frac{\log_q |\mathcal{C}_2|}{m \cdot \sum_{i=1}^{n} \ell(\bar{\mathcal{C}}_{2,i})} = \frac{\log_2 63}{2 \cdot (2+2+2)} = 0{,}498.$$

Portanto, sob o contexto do controle de erros em codificação de rede, o código $\mathcal{C}_2$ tem um desempenho inferior ao de $\mathcal{C}_1$, em termos de taxa de informação. Isso se dá devido à presença de subespaços de dimensão 2 no código $\mathcal{C}_2$ (note que o código $\mathcal{C}_1$ possui apenas subespaços de dimensão 1), o que aumenta o preço da transmissão das palavras-código. O problema pode ser contornado eliminando-se de $\mathcal{C}_2$ todas as palavras-código que contêm o subespaço W. Fazendo isso, obtém-se o código

$$\begin{aligned} \mathcal{C}_2' \quad = \quad & \{O\} \times \{O\} \times \{S_1, S_2, S_3\} \cup \\ & \{O\} \times \{S_1, S_2, S_3\} \times \{O\} \cup \\ & \{S_1, S_2, S_3\} \times \{O\} \times \{O\} \cup \\ & \{S_1, S_2, S_3\} \times \{S_1, S_2, S_3\} \times \{S_1, S_2, S_3\}, \end{aligned}$$

com $|\mathcal{C}_2'| = 3 + 3 + 3 + 3 \cdot 3 \cdot 3 = 36$. Tal código dá origem a um código matricial $\mathcal{X}_2'$ de taxa

$$R(\mathcal{X}_2') = \frac{\log_q |\mathcal{C}_2'|}{m \cdot \sum_{i=1}^{n} \ell(\bar{\mathcal{C}}_{2,i}')} = \frac{\log_2 36}{2 \cdot (1+1+1)} = 0{,}862.$$

Portanto, a perda de 27 palavras-código foi compensada pela diminuição da dimensão máxima dos subespaços de 2 para 1. Além disso, uma vez que a taxa desse novo código é superior àquela do código *one-shot*, o uso de codificação de subespaço *multishot* é justificado no controle de erros em codificação de rede não-coerente. □

Outro ponto importante a ser considerado na aplicação de códigos de subespaço *multishot* no controle de erros em codificação de rede não-coerente é a relação entre a distância mínima do código e o êxito do código matricial *multishot* associado quando utilizado no canal matricial definido pelos parâmetros ρ (máxima deficiência de posto) e τ (máximo número de vetores de erro). Uma generalização do Teorema 4.9 (do caso *one-shot* para o caso *multishot*) ainda é um problema em aberto.

CAPÍTULO 6

Conclusão

O presente trabalho se dividiu em duas partes. A primeira apresentou uma introdução à teoria da *codificação de rede multidifusão*, cenário no qual toda a dissertação é focada. Foram apresentados resultados pertinentes que serviram de base para o restante do trabalho. A segunda parte abordou a área de *codificação no espaço projetivo*. Apesar de ter emergido como uma aplicação no controle de erros em codificação de rede, codificação no espaço projetivo pode ser vista como um campo de estudo à parte.

A ponte entre esses dois assuntos foi o *canal de comunicação matricial*, um modelo unificado para o combate das adversidades em codificação de rede multidifusão. Códigos para o canal matricial podem ser usados para contornar problemas relevantes presentes em redes de comunicação: desconhecimento da topologia, imperfeição do código de rede e erros nos canais. Códigos no espaço projetivo, isto é, *códigos de subespaço*, podem ser transformados em códigos matriciais, com aplicação específica sobre codificação de rede não-coerente.

A contribuição deste trabalho foi o estudo de códigos que utilizam o canal de subespaço múltiplas vezes; nesse contexto, foram definidos códigos de bloco sobre o espaço projetivo, aqui denominados de *códigos de subespaço multishot*. Codificação de subespaço *multishot* é uma nova área que reúne a teoria da codificação clássica com a teoria da

codificação de subespaço. Mais especificamente, os seguintes tópicos foram abordados:

(*i*) Definições da distância de subespaço estendida e de códigos de subespaço *multishot*.

(*ii*) Motivação para o uso de códigos *multishot*.

(*iii*) Relação dos códigos de subespaço *multishot* com os códigos de subespaço *one-shot*.

(*iv*) Obtenção de limitantes análogos aos de Hamming, Singleton e Gilbert-Varshamov para o caso em questão.

(*v*) Sugestão de um método para a construção de códigos *multishot*, incluindo exemplos de códigos e famílias de códigos.

(*vi*) Aplicação dos códigos de subespaço *multishot* no controle de erros em codificação de rede não-coerente.

Trabalhos futuros

A seguir são sugeridas futuras direções de pesquisa e indicadas algumas lacunas presentes neste trabalho.

(*i*) O estudo de códigos de subespaço *multishot* utilizando as *métricas de injeção* e *de posto*. A primeira métrica é mais adequada ao problema de controle de erros em codificação de rede, como mostra o artigo [47]. Já a segunda métrica citada, de acordo com [45], permite que várias ferramentas já abordadas na literatura sejam utilizadas no problema em consideração.

(*ii*) O uso de *códigos convolucionais* no lugar dos códigos de bloco para o caso de múltiplos *shots*, isto é, a adoção da modulação codificada convolucional de Ungerboeck (*cf.* [52]).

(*iii*) A *comparação* dos limitantes obtidos para o caso *multishot*, em $\mathcal{P}(\mathbb{F}_q^m)^n$, com os limitantes já conhecidos na literatura para o caso *one-shot*, em $\mathcal{P}(\mathbb{F}_q^{mn})$. Em adição, a comparação da *complexidade computacional* requerida no projeto, codificação e decodificação de códigos *multishot* em $\mathcal{P}(\mathbb{F}_q^m)^n$ e códigos *one-shot* em $\mathcal{P}(\mathbb{F}_q^{mn})$.

(*iv*) A determinação do *comportamento assintótico* dos limitantes obtidos para os códigos de subespaço *multishot* em $\mathcal{P}(\mathbb{F}_q^m)^n$, em termos do número de usos do canal, n, do tamanho do corpo finito, q, e do comprimento do vetor, m.

(*v*) A realização de *simulações computacionais* de sistemas que utilizam codificação de subespaço (*one-shot* e *multishot*) comparando as técnicas entre si e os resultados simulados com os teóricos. Em particular, a verificação da capacidade de correção de erros e a obtenção da taxa de erro alcançada em função da severidade da perturbação causada pelo adversário.

APÊNDICE A

Conceitos matemáticos

Este apêndice apresenta conceitos matemáticos diversos, necessários no decorrer da dissertação.

A.1 Grafos

Um *grafo simples* é definido por um par $G = (\mathcal{V}, \mathcal{E})$ no qual $\mathcal{V}$ é o conjunto dos *vértices* e $\mathcal{E}$ é o conjunto das *arestas*. As arestas são pares ordenados de vértices, no caso de um grafo *direcionado* ou pares não-ordenados de vértices, no caso de um grafo *não-direcionado*.

No caso de um *grafo composto* (direcionado ou não), $\mathcal{E}$ é na verdade um "multiconjunto", isto é, um conjunto no qual a multiplicidade (mas não a ordem) dos elementos é relevante. Intuitivamente, grafos compostos são tais que permitem "arestas paralelas".

Notação. Sejam $G = (\mathcal{V}, \mathcal{E})$ um grafo direcionado e $e \in \mathcal{E}$ uma aresta desse grafo dada por $e = (v_1, v_2)$ em que $v_1, v_2 \in \mathcal{V}$. Definem-se

$$\begin{aligned} \text{head}(e) &= v_2, \\ \text{tail}(e) &= v_1. \end{aligned}$$

Além disso, para $v \in \mathcal{V}$, definem-se

$$\begin{aligned} \mathrm{In}(v) &= \{e \in \mathcal{E} : \mathrm{head}(e) = v\}, \\ \mathrm{Out}(v) &= \{e \in \mathcal{E} : \mathrm{tail}(e) = v\}, \end{aligned}$$

ou seja, $\mathrm{In}(v)$ (resp., $\mathrm{Out}(v)$) é o conjunto de todas as arestas que entram (resp., saem) do nó v.

Caminhos e ciclos. Dados dois vértices $v_1, v_2 \in \mathcal{V}$, um *caminho direcionado* de v_1 a v_2 é uma sequência $(e_1, \ldots, e_n)$ de arestas tal que $v_1 = \mathrm{tail}(e_1)$, $\mathrm{head}(e_1) = \mathrm{tail}(e_2)$, ..., $\mathrm{head}(e_{n-1}) = \mathrm{tail}(e_n)$ e $\mathrm{head}(e_n) = v_2$. O *comprimento* do caminho é n. Um *ciclo direcionado* é um caminho direcionado tal que $v_1 = v_2$. Um ciclo de comprimento 1 é denominado de *loop*.

Grafos acíclicos. Grafos que não possuem nenhum ciclo são chamados de *grafos acíclicos*. A aciclicidade de um grafo permite que seus vértices sejam ordenados de acordo com um ordenamento parcial (*cf.* Apêndice A.2) chamado de *ordem ancestral*. Essa tem a seguinte propriedade: $v_1 \in \mathcal{V}$ precede $v_2 \in \mathcal{V}$ se e somente se existe um caminho direcionado de v_1 a v_2.

Problemas de fluxo máximo e corte mínimo. Tanto o *problema do fluxo máximo* quanto o *problema do corte mínimo* têm como entrada

(*i*) um grafo direcionado $G = (\mathcal{V}, \mathcal{E})$;

(*ii*) uma função $u : \mathcal{E} \longrightarrow \mathbb{R}$, com a interpretação de que $u(e)$ é a *capacidade* da aresta $e \in \mathcal{E}$;

(*iii*) dois nós quaisquer $s, t \in \mathcal{V}$, denominados, respectivamente, de *nó fonte* e *nó dreno*.

Fluxos. Um *fluxo* de s a t é uma atribuição $f : \mathcal{E} \longrightarrow \mathbb{R}$ tal que

$$\forall e \in \mathcal{E} : \quad 0 \leq f(e) \leq u(e),$$

$$\forall v \in \mathcal{V} - \{s, t\} : \quad \sum_{e \in \mathrm{In}(v)} f(e) = \sum_{e \in \mathrm{Out}(v)} f(e).$$

O *valor* do fluxo f é definido por

$$\mathrm{val}(f) = \sum_{e \in \mathrm{Out}(s)} f(e) = \sum_{e \in \mathrm{In}(t)} f(e).$$

O máximo valor de um fluxo entre s e t é denotado por $\mathrm{maxflow}(s,t)$:

$$\mathrm{maxflow}(s,t) = \max\{\mathrm{val}(f) : f \text{ é um fluxo}\}.$$

Cortes. Um *corte* entre s e t é um subconjunto $\mathcal{C} \subseteq \mathcal{V}$ tal que $s \in \mathcal{C}$ e $t \in \bar{\mathcal{C}}$, em que $\bar{\mathcal{C}} = \mathcal{V} - \mathcal{C}$ é o subconjunto de vértices complementar a $\mathcal{C}$. Dada qualquer aresta $e \in \mathcal{E}$ e um corte $\mathcal{C} \subseteq \mathcal{V}$, uma e somente uma das seguintes possibilidades ocorre.

(*i*) $\mathrm{tail}(e) \in \mathcal{C}$ e $\mathrm{head}(e) \in \mathcal{C}$;

(*ii*) $\mathrm{tail}(e) \in \mathcal{C}$ e $\mathrm{head}(e) \in \bar{\mathcal{C}}$; nesse caso, a aresta e é dita estar *saindo* do corte e escreve-se $e \in \Gamma^+_{\mathcal{C}}$;

(*iii*) $\mathrm{tail}(e) \in \bar{\mathcal{C}}$ e $\mathrm{head}(e) \in \mathcal{C}$; nesse caso, a aresta e é dita estar *entrando* no corte e escreve-se $e \in \Gamma^-_{\mathcal{C}}$;

(*iv*) $\mathrm{tail}(e) \in \bar{\mathcal{C}}$ e $\mathrm{head}(e) \in \bar{\mathcal{C}}$.

O *valor de um corte* $\mathcal{C}$ é definido pela soma das capacidades das arestas *saindo* do corte:

$$\mathrm{val}(\mathcal{C}) = \sum_{e \in \Gamma^+_{\mathcal{C}}} u(e).$$

O mínimo valor de um corte entre s e t é denotado por $\mathrm{mincut}(s,t)$:

$$\mathrm{mincut}(s,t) = \min\{\mathrm{val}(\mathcal{C}) : \mathcal{C} \text{ é um corte}\}.$$

Teorema maxflow-mincut. O *Teorema maxflow-mincut* afirma que

$$\mathrm{maxflow}(s,t) = \mathrm{mincut}(s,t).$$

Esse resultado foi provado pela primeira vez por Elias, Feinstein & Shannon em [8] e independentemente por Ford Jr. & Fulkerson em [13]. Vários métodos existem para se encontrar uma atribuição f tal que $\mathrm{val}(f) = \mathrm{maxflow}(s,t) = \mathrm{mincut}(s,t)$, sendo o mais famoso deles o *algoritmo de Ford-Fulkerson* [14].

A.2 Relações de ordem

Relações de ordem parciais e totais. Seja $\mathcal{X}$ um conjunto e $\preceq$ uma relação binária em $\mathcal{X}$. Se, para todo $x, y, z \in \mathcal{X}$, a relação $\preceq$ satisfizer

(i) $x \preceq x$ (reflexividade),

(ii) se $x \preceq y$ e $y \preceq x$ então $x = y$ (anti-simetria),

(iii) se $x \preceq y$ e $y \preceq z$ então $x \preceq z$ (transitividade),

então $\preceq$ é dita ser uma *relação de ordem* e o conjunto $\mathcal{X}$ é dito ser *ordenado* por $\preceq$. Se, em adição, a relação $\preceq$ satisfizer

(iv) $x \preceq y$ ou $y \preceq x$ (totalidade)

para todo $x, y \in \mathcal{X}$, então $\preceq$ é dita ser uma *relação de ordem total*. Caso contrário, tem-se uma *relação de ordem parcial*.

Elementos comparáveis. Se nem $x \preceq y$ nem $y \preceq x$ então x e y são ditos *incomparáveis*. Uma relação de ordem total é, portanto, uma relação de ordem na qual todos os pares de elementos são comparáveis.

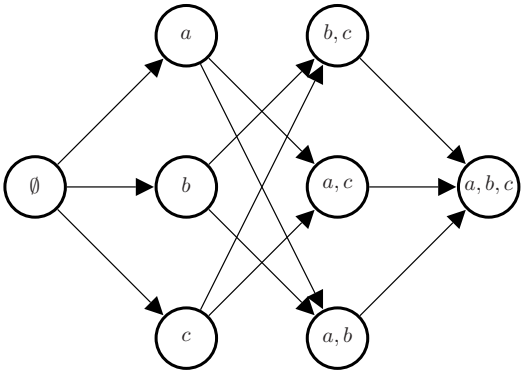


Figura A.1: Diagrama de Hasse de $2^{\{a,b,c\}}$ ordenado por inclusão.

Diagrama de Hasse. Todo conjunto finito $\mathcal{X}$ ordenado por $\preceq$ pode ser representado graficamente através de um grafo direcionado chamado de *diagrama de Hasse*. O conjunto dos vértices é o próprio conjunto $\mathcal{X}$

e (x_1, x_2) é uma aresta (direcionada) se e somente se $x_1 \preceq x_2$ e não existir $x \in \mathcal{X}$ tal que $x_1 \preceq x \preceq x_2$. O diagrama de Hasse de qualquer relação de ordem total finita consiste em um único caminho.

Exemplo. Seja $\mathcal{X} = 2^{\{a,b,c\}}$ o conjunto de todos os subconjuntos de $\{a, b, c\}$. O conjunto $\mathcal{X}$ é parcialmente ordenado por $\preceq$ em que $x \preceq y$ se e somente se $x \subseteq y$. O diagrama de Hasse de $\mathcal{X}$ ordenado por $\preceq$ é mostrado na Figura (A.1).

A.3 Corpos finitos

Definição de um corpo. Um *corpo* é uma estrutura algébrica na qual as quatro operações algébricas (adição, subtração, multiplicação e divisão) estão bem definidas e respeitam determinadas propriedades (associatividade da adição e multiplicação; comutatividade da adição e multiplicação; existência de elementos neutros; existência de elementos opostos e inversos; e distributividade da multiplicação sobre a soma). Exemplos familiares de corpos incluem os números racionais, reais e complexos. Os números naturais não constituem um corpo (por falta do elemento oposto, por exemplo), tampouco os inteiros (por falta do elemento inverso).

Corpos finitos Um *corpo finito* é um corpo que possui um número finito de elementos. São também chamados de *corpos de Galois*. Dois resultados são importantes:

(*i*) Se dois corpos finitos possuem o mesmo número de elementos então eles são isomórficos. Assim, denota-se "o" corpo finito com q elementos por $\mathbb{F}_q$.

(*ii*) Se $\mathbb{F}_q$ é um corpo finito então $q = p^k$ para algum p primo e k inteiro positivo.

A.4 Polinômios multivariáveis

Definições. Um *polinômio (multivariável)* nas variáveis $X_1, \ldots, X_\eta$ sobre um anel comutativo[*] R é definido por coeficientes $c_{\mathbf{i}} \in R$, em que $\mathbf{i} = (i_1, \ldots, i_\eta) \in \mathbb{N}^\eta$ e apenas um número finito desses é não-nulo.

[*]Um anel comutativo é uma estrutura algébrica com as mesmas propriedades de um corpo, com exceção da existência de elemento inverso para a multiplicação

O *polinômio nulo* é aquele que tem todos os seus coeficientes nulos, isto é, $c_{\mathbf{i}} = 0$ para todo $\mathbf{i} \in \mathbb{N}^\eta$.

Anel polinomial. O conjunto de todos os polinômios nas variáveis $X_1, \ldots, X_\eta$ sobre o anel R constitui também um anel, chamado de *anel polinomial* e denotado por $R[X_1, \ldots, X_\eta]$.

Notação e avaliação de um polinômio. Um polinômio é representado formalmente pela notação

$$P(X_1, \ldots, X_\eta) = \sum_{\mathbf{i} \in \mathbb{N}^\eta} c_{\mathbf{i}} \mathbf{X}^{\mathbf{i}} \in R[X_1, \ldots, X_\eta],$$

em que a notação $\mathbf{X}^{\mathbf{i}}$ significa $X_1^{i_1} \cdots X_\eta^{i_\eta}$. *Avaliar* esse polinômio em um ponto $\mathbf{a} = (a_1, \ldots, a_\eta) \in R^\eta$ significa calcular o valor da expressão

$$P(a_1, \ldots, a_\eta) = \sum_{\mathbf{i} \in \mathbb{N}^\eta} c_{\mathbf{i}} \mathbf{a}^{\mathbf{i}} \in R,$$

em que $\mathbf{a}^{\mathbf{i}} = a_1^{i_1} \cdots a_\eta^{i_\eta}$.

Enfoque recursivo. Um polinômio multivariável em $X_1, \ldots, X_\eta$ sobre o anel R pode ser visto como um polinômio univariável em X_j sobre o anel $R[X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_\eta]$. Dessa forma, existem coeficientes

$$Q_i(X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_\eta) \in R[X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_\eta],$$

em que $i \in \mathbb{N}$ e apenas um número finito desses é não-nulo, tais que

$$P(X_1, \ldots, X_\eta) = \sum_{i \in \mathbb{N}} Q_i(X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_\eta) X_j^i.$$

Sob essa ótica, $R[X_1, \ldots, X_\eta]$ e $R[X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_\eta][X_j]$ são essencialmente o mesmo objeto.

Grau de uma variável. Seja $P(X_1, \ldots, X_\eta)$ um polinômio nas variáveis $X_1, \ldots, X_\eta$ sobre um anel R. O *grau de uma variável* X_j no

(*i.e.*, a operação de divisão não é definida). Desse modo, todo corpo é um anel comutativo.

polinômio $P(X_1, \ldots, X_\eta)$ é definido como

$$d_j = \max\{i \in \mathbb{N} : Q_i(X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_\eta) \neq 0\}.$$

O *máximo grau do polinômio* $P(X_1, \ldots, X_\eta)$ é definido por

$$d = \max\{d_1, \ldots, d_\eta\}.$$

Funções polinomiais. A cada polinômio $P(X_1, \ldots, X_\eta)$ sobre um anel R é natural associar uma função polinomial P com domínio R^η e contradomínio R: o valor assumido pela função em um ponto $\mathbf{a} \in R^\eta$ é dado pelo valor obtido ao avaliar o polinômio $P(X_1, \ldots, X_\eta)$ no ponto $\mathbf{a}$. Polinômio e função polinomial são dois conceitos relacionados mas diferentes. A notação, no entanto, se confunde. Neste trabalho, o anel comutativo R é, na verdade, um corpo finito $\mathbb{F}_q$. Nesse caso, a distinção entre o polinômio $P(X_1, \ldots, X_\eta) \in \mathbb{F}_q[X_1, \ldots, X_\eta]$ e a função polinomial $P : \mathbb{F}_q^\eta \longrightarrow \mathbb{F}_q$ associada é ainda mais importante, como mostra a discussão a seguir.

Polinômios sobre corpos finitos. Obviamente, o polinômio nulo dá origem a uma função polinomial identicamente nula, isto é, à função $\mathbf{a} \longmapsto 0$, para todo $\mathbf{a} \in \mathbb{F}_q^\eta$. Entretanto, polinômios não-nulos também podem dar origem a funções identicamente nulas. Por exemplo, $P(X) = X + X^2 \in \mathbb{F}_2[X]$ não é, obviamente, o polinômio nulo. Contudo, a função polinomial associada, $P : \mathbb{F}_2 \longrightarrow \mathbb{F}_2, a \longmapsto a + a^2 = 0$ é a função identicamente nula de domínio $\mathbb{F}_2$. O próximo lema, no entanto, garante que isto é impossível de ocorrer se o tamanho do corpo finito é suficientemente grande. Por exemplo, o polinômio anterior $P(X) = X + X^2$ dá origem a uma função polinomial identicamente nula em $\mathbb{F}_2$, mas isto não ocorre em $\mathbb{F}_3$ ou $\mathbb{F}_4$.

Lema dos zeros esparsos. Seja $P(X_1, \ldots, X_\eta)$ um polinômio não-nulo nas variáveis $X_1, \ldots, X_\eta$ sobre um corpo $\mathbb{F}_q$ e d o máximo grau desse polinômio. Se $q > d$ então existe $\mathbf{a} = (a_1, \ldots, a_\eta) \in \mathbb{F}_q^\eta$ tal que $P(\mathbf{a}) \neq 0$.

Demonstração. (Baseada em [56, 17]) A prova se dá por indução sobre o número η de variáveis.

(i) O caso base $\eta = 1$ segue do fato de que um polinômio univariá-

vel $P(X_1)$ não-nulo de grau d possui no máximo d raízes distintas; como $|\mathbb{F}_q| = q > d$, é sempre possível escolher $a_1 \in \mathbb{F}_q$ tal que $P(a_1) \neq 0$.

(*ii*) Assuma agora que o teorema é válido para $\eta - 1$ e considere um polinômio $P(X_1, \ldots, X_\eta)$ não-nulo nas variáveis $X_1, \ldots, X_\eta$, sobre $\mathbb{F}_q$. Esse polinômio multivariável pode ser expresso como um polinômio univariável em X_η sobre $\mathbb{F}_q[X_1, \ldots, X_{\eta-1}]$:

$$P(X_1, \ldots, X_\eta) = \sum_{i=0}^{d_\eta} Q_i(X_1, \ldots, X_{\eta-1}) X_\eta^i,$$

em que d_η é o grau da variável X_η no polinômio $P(X_1, \ldots, X_\eta)$ e $Q_{d_\eta}(X_1, \ldots, X_{\eta-1}) \in \mathbb{F}_q[X_1, \ldots, X_{\eta-1}]$ é obrigatoriamente não-nulo. Então, pela hipótese de indução, existe $(a_1, \ldots, a_{\eta-1}) \in \mathbb{F}_q^{\eta-1}$ tal que $Q_{d_\eta}(a_1, \ldots, a_{\eta-1}) \neq 0$. Desse modo, o polinômio univariável $P(a_1, \ldots, a_{\eta-1}, X_\eta) \in \mathbb{F}_q[X_\eta]$ é não-nulo e tem grau $d_\eta \leq d$. Pelo mesmo argumento do caso base, é possível escolher $a_\eta \in \mathbb{F}_q$ tal que $P(a_1, \ldots, a_{\eta-1}, a_\eta) \neq 0$.

O resultado segue. □

Lema da probabilidade dos zeros. Seja $P(X_1, \ldots X_\eta)$ um polinômio não-nulo nas variáveis $X_1, \ldots, X_\eta$ sobre um corpo $\mathbb{F}_q$ e d o máximo grau desse polinômio. Se $q > d$ então

$$\Pr\left[P(\mathbf{a}) \neq 0\right] \geq \left(1 - \frac{d}{q}\right)^\eta,$$

em que $\mathbf{a} = (a_1, \ldots, a_\eta) \in \mathbb{F}_q^\eta$ é escolhido aleatoria e uniformemente.

Demonstração. (Baseada em [17]) Novamente, a prova se dá por indução sobre o número η de variáveis.

(*i*) O caso base $\eta = 1$ segue do fato de que um polinômio univariável $P(X_1)$ não-nulo de grau d possui no máximo d raízes distintas; assim, se $a_1 \in \mathbb{F}_q$ é escolhido aleatoria e uniformemente então

$$\Pr\left[P(a_1) = 0\right] \leq \frac{d}{q}.$$

(*ii*) Assuma agora que o teorema é válido para $\eta - 1$ e considere um polinômio $P(X_1, \ldots, X_\eta)$ não-nulo nas variáveis $X_1, \ldots, X_\eta$, sobre $\mathbb{F}_q$. Esse polinômio multivariável pode ser expresso como um polinômio univariável em X_η sobre $\mathbb{F}_q[X_1, \ldots, X_{\eta-1}]$:

$$P(X_1, \ldots, X_\eta) = \sum_{i=0}^{d_\eta} Q_i(X_1, \ldots, X_{\eta-1}) X_\eta^i,$$

em que d_η é o grau da variável X_η no polinômio $P(X_1, \ldots, X_\eta)$ e $Q_{d_n}(X_1, \ldots, X_{\eta-1}) \in \mathbb{F}_q[X_1, \ldots, X_{\eta-1}]$ é obrigatoriamente não-nulo. Seja $\mathbf{a} = (a_1, \ldots, a_\eta) \in \mathbb{F}_q^\eta$ escolhido aleatoria e uniformemente. Denote por A o evento "$P(a_1, \ldots, a_\eta) = 0$" e por B o evento "$Q_{d_n}(a_1, \ldots, a_{\eta-1}) = 0$". Então,

$$\begin{aligned}
\Pr[A] &\overset{(a)}{=} \Pr[A|B]\Pr[B] + \Pr[A|\bar{B}]\Pr[\bar{B}] \\
&\overset{(b)}{\leq} \Pr[B] + \Pr[A|\bar{B}]\Pr[\bar{B}] \\
&\overset{(c)}{\leq} \Pr[B] + \frac{d}{q}\Pr[\bar{B}] \\
&\overset{(d)}{=} \Pr[B] + \frac{d}{q}(1 - \Pr[B]) \\
&= \Pr[B]\left(1 - \frac{d}{q}\right) + \frac{d}{q} \\
&\overset{(e)}{\leq} \left[1 - \left(1 - \frac{d}{q}\right)^{\eta-1}\right]\left(1 - \frac{d}{q}\right) + \frac{d}{q} \\
&= 1 - \left(1 - \frac{d}{q}\right)^{\eta},
\end{aligned}$$

em que (a) segue da lei da probabilidade total; (b) segue de $\Pr[A|B] \leq 1$; (c) segue de $\Pr[A|\bar{B}] \leq d/q$, que, por sua vez, segue do caso base, visto que se $\bar{B}$ então $P(a_1, \ldots, a_{\eta-1}, X_\eta) \in \mathbb{F}_q[X_\eta]$ é um polinômio univariável não-nulo de grau $d_\eta \leq d$; (d) segue de $\Pr[B] + \Pr[\bar{B}] = 1$; e (e) segue da hipótese de indução:

$$\Pr[B] \leq 1 - \left(1 - \frac{d}{q}\right)^{\eta-1}.$$

O resultado segue pelo princípio da indução matemática. □

A.5 Espaços projetivos

Espaços projetivos e grassmannianos sobre corpos finitos. Seja W um espaço vetorial de dimensão m sobre um corpo finito $\mathbb{F}_q$. O *espaço projetivo* de W é definido como o conjunto de todos os subespaços vetoriais de W e é denotado por $\mathcal{P}(W)$. Além disso, o conjunto de todos os subespaços com uma dada dimensão k é denominado de *grassmanniano* e denotado por $\mathcal{P}(W, k)$.[†] Obviamente,

$$\mathcal{P}(W) = \bigcup_{k=0}^{m} \mathcal{P}(W, k).$$

Soma e interseção de subespaços. É possível combinar elementos do espaço projetivo de modo a obter outros elementos desse mesmo espaço.

(*i*) A *interseção de subespaços* $V_1 \cap V_2$, definida como no sentido usual da teoria dos conjuntos,

$$V_1 \cap V_2 = \{v : v \in V_1,\ v \in V_2\},$$

é um subespaço vetorial de W. De fato, a interseção $V_1 \cap V_2$ é o maior subespaço de $\mathcal{P}(W)$ que está contido simultaneamente em V_1 e V_2.

Em geral, a união (no sentido usual da teoria dos conjuntos) de subespaços vetoriais não é um subespaço vetorial. Entretanto, a seguinte definição é útil.

(*ii*) A *soma de subespaços* $V_1 \dot{+} V_2$, definida como o espaço gerado pelo conjunto $V_1 \cup V_2$,

$$V_1 \dot{+} V_2 = \{v_1 + v_2 : v_1 \in V_1,\ v_2 \in V_2\},$$

é um subespaço vetorial de W. De fato, a soma $V_1 \dot{+} V_2$ é o menor subespaço de $\mathcal{P}(W)$ que contém simultaneamente V_1 e V_2.

[†]O termo "espaço projetivo", na matemática, costuma significar o conjunto de todos os subespaços vetoriais *unidimensionais* de um dado espaço vetorial (*i.e.*, um grassmanniano com $k = 1$). No entanto, esta dissertação adota a nomenclatura imposta por Etzion & Vardy em [12] e considera o espaço projetivo como sendo o conjunto de todos os subespaços vetoriais (de quaisquer dimensões) de um dado espaço vetorial.

As dimensões do subespaço interseção e do subespaço soma se relacionam através de

$$\dim(V_1 \dotplus V_2) = \dim V_1 + \dim V_2 - \dim(V_1 \cap V_2). \qquad \text{(A.1)}$$

Isomorfismo. Todo espaço vetorial W de dimensão m sobre um corpo finito $\mathbb{F}_q$ é isomórfico a $\mathbb{F}_q^m$. Desse modo, sem perda de generalidade, a consideração $W = \mathbb{F}_q^m$ pode ser feita. Nesse caso, vetores do espaço vetorial são m-tuplas com elementos em $\mathbb{F}_q$.

Produto interno. Sejam $\mathbf{v} = (v_1, \ldots, v_m)$ e $\mathbf{u} = (u_1, \ldots, u_m)$ dois elementos do espaço vetorial $\mathbb{F}_q^m$. Define-se o *produto interno* desses vetores por

$$\mathbf{v} \cdot \mathbf{u} = \sum_{i=1}^{m} v_i u_i.$$

Uma observação importante é que, ao contrário de espaços vetoriais sobre $\mathbb{R}$ ou $\mathbb{C}$, esse produto interno não apresenta a propriedade da positividade, isto é, $\mathbf{v} \cdot \mathbf{v}$ pode ser zero mesmo com $\mathbf{v} \neq 0$. Por exemplo, se $\mathbf{v} = (1, 1) \in \mathbb{F}_2^2$ então $\mathbf{v} \cdot \mathbf{v} = 0$. Produtos internos nos quais isso ocorre são ditos *degenerados*. Se $\mathbf{v} \cdot \mathbf{u} = 0$ então $\mathbf{v}$ e $\mathbf{u}$ são ditos *ortogonais*.

Subespaço ortogonal. Se V é um subespaço vetorial de $\mathbb{F}_q^m$, o *subespaço ortogonal* a V, denotado por $V^\perp$, consiste nos vetores de $\mathbb{F}_q^m$ ortogonais aos vetores de V:

$$V^\perp = \{\mathbf{w} \in \mathbb{F}_q^m : \mathbf{w} \cdot \mathbf{v} = 0,\ \mathbf{v} \in V\}.$$

O subespaço ortogonal é mesmo um subespaço vetorial de $\mathbb{F}_q^m$. Ainda que $V \dotplus V^\perp$ não seja necessariamente $\mathbb{F}_q^m$, tem-se sempre

$$\dim V + \dim V^\perp = m.$$

Enumeração de subespaços vetoriais. O número de subespaços vetoriais de $\mathbb{F}_q^m$ com dimensão k é dado por

$$\binom{m}{k}_q = \prod_{i=0}^{k-1} \frac{q^m - q^i}{q^k - q^i} = \frac{(q^m - 1)(q^m - q) \cdots (q^m - q^{k-1})}{(q^k - 1)(q^k - q) \cdots (q - q^{k-1})}, \qquad \text{(A.2)}$$

quantidade conhecida como *coeficiente binomial gaussiano*. (Na demonstração a seguir, o termo *k-base* é equivalente a uma k-tupla ordenada de vetores linearmente independentes de $\mathbb{F}_q^m$.)

Demonstração. (Baseada em [6]) Existem

$$(q^m - 1)(q^m - q) \cdots (q^m - q^{k-1})$$

k-bases diferentes em $\mathbb{F}_q^m$:

(*i*) o primeiro vetor da k-base pode ser qualquer um dos q^m vetores de $\mathbb{F}_q^m$, exceto o vetor nulo;

(*ii*) o segundo vetor da k-base pode ser qualquer um dos q^m vetores de $\mathbb{F}_q^m$, exceto os q múltiplos do vetor já escolhido;

$$\vdots$$

(*iii*) o k-ésimo vetor da k-base pode ser qualquer um dos q^m vetores de $\mathbb{F}_q^m$, exceto os q^{k-1} vetores do espaço gerado pelos $k-1$ vetores já escolhidos.

Assim, cada uma das $(q^m - 1)(q^m - q) \cdots (q^m - q^{k-1})$ k-bases de $\mathbb{F}_q^m$ gera algum subespaço vetorial de $\mathbb{F}_q^m$ com dimensão k. Entretanto, k-bases diferentes podem gerar o mesmo subespaço. Com efeito, seja V um subespaço vetorial qualquer de $\mathbb{F}_q^m$ com dimensão k. Existem

$$(q^k - 1)(q^k - q) \cdots (q^k - q^{k-1})$$

possíveis k-bases que geram V:

(*i*) o primeiro vetor da k-base pode ser qualquer um dos q^k vetores de V, exceto o vetor nulo;

(*ii*) o segundo vetor da k-base pode ser qualquer um dos q^k vetores de V, exceto os q múltiplos do vetor já escolhido;

$$\vdots$$

(*iii*) o k-ésimo vetor da k-base pode ser qualquer um dos q^k vetores de V, exceto os q^{k-1} vetores do espaço gerado pelos $k-1$ vetores já escolhidos.

Portanto, eliminando as repetições, existem

$$\frac{(q^m-1)(q^m-q)\cdots(q^m-q^{k-1})}{(q^k-1)(q^k-q)\cdots(q^k-q^{k-1})}$$

subespaços vetoriais de $\mathbb{F}_q^m$ com dimensão k. □

Cardinalidade do espaço projetivo e do grassmanniano. Como consequência imediata, tem-se que a cardinalidade de um grassmanniano de $\mathbb{F}_q^m$ com dimensão k é

$$\left|\mathcal{P}(\mathbb{F}_q^m, k)\right| = \binom{m}{k}_q \tag{A.3}$$

e a cardinalidade do espaço projetivo de $\mathbb{F}_q^m$ é

$$\left|\mathcal{P}(\mathbb{F}_q^m)\right| = \sum_{k=0}^{m} \binom{m}{k}_q. \tag{A.4}$$

Limitantes. Por fim, os seguintes limitantes são válidos:

$$q^{k(m-k)} < \left|\mathcal{P}(\mathbb{F}_q^m, k)\right| < 4q^{k(m-k)} \tag{A.5}$$

para $0 < k < m$ e

$$q^{\frac{m^2}{4}} < \left|\mathcal{P}(\mathbb{F}_q^m)\right| < 4(m+1)q^{\frac{m^2}{4}}. \tag{A.6}$$

Demonstração. Uma prova de (A.5) se encontra em [33, Lemma 4]. Já (A.6) segue de (A.5) e do fato de que $q^{k(m-k)}$ atinge seu máximo em $k = m/2$. Assim, tem-se

$$q^{\frac{m^2}{4}} < \sum_{k=0}^{m} q^{k(m-k)} < \sum_{k=0}^{m} \binom{m}{k}_q < \sum_{k=0}^{m} 4q^{k(m-k)} < 4(m+1)q^{\frac{m^2}{4}},$$

como desejado. □

A.6 Espaços métricos

Métricas e espaços métricos. Espaços métricos são conjuntos nos quais existe a noção de "distância" entre elementos. Seja $\mathcal{M}$ um conjunto não-vazio qualquer. Uma *métrica* ou *distância* em $\mathcal{M}$ é uma função $d : \mathcal{M} \times \mathcal{M} \to [0, \infty)$ tal que, para todo $x, y, z \in \mathcal{M}$, as seguintes propriedades são todas satisfeitas:

(*i*) $d(x, y) = 0$ se e somente se $x = y$,

(*ii*) $d(x, y) = d(y, x)$ (*simetria*) e

(*iii*) $d(x, y) \leq d(x, z) + d(z, y)$ (*desigualdade triangular*).

O conjunto $\mathcal{M}$ com a métrica $d(\cdot, \cdot)$ é dito ser um *espaço métrico*.

Exemplos. A seguir são apresentados alguns exemplos de espaços métricos.

(*i*) A função $d_{\mathrm{E}}(\cdot, \cdot)$ definida por

$$d_{\mathrm{E}}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=0}^{n} (x_i - y_i)^2},$$

em que $\mathbf{x} = (x_1, \ldots, x_n)$ e $\mathbf{y} = (y_1, \ldots, y_n)$ é uma métrica em $\mathbb{R}^n$, conhecida como *distância euclidiana*.

O interesse deste trabalho está nos *espaços métricos finitos* (isto é, com um número finito de elementos) cuja distância assume apenas valores inteiros.

(*ii*) A função $d_{\mathrm{H}}(\cdot, \cdot)$ definida por

$$d_{\mathrm{H}}(\mathbf{x}, \mathbf{y}) = \sum_{i=0}^{n} \delta_{\mathrm{H}}(x_i, y_i),$$

em que $\mathbf{x} = (x_1, \ldots, x_n)$ e $\mathbf{y} = (y_1, \ldots, y_n)$ e

$$\delta_{\mathrm{H}}(x, y) = \begin{cases} 1, & \text{se } x \neq y, \\ 0, & \text{caso contrário} \end{cases}$$

é uma métrica em $\mathbb{Z}_q^n$, conhecida como *distância de Hamming*.

(*iii*) A função $d_{\mathrm{L}}(\cdot,\cdot)$ definida por

$$d_{\mathrm{L}}(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n} \delta_{\mathrm{L}}(x_i, y_i),$$

em que $\mathbf{x} = (x_1, \ldots, x_n)$ e $\mathbf{y} = (y_1, \ldots, y_n)$ e

$$\delta_{\mathrm{L}}(x,y) = \min\{(y-x) \bmod q, (x-y) \bmod q\}$$

é uma métrica em $\mathbb{Z}_q^n$, conhecida como *distância de Lee*. Essa distância coincide com a de Hamming nos casos binário e ternário ($q = 2$ ou 3).

Distância em grafos. Dado um grafo $G = (\mathcal{V}, \mathcal{E})$ simples, não-direcionado, conectado e sem *loops*, é sempre possível definir uma distância $d_G(v_1, v_2)$ entre dois vértices $v_1, v_2 \in \mathcal{V}$ como o comprimento de uma *geodésica* (caminho de menor distância) entre v_1 e v_2.

Potência cartesiana de um espaço métrico. Se $\mathcal{M}$ é um espaço métrico com métrica $\delta(\cdot,\cdot)$ então $\mathcal{M}^n$ também é, com métrica $d(\cdot,\cdot)$ dada por

$$d(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n} \delta(x_i, y_i),$$

em que $\mathbf{x} = (x_1, \ldots, x_n)$ e $\mathbf{y} = (y_1, \ldots, y_n)$.

Esferas. Dado um espaço métrico finito $\mathcal{M}$ com distância $d(\cdot,\cdot)$, define-se a *esfera* de centro x_0 e raio r como o conjunto de todos os pontos de $\mathcal{M}$ cuja distância até x_0 é no máximo r, isto é,

$$\mathcal{B}_{(\mathcal{M},d)}(x_0, r) = \{x \in \mathcal{M} : d(x, x_0) \leq r\}.$$

O *volume* dessa esfera é definido por

$$\mathrm{Vol}_{(\mathcal{M},d)}(x_0, r) = \left|\mathcal{B}_{(\mathcal{M},d)}(x_0, r)\right|.$$

O espaço métrico $\mathcal{M}$ é dito ser *regular* se o volume de esferas em $\mathcal{M}$ não depende dos centros das mesmas; se for esse o caso, utiliza-se a notação $\mathrm{Vol}_{(\mathcal{M},d)}(r)$.

Isometria. Um mapeamento bijetivo entre dois espaços métricos é dito ser uma *isometria* se preserva a distância entre pontos. Ou seja, uma bijeção $f : \mathcal{M}_1 \longrightarrow \mathcal{M}_2$ é uma isometria se, para todo $x, y \in \mathcal{M}_1$,

$$d_1(x, y) = d_2(f(x), f(y)),$$

em que $d_1(\cdot, \cdot)$ é a métrica de $\mathcal{M}_1$ e $d_2(\cdot, \cdot)$ é a métrica de $\mathcal{M}_2$.

A.7 Partições

Definição. Dado um conjunto $\mathcal{S}$, uma *partição* de $\mathcal{S}$ consiste em uma coleção $\Gamma = \{\mathcal{S}_1, \ldots, \mathcal{S}_p\}$ de p subconjuntos não-vazios de $\mathcal{S}$ tais que

(i) a união de todos os elementos da partição é igual ao conjunto $\mathcal{S}$:

$$\mathcal{S}_1 \cup \cdots \cup \mathcal{S}_p = \mathcal{S} \quad \text{e}$$

(ii) a interseção de quaisquer dois elementos da partição é vazia:

$$\mathcal{S}_i \cap \mathcal{S}_j = \emptyset, \quad i \neq j.$$

Refinamentos. Uma partição Γ' é dita ser um *refinamento* de outra partição Γ se todo elemento de Γ' é subconjunto de algum elemento de Γ.

Exemplo. O conjunto $\mathcal{S} = \{a, b, c, d\}$ possui 15 partições, a saber:

$$\Gamma_0 = \{\{a, b, c, d\}\},$$
$$\Gamma_1 = \{\{a\}, \{b, c, d\}\},\ \Gamma_2 = \{\{b\}, \{a, c, d\}\},$$
$$\Gamma_3 = \{\{c\}, \{a, b, d\}\},\ \Gamma_4 = \{\{d\}, \{a, b, c\}\},$$
$$\Gamma_5 = \{\{a, b\}, \{c, d\}\},\ \Gamma_6 = \{\{a, c\}, \{b, d\}\},\ \Gamma_7 = \{\{a, d\}, \{b, c\}\},$$
$$\Gamma_8 = \{\{a\}, \{b\}, \{c, d\}\},\ \Gamma_9 = \{\{a\}, \{c\}, \{b, d\}\},\ \Gamma_{10} = \{\{a\}, \{d\}, \{b, c\}\},$$
$$\Gamma_{11} = \{\{a, b\}, \{c\}, \{d\}\},\ \Gamma_{12} = \{\{a, c\}, \{b\}, \{d\}\},\ \Gamma_{13} = \{\{a, d\}, \{b\}, \{c\}\},$$
$$\Gamma_{14} = \{\{a\}, \{b\}, \{c\}, \{d\}\}.$$

Tem-se que:

(i) As partições Γ_8, Γ_{11} e Γ_{14} são refinamentos da partição Γ_5.

(ii) A partição Γ_8 é refinamento das partições Γ_0, Γ_1, Γ_2 e Γ_5.

(*iii*) Qualquer partição é refinamento da partição Γ_0.

(*iv*) A partição Γ_{14} é refinamento de qualquer partição.

Ordenamento parcial. O conjunto $\mathcal{X}$ de todas as partições de um determinado conjunto $\mathcal{S}$ pode ser parcialmente ordenado por $\preceq$, em que $\Gamma_i \preceq \Gamma_j$ se e somente se Γ_j é um refinamento de Γ_i. Esse fato, porém, não é utilizado na dissertação.

APÊNDICE B

Códigos corretores de erros

Este apêndice apresenta, brevemente, conceitos relativos à teoria da codificação para controle de erros [27] que são utilizados neste trabalho. Em particular, são abordados os limitantes esféricos de Hamming e Gilbert-Varshamov e o puncionamento de códigos, resultando no limitante de Singleton.

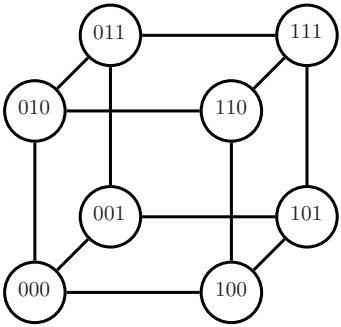


Figura B.1: Cubo de Hamming representando o espaço métrico $\mathbb{Z}_2^3$.

Inicialmente, a teoria da codificação foi desenvolvida no espaço métrico particular conhecido como *hipercubo de Hamming*, que nada mais é que $\mathbb{Z}_2^n = \{0,1\}^n$ (ou seja, o conjunto de todas as n-tuplas binárias)

com a métrica de Hamming. O hipercubo de Hamming com $n = 3$ é mostrado na Figura B.1. Posteriormente, essa teoria foi expandida para o caso não-binário, isto é, codificação em $\mathbb{Z}_q^n$ com a métrica de Hamming ou com a métrica de Lee, essa última mais adequada a sistemas de comunicação que operam com modulação de fase.

Os resultados deste apêndice são válidos para quaisquer espaços métricos finitos com métrica inteira e são particularizados aqui para $\mathbb{Z}_q^n$ com a métrica de Hamming. Os Capítulos 4 e 5 aplicam a teoria a $\mathcal{P}(\mathbb{F}_q^m)$ e $\mathcal{P}(\mathbb{F}_q^m)^n$, respectivamente, tendo como métrica a *distância de subespaço*, a ser definida em momento oportuno.

B.1 Códigos corretores e detectores de erro

Seja $\mathcal{M}$ um espaço métrico finito com distância associada $d(\cdot,\cdot)$ assumindo apenas valores inteiros. Em adição, considere um canal de comunicação que opera sobre $\mathcal{M}$, isto é, um sistema que tem como entrada um símbolo $x \in \mathcal{M}$ e saída um símbolo $y \in \mathcal{M}$. Na ausência de erros, o símbolo transmitido x é igual ao símbolo recebido y. Caso contrário, é dito que um erro de *peso* $d(x, y)$ ocorreu.

É considerado que qualquer erro de peso $t \in \mathbb{N}$ ou menos pode ocorrer (mas nenhum erro de peso superior a t). Para permitir uma comunicação mesmo nessas circunstâncias, limita-se os símbolos na entrada do canal a um subconjunto particular de $\mathcal{M}$. Desse modo, é possível detectar e até mesmo corrigir eventuais erros ocorridos, como mostrado a seguir.

Códigos. Um *código* $\mathcal{C}$ sobre o espaço métrico $\mathcal{M}$ é qualquer subconjunto não-vazio de $\mathcal{M}$. Os elementos de um código são chamados de *palavras-código*.

Distância mínima. A *distância mínima* $d(\mathcal{C})$ de um código $\mathcal{C} \subseteq \mathcal{M}$ é definida como a menor distância entre duas palavras-código distintas, ou seja,

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

Capacidade de controle de erros. Um código é dito ser t-detector (resp., t-corretor) se ele for capaz de detectar (resp., corrigir) todos os possíveis erros de peso t ou menos. Assim, utilizando decodificação de

mínima distância,

$$\hat{x} = \operatorname*{argmin}_{c \in \mathcal{C}} d(c, y),$$

prova-se que as seguintes afirmações acerca de um código $\mathcal{C}$ são equivalentes.

(i) $d(\mathcal{C}) = d$.

(ii) $\mathcal{C}$ é $(d-1)$-detector.

(iii) $\mathcal{C}$ é $\lfloor \frac{d-1}{2} \rfloor$-corretor.

Taxa. A *taxa* $R(\mathcal{C})$ de um código $\mathcal{C} \subseteq \mathcal{M}$ é definida pela razão entre número de dígitos q-ários "aproveitados" e o de dígitos q-ários totais:

$$R(\mathcal{C}) = \frac{\log_q |\mathcal{C}|}{\log_q |\mathcal{M}|}.$$

No caso de códigos sobre $\mathbb{Z}_q^n$, a taxa se traduz em $\log_q |\mathcal{C}| / n$.

Compromisso entre a taxa e a capacidade de controle de erros. Evidentemente, é desejável que a tanto taxa (ou, equivalentemente, a cardinalidade) quanto a capacidade de controle de erros (ou, equivalentemente, a distância mínima) de um código sejam as maiores possíveis. Esses objetivos, entretanto, são conflitantes. As duas seções restantes apresentam regiões do "plano" $d(\mathcal{C}) \times |\mathcal{C}|$ nas quais a existência ou inexistência de códigos são asseguradas.

B.2 Limitantes esféricos

Definem-se

$$\begin{aligned}
\mathrm{Vol}^{\min}(r) &= \min_{x_0 \in \mathcal{M}} \mathrm{Vol}(x_0, r), \\
\mathrm{Vol}^{\mathrm{med}}(r) &= \frac{1}{|\mathcal{M}|} \sum_{x_0 \in \mathcal{M}} \mathrm{Vol}(x_0, r), \\
\mathrm{Vol}^{\max}(r) &= \max_{x_0 \in \mathcal{M}} \mathrm{Vol}(x_0, r).
\end{aligned}$$

Limitante superior de Hamming. Seja $\mathcal{C} \subseteq \mathcal{M}$ um código qualquer com distância mínima $d(\mathcal{C}) = d$. Considere esferas de raio $r =$

$\left\lfloor \frac{d-1}{2} \right\rfloor$ centradas nas palavras-código de $\mathcal{C}$, como ilustrado na Figura B.2 para $d = 3$.

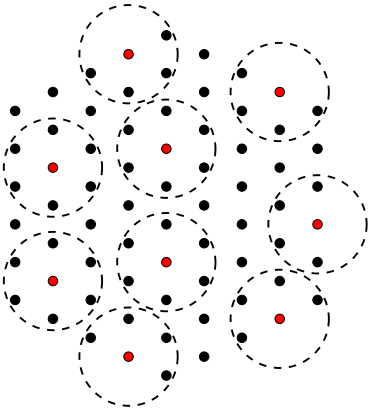


Figura B.2: Ilustração do limitante de Hamming.

As esferas são todas disjuntas, de modo que

$$\begin{aligned} |\mathcal{M}| &\geq \sum_{c \in \mathcal{C}} \mathrm{Vol}(c, r) \\ &\geq \sum_{c \in \mathcal{C}} \mathrm{Vol}^{\min}(r) \\ &= |\mathcal{C}| \, \mathrm{Vol}^{\min}(r) \end{aligned}$$

e, portanto, a cardinalidade de $\mathcal{C}$ é limitada por

$$|\mathcal{C}| \leq \frac{|\mathcal{M}|}{\mathrm{Vol}^{\min}(\lfloor \frac{d-1}{2} \rfloor)}.$$

Esse resultado é também conhecido como *limitante do empacotamento de esferas* (em inglês *sphere-packing bound*).

Limitante inferior de Gilbert-Varshamov. Considere o seguinte algoritmo *greedy* para a construção de um código $\mathcal{C} \subseteq \mathcal{M}$ com distância mínima $d(\mathcal{C}) = d$. Inicia-se com $\mathcal{C} = \{c_1\}$, em que c_1 é qualquer ponto de $\mathcal{M}$. Posteriormente, adicionam-se palavras-código $c_2, c_3, \ldots$, uma a uma, desde que c_i esteja fora da união das esferas de raio $r = d - 1$

centradas nas palavras-código já adicionadas:

$$c_i \notin \bigcup_{j=1}^{i-1} \mathcal{B}(c_j, r).$$

Assim, ao final desse procedimento, tem-se um código com distância mínima d que satisfaz

$$\begin{aligned} |\mathcal{M}| &\leq \sum_{c \in \mathcal{C}} \mathrm{Vol}(c, r) \\ &\leq \sum_{c \in \mathcal{C}} \mathrm{Vol}^{\max}(r) \\ &= |\mathcal{C}|\, \mathrm{Vol}^{\max}(r), \end{aligned}$$

ou, equivalentemente,

$$|\mathcal{C}| \geq \frac{|\mathcal{M}|}{\mathrm{Vol}^{\max}(d-1)}.$$

Esse resultado, ilustrado na Figura B.3 para $d = 3$, é também conhecido como limitante da cobertura por esferas (em inglês *sphere-covering bound*).

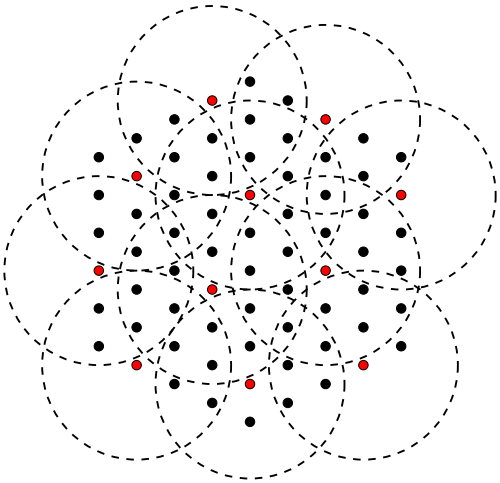


Figura B.3: Ilustração do limitante de Gilbert-Varshamov.

De acordo com Gu & Fuja em [21], $\mathrm{Vol}^{\max}(\cdot)$ pode ser substituído

por $\mathrm{Vol}^{\mathrm{med}}(\cdot)$ na desigualdade acima. Ou seja, existe código $\mathcal{C}$ com distância mínima $d(\mathcal{C}) = d$ tal que

$$|\mathcal{C}| \geq \frac{|\mathcal{M}|}{\mathrm{Vol}^{\mathrm{med}}(d-1)}.$$

Esse resultado é conhecido como o *limitante generalizado de Gilbert-Varshamov*.

Particularização para $\mathbb{Z}_q^n$ com a métrica de Hamming. Se o espaço métrico $\mathcal{M}$ em consideração for $\mathbb{Z}_q^n$ com a métrica de Hamming, então $|\mathcal{M}| = q^n$. Em adição, pode-se mostrar que esse espaço métrico é regular (*i.e.*, o volume de esferas independe do centro) e que

$$\mathrm{Vol}(r) = \sum_{j=0}^{r} \binom{n}{j} (q-1)^j.$$

Assim, o limitante de Hamming afirma que *para todo* código $\mathcal{C} \subseteq \mathbb{Z}_q^n$ com distância mínima $d(\mathcal{C}) = d$, vale

$$|\mathcal{C}| \leq \frac{q^n}{\sum_{j=0}^{\lfloor \frac{d-1}{2} \rfloor} \binom{n}{j} (q-1)^j},$$

enquanto que o limitante de Gilbert-Varshamov afirma que *existe* um código $\mathcal{C} \subseteq \mathbb{Z}_q^n$ com distância mínima $d(\mathcal{C}) = d$ tal que

$$|\mathcal{C}| \geq \frac{q^n}{\sum_{j=0}^{d-1} \binom{n}{j} (q-1)^j}.$$

B.3 Puncionamento de códigos

O conceito de *puncionamento* aqui considerado consiste, intuitivamente, em "transportar" um código de um espaço métrico a outro, estando esse último espaço "contido" no primeiro. Essa operação deve ter a propriedade de preservar o tamanho do código (sob certas condições); já a distância mínima é permitida que sofra alguma redução. A Figura B.4 ilustra a ideia.

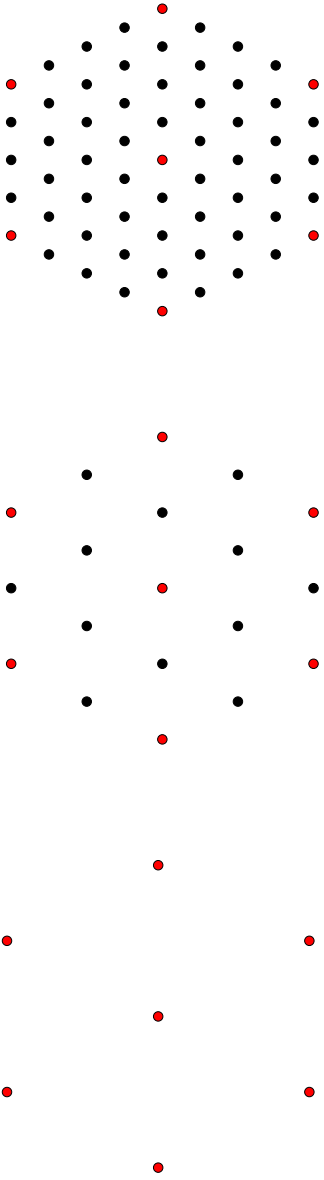


Figura B.4: Ilustração da operação de puncionamento.

Operação de puncionamento. Um *puncionamento* é uma operação $(\cdot)^{\blacktriangledown}$ que leva o código $\mathcal{C} \subseteq \mathcal{M}$ no código $\mathcal{C}^{\blacktriangledown} \subseteq \mathcal{M}'$ e que satisfaz a seguinte propriedade:

$$d(\mathcal{C}) > \Delta \implies \left(|\mathcal{C}^{\blacktriangledown}| = |\mathcal{C}|\right) \wedge \left(d(\mathcal{C}) - d(\mathcal{C}^{\blacktriangledown}) \leq \Delta\right),$$

para algum $\Delta \in \mathbb{N}$. Note o abuso de notação: d é ora a métrica de $\mathcal{M}$, ora a métrica de $\mathcal{M}'$.

Limitante de Singleton. Suponha que exista uma sequência de p puncionamentos $\{(\cdot)^{\blacktriangledown_i}\}_{i=0}^{p-1}$, em que $(\cdot)^{\blacktriangledown_i}$ leva códigos de $\mathcal{M}_i$ em códigos de $\mathcal{M}_{i+1}$. Seja Δ o parâmetro de todos os puncionamentos e considere $p = \left\lfloor \frac{d-1}{\Delta} \right\rfloor$. Seja $\mathcal{C} \subseteq \mathcal{M}_0$ um código com distância mínima $d(\mathcal{C}) = d$ qualquer. Aplicando os puncionamentos seguidamente, obtém-se o código $\mathcal{C}' = \mathcal{C}^{\blacktriangledown_0 \cdots \blacktriangledown_{p-1}} \subseteq \mathcal{M}_p$ satisfazendo

$$d(\mathcal{C}') \geq d - p\Delta = d - \left\lfloor \frac{d-1}{\Delta} \right\rfloor \Delta \geq d - \left(\frac{d-1}{\Delta}\right)\Delta = 1$$

e, portanto,

$$|\mathcal{C}| = |\mathcal{C}'| \leq |\mathcal{M}_p|.$$

Particularização para $\mathbb{Z}_q^n$ com a métrica de Hamming. A operação natural de puncionamento em $\mathbb{Z}_q^n$ consiste na remoção de coordenadas das palavras-código. Por exemplo, seja a operação de puncionamento de uma *tupla* dada por

$$\begin{array}{rcl} (\cdot)^{\blacktriangledown} : \mathbb{Z}_q^n & \longrightarrow & \mathbb{Z}_q^{n-1} \\ \mathbf{x} = (x_1, \ldots, x_n) & \longmapsto & \mathbf{x}^{\blacktriangledown} = (x_1, \ldots, x_{n-1}) \end{array}$$

e a operação de puncionamento de um *código* $\mathcal{C} \subseteq \mathbb{Z}_q^n$ dada por

$$\mathcal{C}^{\blacktriangledown} = \{\mathbf{c}^{\blacktriangledown} : \mathbf{c} \in \mathcal{C}\} \subseteq \mathbb{Z}_q^{n-1}.$$

Dessa maneira, se um código $\mathcal{C}$ tiver distância mínima $d(\mathcal{C}) = d > 1$, sua versão puncionada $\mathcal{C}^{\blacktriangledown}$ tem distância mínima $d(\mathcal{C}^{\blacktriangledown}) \geq d - 1$. Isso porque a remoção de uma coordenada ou mantém a distância entre duas palavras-código ou reduz essa distância de 1.

Em termos do recém exposto, tem-se

$$\Delta = 1,$$

$$p = \left\lfloor \frac{d-1}{\Delta} \right\rfloor = d - 1,$$
$$\mathcal{M}_i = \mathbb{Z}_q^{n-i}, \quad i = 0, \ldots, p,$$

de modo que $(\cdot)^{\blacktriangledown_i}$ é o puncionamento que leva códigos de $\mathbb{Z}_q^{n-i}$ em códigos de $\mathbb{Z}_q^{n-(i+1)}$.

Assim, o limitante de Singleton aplicado a $\mathbb{Z}_q^n$ com a métrica de Hamming afirma que todo código $\mathcal{C} \subseteq \mathbb{Z}_q^n$ com distância mínima $d(\mathcal{C}) = d$ satisfaz

$$|\mathcal{C}| \leq q^{n-d+1}.$$

Referências bibliográficas

[1] R. Ahlswede, N. Cai, R. Li, and R. Yeung, "Network Information Flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[2] N. Cai and R. Yeung, "Network Coding and Error Correction," in *Proceedings of the 2002 IEEE Information Theory Workshop (ITW'02)*, Bangalore, India, Oct. 2002, pp. 119–112.

[3] ——, "Network Error Correction, Part II: Lower Bounds," *Communications in Information and Systems*, vol. 6, no. 1, pp. 37–54, 2006.

[4] R. Calderbank, "Multilevel Codes and Multistage Decoding," *IEEE Transactions on Communications*, vol. 37, no. 3, pp. 222–229, Mar. 1989.

[5] P. Chou, Y. Wu, and K. Jain, "Practical Network Coding," in *Proceedings of the 41st Annual Allerton Conference on Communication, Control, and Computing (Allerton'03)*, Monticello, Illinois, Oct. 2003.

[6] H. Cohn, "Projective Geometry over $\mathbb{F}_1$ and the Gaussian Binomial Coefficients," *American Mathematical Monthly*, vol. 111, pp. 487–495, Jun. 2004.

[7] R. Dougherty, C. Freiling, and K. Zeger, "Insufficiency of Linear Coding in Network Information Flow," *IEEE Transactions on Information Theory*, vol. 51, no. 8, pp. 2745–2759, Aug. 2005.

[8] P. Elias, A. Feinstein, and C. Shannon, "A Note on the Maximum Flow Through a Network," *IRE Transactions on Information Theory*, vol. 2, no. 4, pp. 117–119, Dec. 1956.

[9] E. Erez and M. Feder, "Convolutional Network Codes," in *Proceedings of the 2004 IEEE International Symposium on Information Theory (ISIT'04)*, Chicago, Illinois, Jun. 2004, p. 146.

[10] ——, "Convolutional Network Codes for Cyclic Networks," in *Proceedings of the 1st Workshop on Network Coding, Theory, and Applications (NetCod'05)*, Riva del Garda, Italy, Apr. 2005.

[11] T. Etzion and N. Silberstein, "Error-Correcting Codes in Projective Spaces Via Rank-Metric Codes and Ferrers Diagrams," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 2909–2919, Jul. 2009.

[12] T. Etzion and A. Vardy, "Error-Correcting Codes in Projective Space," in *Proceedings of the 2008 IEEE International Symposium on Information Theory (ISIT'08)*, Toronto, Canada, Jul. 2008, pp. 871–875.

[13] L. Ford Jr. and D. Fulkerson, "Maximal Flow Through a Network," *Canadian Journal of Mathematics*, vol. 8, pp. 399–404, 1956.

[14] ——, "A Simple Algorithm for Finding Maximal Network Flows and an Application to the Hitchcock Problem," *Canadian Journal of Mathematics*, vol. 9, pp. 210–218, 1957.

[15] C. Fragouli and E. Soljanin, "A Connection between Network Coding and Convolutional Codes," in *Proceedings of the 2004 IEEE International Conference on Communications (ICC'04)*, vol. 2, Paris, France, Jun. 2004, pp. 661–666.

[16] ——, *Network Coding Applications*, ser. Foundations and Trends® in Networking. Now Publishers Inc, 2007.

[17] ——, *Network Coding Fundamentals*, ser. Foundations and Trends® in Networking. Now Publishers Inc, 2007.

[18] E. Gabidulin and M. Bossert, "Codes for Network Coding," in *Proceedings of the 2008 IEEE International Symposium on Information Theory (ISIT'08)*, Toronto, Canada, Jul. 2008, pp. 867–870.

[19] M. Gadouleau and Z. Yan, "Construction and Covering Properties of Constant-Dimension Codes," *Computing Research Repository (CoRR)*, vol. abs/0903.2675, Mar. 2009, submitted to IEEE Transactions on Information Theory.

[20] C. Gkantsidis and P. Rodriguez, "Network Coding for Large Scale Content Distribution," in *Proceedings of the 24th IEEE Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'05)*, vol. 4, Miami, Florida, Mar. 2005, pp. 2235–2245.

[21] J. Gu and T. Fuja, "A Generalized Gilbert-Varshamov Bound Derived via Analysis of a Code-search Algorithm," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 1089–1093, May 1993.

[22] N. Harvey, "Deterministic Network Coding by Matrix Completion," Master's thesis, Massachusetts Institute of Technology, May 2005.

[23] N. Harvey, D. Karger, and K. Murota, "Deterministic Network Coding by Matrix Completion," in *Proceedings of the 16th ACM/SIAM Symposium on Discrete Algorithms (SODA'05)*, Vancouver, Canada, Jan. 2005, pp. 489–498.

[24] T. Ho, R. Koetter, M. Médard, D. Karger, and M. Effros, "The Benefits of Coding over Routing in a Randomized Setting," in *Proceedings of the 2003 IEEE International Symposium on Information Theory (ISIT'03)*, Yokohama, Japan, Jun. 2003, p. 442.

[25] T. Ho and D. Lun, *Network Coding: An Introduction.* Cambridge University Press, 2008.

[26] T. Ho, M. Médard, R. Koetter, D. Karger, M. Effros, J. Shi, and B. Leong, "A Random Linear Network Coding Approach to Multicast," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4413–4430, Oct. 2006.

[27] W. Huffman and V. Pless, *Fundamentals of Error-Correcting Codes*. Cambridge University Press, 2003.

[28] H. Imai and S. Hirakawa, "A New Multilevel Coding Method Using Error-Correcting Codes," *IEEE Transactions on Information Theory*, vol. 23, no. 3, pp. 371–377, May 1977.

[29] S. Jaggi, P. Chou, and K. Jain, "Low Complexity Algebraic Multicast Network Codes," in *Proceedings of the 2003 IEEE International Symposium on Information Theory (ISIT'03)*, Yokohama, Japan, Jun. 2003, p. 368.

[30] S. Jaggi, P. Sanders, P. Chou, M. Effros, S. Egner, K. Jain, and L. Tolhuizen, "Polynomial Time Algorithms for Multicast Network Code Construction," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 1973–1982, Jun. 2005.

[31] A. Khaleghi and F. Kschischang, "Projective Space Codes for the Injection Metric," in *Proceedings of the 11th Canadian Workshop on Information Theory (CWIT'09)*, Ottawa, Canada, May 2009, pp. 9–12.

[32] R. Koetter and F. Kschischang, "Coding for Errors and Erasures in Random Network Coding," in *Proceedings of the 2007 IEEE International Symposium on Information Theory (ISIT'07)*, Nice, France, Jun. 2007, pp. 791–795.

[33] ——, "Coding for Errors and Erasures in Random Network Coding," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3579–3591, Aug. 2008.

[34] R. Koetter and M. Médard, "Beyond Routing: An Algebraic Approach to Network Coding," in *Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'02)*, vol. 1, New York, New York, Jul. 2002, pp. 122–130.

[35] ——, "An Algebraic Approach to Network Coding," *IEEE/ACM Transactions on Networking*, vol. 11, no. 5, pp. 782–795, Oct. 2003.

[36] A. Kohnert and S. Kurz, "Construction of Large Constant Dimension Codes with a Prescribed Minimum Distance," in *Mathematical Methods in Computer Science*, J. Calmet, W. Geiselmann, and

J. Müller-Quade, Eds. Berlin, Germany: Springer-Verlag, Dec. 2008, vol. 5393, ch. Lecture Notes in Computer Science, pp. 31–42.

[37] F. Kschischang, R. Koetter, and D. Silva, "Nonlinear Coding for Linear Network Coding: Robust Multicasting with Vector Spaces," Presented on the 26th Brazilian Symposium on Telecommunications (SBrT'08), Rio de Janeiro, Brazil, Sep. 2008, Plenary 1.

[38] R. Li and R. Yeung, "On Convolutional Network Coding," in *Proceedings of the 2006 IEEE International Symposium on Information Theory (ISIT'06)*, Seattle, Washington, Jul. 2006, pp. 1743–1747.

[39] R. Li, R. Yeung, and N. Cai, "Linear Network Coding," *IEEE Transactions on Information Theory*, vol. 49, no. 2, pp. 371–381, Feb. 2003.

[40] Z. Li, B. Li, and L. C. Lau, "A Constant Bound on Throughput Improvement of Multicast Network Coding in Undirected Networks," *IEEE Transactions on Information Theory*, vol. 55, no. 3, pp. 1016–1026, Mar. 2009.

[41] A. Montanari and R. Urbanke, "Coding for Network Coding," *Computing Research Repository (CoRR)*, vol. abs/0711.3935, Nov. 2007.

[42] R. Nóbrega and B. Uchôa-Filho, "Multishot Codes for Network Coding: Bounds and a Multilevel Construction," in *Proceedings of the 2009 IEEE International Symposium on Information Theory (ISIT'09)*, Seoul, South Korea, Jun. 2009.

[43] A. Rasala Lehman, "Network Coding," Ph.D. dissertation, Massachusetts Institute of Technology, Feb. 2005.

[44] P. Sanders, S. Egner, and L. Tolhuizen, "Polynomial Time Algorithms for Network Information Flow," in *Proceedings of the 15th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA'03)*, San Diego, California, Jun. 2003, pp. 286–294.

[45] D. Silva and F. Kschischang, "Using Rank-Metric Codes for Error Correction in Random Network Coding," in *Proceedings of the 2007 IEEE International Symposium on Information Theory (ISIT'07)*, Nice, France, Jun. 2007, pp. 796–800.

[46] ——, "Adversarial Error Correction for Network Coding: Models and Metrics," in *Proceedings of the 46st Annual Allerton Conference on Communication, Control, and Computing (Allerton'08)*, Monticello, Illinois, Sep. 2008, pp. 1246–1253.

[47] ——, "On Metrics for Error Correction in Network Coding," *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5479–5490, Dec. 2009.

[48] D. Silva, F. Kschischang, and R. Koetter, "A Rank-Metric Approach to Error Control in Random Network Coding," *IEEE Transactions on Information Theory*, vol. 54, no. 9, pp. 3951–3967, Sep. 2008.

[49] ——, "Capacity of Random Network Coding under a Probabilistic Error Model," in *Proceedings of the 24th Biennial Symposium on Communications*, Kingston, Canada, Jun. 2008, pp. 9–12.

[50] ——, "Capacity of Random Network Coding under a Probabilistic Error Model," *Computing Research Repository (CoRR)*, vol. abs/0807.1372, Jul. 2008, submitted to IEEE Transactions on Information Theory.

[51] V. Skachek, "Recursive Code Construction for Random Networks," *Computing Research Repository (CoRR)*, vol. abs/0806.3650v2, Jul. 2008.

[52] G. Ungerboeck, "Channel Coding with Multilevel/Phase Signals," *IEEE Transactions on Information Theory*, vol. 28, no. 1, pp. 55–67, Jan. 1982.

[53] S. Yang and R. Yeung, "Characterizations of Network Error Correction/Detection and Erasure Correction," in *Proceedings of the 3rd Workshop on Network Coding, Theory, and Applications (NetCod'07)*, San Diego, California, Jan. 2007.

[54] R. Yeung, *Information Theory and Network Coding*. Springer, 2008.

[55] R. Yeung and N. Cai, "Network Error Correction, Part I: Basic Concepts and Upper Bounds," *Communications in Information and Systems*, vol. 6, no. 1, pp. 19–36, 2006.

[56] R. Yeung, R. Li, N. Cai, and Z. Zhang, *Network Coding Theory*, ser. Foundations and Trends® in Communications and Information Theory. Now Publishers Inc, 2006.

[57] Z. Zhang, "Network Error Correction Coding in Packetized Networks," in *Proceedings of the 2006 IEEE Information Theory Workshop (ITW'06b)*, Chengdu, China, Oct. 2006, pp. 433–437.

[58] ——, "Linear Network Error Correction Codes in Packet Networks," *IEEE Transactions on Information Theory*, vol. 54, no. 1, pp. 209–218, Jan. 2008.

This page intentionally left blank.

Esta versão foi compilada em 3 de dezembro de 2009.
A última versão pode ser encontrada em
http://sites.google.com/site/rwnobrega/