Universidade Federal do Rio Grande do Sul

Instituto de Matemática e Estatística

Programa de Pós-Graduação em Matemática

Teoria da informação quântica: capacidade de canais e a conjectura da aditividade

Dissertação de Mestrado

Juliano Bianchi Tresoldi

Porto Alegre, Janeiro de 2026.

Dissertação submetida por Juliano Bianchi Tresoldi[1] como requisito parcial para a obtenção do grau de Mestre em Matemática pelo Programa de Pós-Graduação em Matemática do Instituto de Matemática e Estatística da Universidade Federal do Rio Grande do Sul.

Professor Orientador:

Prof. Dr. Carlos Felipe Lardizabal (PPGMAT/UFRGS)

Banca Examinadora:

Prof. Dr. Leonardo Guerini De Souza (UFSM)
Prof. Dr. Marcelo Terra Cunha (UNICAMP)
Prof. Dr. Artur Oscar Lopes (PPGMAT/UFRGS)
Prof. Dr. Newton Loebens (UFRGS)
Prof. Dr. Josué Knorst (UFRGS)

Data da Apresentação: 20 de Fevereiro de 2026.

[1]Bolsista da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES. Contato: juliano.bianchit@gmail.com.

Agradecimentos

Agradeço à minha mãe, Deise, meu pai, Dimi, e meus irmãos, Angelo, Théo e Davi. Agradeço igualmente à minha família adotiva de quatro patas.

Agradeço a todos do Programa de Pós-Graduação em Matemática Pura da UFRGS, especialmente a meu orientador, Carlos Felipe Lardizabal Rodrigues, e à CAPES, pelo auxílio financeiro.

Resumo

O tema central deste trabalho é a comunicação clássica por canais quânticos. Introduzimos os pré-requisitos mínimos necessários de teoria da informação e teoria quântica para formular e demonstrar o Teorema da Comunicação de Canais e o Teorema de Holevo-Schumacher-Westmoreland que caracterizam, no caso clássico e quântico, respectivamente, a capacidade de um canal. A capacidade de um canal clássico é o máximo das taxas de transmissão para as quais a probabilidade de erro tende assintoticamente à zero dados usos repetidos e independentes do canal. O Teorema da Codificação de Canais fornece uma expressão para a capacidade em termos da informação mútua entre entrada e saída de um único uso do canal. A capacidade clássica de um canal quântico é definida similarmente, e o Teorema de Holevo-Schumacher-Westmoreland descreve-o como um limite envolvendo a dita quantidade de Holevo de produtos tensoriais do canal. Se essa quantidade assintótica se reduz, como no caso clássico, a uma quantidade envolvendo um único uso do canal, permaneceu uma questão em aberto até recentemente, e deu origem à conjectura da aditividade. Apresentamos as diversas maneiras de como informação clássica pode ser transmitida por um canal quântico, cada uma com uma noção adequada de capacidade, e a relação entre elas. Vemos que, apesar de falsa no geral, a conjectura da aditividade é válida se restrita a certas classes de canais quânticos.

Palavras-chave: Canal Quântico. Capacidade Clássica. Teorema de Holevo-Schumacher-Westmoreland. Conjectura da Aditividade.

Abstract

The central theme of this work is the classical communication through quantum channels. We introduce the minimum necessary pre-requisites of information theory and quantum theory to formulate and prove the Channel Coding Theorem and the Holevo-Schumacher-Westmoreland Theorem which caracterize, in the classical and quantum cases respectively, the capacity of a channel. The capacity of a classical channel is the maximum rate of transmission for which the error probability asymptotically tend to zero given repeated and independent uses of the channel. The Channel Coding Theorem gives an expression to the capacity in terms of the mutual information between the input and output of a single use of the channel. The classical capacity of a quantum channel is defined similarly, and the Holevo-Schumacher-Westmoreland Theorem describes it as a limit involving the so called Holevo's quantity of the tensor products of the channel. Whether this quantity reduces itself, as in the classical case, to a quantity involving a single use of the channel, remained an open problem until recently, and gave origin to the additivity conjecture. We present the various ways in which classical information can be transmitted by a quantum channel, each one with an adequate notion of capacity, and the relation between them. We see how, despite false in general, the additivity conjecture holds if restricted to certain classes of quantum channel.

Keywords: Quantum Channel. Classical Capacity. Holevo-Schumacher-Westmoreland Theorem. Additivity Conjecture.

Sumário

Introdução **1**

1 Teoria da Informação Clássica **3**
1.1 Entropia de Shannon . . . 3
1.1.1 Definição e Exemplos . . . 3
1.1.2 Entropia Conjunta e Entropia Condicional . . . 7
1.1.3 Informação Mútua . . . 8
1.1.4 Entropia Relativa . . . 10
1.1.5 Propriedade da Equipartição Assintótica . . . 11
1.2 Canais Clássicos . . . 14
1.2.1 Definições e Exemplos . . . 14
1.2.2 Códigos de Canal . . . 15
1.2.3 Tipicalidade Condicional . . . 21
1.3 Teorema da Codificação de Canais . . . 23
1.3.1 Parte Direta . . . 24
1.3.2 Parte Recíproca . . . 27
1.4 Aditividade da Informação Mútua . . . 29

2 A Teoria Quântica **33**
2.1 Sistemas, Estados e Medições . . . 33
2.1.1 Sistemas quânticos . . . 33
2.1.2 Estados quânticos . . . 37
2.1.3 Evolução Unitária . . . 39
2.1.4 Medições . . . 39
2.2 Sistemas Compostos . . . 40
2.2.1 Produto Tensorial . . . 41
2.2.2 Decomposição de Schmidt . . . 43
2.2.3 Emaranhamento quântico . . . 44
2.2.4 Traço Parcial . . . 45
2.2.5 Medições Locais . . . 46
2.2.6 Purificação . . . 47
2.3 Canais Quânticos . . . 48
2.3.1 Definições e Exemplos . . . 48
2.3.2 Isomorfismo de Choi-Jamiołkowski . . . 49
2.3.3 Representação de Choi-Kraus de mapas completamente positivos . . . 50
2.3.4 Representação de Stinespring . . . 53
2.3.5 Relação entre as representações . . . 55
2.3.6 Canais *Entanglement-Breaking* . . . 56

3 Teoria da Informação Quântica **59**
3.1 Informação Clássica em Sistemas Quânticos 59
3.1.1 Estados Distinguíveis 59
3.1.2 Estados Clássico-Quânticos 60
3.1.3 Canais Clássico-Clássicos (C-C) 61
3.1.4 Canais Clássico-Quânticos (C-Q) 62
3.1.5 Canais Quântico-Clássicos (Q-C) 63
3.1.6 Caracterização de Canais *Entanglement-Breaking* 63
3.2 Entropia Quântica 64
3.2.1 Definição 64
3.2.2 Entropia Relativa Quântica 66
3.2.3 Informação Mútua Quântica 67
3.2.4 A Quantidade de Holevo 68
3.2.5 O Teorema de Holevo 68

4 Comunicação Clássica por Canais Quânticos **71**
4.1 Capacidade Clássica de Canais Quânticos 71
4.1.1 Esquemas de Codificação 76
4.1.2 Capacidade Clássica de Canais C-Q 80
4.1.3 Capacidade Clássica de Canais Quânticos 82
4.2 A Conjectura da Aditividade 84
4.2.1 Capacidade Clássica de Canais EB 84
4.2.2 Conjecturas Equivalentes 86
4.2.3 O Contraexemplo de Hastings 88

5 O Teorema de Holevo-Schumacher-Westmoreland **90**
5.1 Parte Recíproca 90
5.2 Tipicalidade Quântica 91
5.2.1 Subespaços Típicos 91
5.2.2 Tipicalidade Condicional Quântica 92
5.3 Parte Direta 94

Referências Bibliográficas **100**

Introdução

O tema central deste trabalho é a transmissão de informação clássica por meio de canais quânticos, tópico da teoria da informação quântica. O termo *informação* como termo matemático foi introduzido por Claude Shannon no artigo "A *Mathematical Theory of Communication*"[1], com o intuito de desenvolver uma teoria quantitativa de processamento de sinais, como ondas eletromagnéticas, nuvens de fumaça ou símbolos em uma página, dando início à teoria da informação. Para isso, Shannon percebeu que era necessário lidar com sinais em termos estatísticos, modelando uma fonte de sinais por um processo estocástico, e canais de comunicação como mapas entre distribuições de probabilidade. A informação contida em um sinal está associada a quanto o sinal pode ser comprimido, ou seja, o quanto de redundância pode ser eliminada do sinal sem que ele se torne irreconhecível. Os dois principais problemas tratados por Shannon foram o de quanto um sinal pode ser comprimido sem que haja perda de informação, e de quanta informação pode ser fielmente transmitida por um canal de comunicação ruidoso, sujeito a erros aleatórios. Para comprimir um sinal, deve-se eliminar redundâncias a fim de expressar a mesma informação com menos recursos. Já para transmitir um sinal por um canal ruidoso, deve-se introduzir redundâncias a fim de que o sinal não seja corrompido por erros, ideia por trás dos *códigos corretores de erros*.

A solução de Shannon para esses problemas se deu na forma dos dois teoremas fundamentais da teoria da informação: o Teorema da Compressão de Dados, que quantifica o máximo de compressão atingível para um sinal, a sua dita *entropia de Shannon*, e o Teorema da Codificação de Canais, que dá uma expressão para a *capacidade* de um canal, a maior taxa de transmissão de informação atingível através de códigos corretores de erros.

Com o advento da computação e criptografia quânticas, surgiu a necessidade adaptar a teoria da informação ao contexto dos fenômenos quânticos [2, 3], levando às noções de *informação quântica* e *canais quânticos*. A informação quântica difere daquela considerada por Shannon, à qual nos referimos por *informação clássica* em diversos aspectos. Por exemplo, a unidade elementar de informação clássica é o *bit*, que corresponde à informação contida em um sistema que se encontra em um de dois possíveis estados distinguíveis entre si. Já a unidade elementar de informação quântica é o *qubit*, que corresponde à informação contida em um sistema quântico que pode estar em um de dois estados distinguíveis, mas também em uma infinidade de estados intermediários, ditos *estados de superposição*.

Foi demonstrado em [4] o análogo quântico do Teorema da Compressão de Dados, o dito *Teorema da Compressão de Schumacher* (para uma demonstração, ver [5]). Já o Teorema da Codificação de Canais pode ser estendido para o contexto quântico de várias formas, já que canais quânticos podem transmitir tanto informação clássica como quântica. Neste trabalho, vamos estudar apenas a transmissão de informação clássica

por canais quânticos. Em particular, vamos demonstrar o célebre Teorema de Holevo-Schumacher-Westmoreland (HSW) [6, 7], que dá uma expressão para a maior taxa atingível de transmissão de informação clássica por um canal quântico, a dita *capacidade clássica* do canal quântico. Porém, ao contrário do Teorema da Codificação de Canais, a expressão para capacidade clássica dada por este teorema é computacionalmente intratável. Acreditava-se que esta expressão sempre se reduz à dita *quantidade de Holevo*, mais fácil de computar, hipótese que ficou conhecida como *conjectura da aditividade* e permaneceu um problema em aberto até 2009, quando Hastings demonstrou a existência de um contraexemplo [8].

O objetivo deste trabalho é introduzir o leitor à teoria da informação clássica e quântica, apresentando demonstrações do Teorema da Codificação de Canais e do Teorema de HSW, destacando as semelhanças entre elas.

O Capítulo 1, que segue principalmente [9, 10], serve como uma introdução à teoria da informação clássica. Na Seção 1.1, apresentamos as definições e resultados básicos da teoria. Na Seção 1.2 apresentamos os canais clássicos e os códigos de canais. Na Seção 1.3, é dada a demonstração do Teorema da Codificação de Canais. A Seção 1.4 oferece uma discussão sobre a aditividade da capacidade de canais, que se torna relevante no contexto da teoria da informação quântica.

No Capítulo 2 apresentamos os pré-requisitos da teoria quântica. Nas Seções 2.1 e 2.2, que seguem principalmente [11, 12, 10], introduzimos os aspectos mais elementares da teoria quântica, como a noções de estado, medição e emaranhamento. Leitores com experiência em teoria quântica podem se arriscar a pular essas seções. Na Seção 2.3, seguindo [12, 13], introduzimos os canais quânticos e suas diversas representações, em particular a importante classe de canais *entanglement-breaking*.

No Capítulo 3, a Seção 3.1 introduz à forma como informação clássica pode ser representada em sistemas quânticos, e os canais quânticos com entrada ou saída clássicas. Na Seção 3.2, apresentamos a entropia quântica e as quantidades análogas, e demonstramos o importante Teorema de Holevo.

No Capítulo 4, começamos apresentando na Seção 4.1 os diversos esquemas de codificação de canais quânticos para a transmissão de informação clássica, com suas noções apropriadas de capacidade, seguindo [13, 14]. Aqui enunciamos o Teorema de HSW e exploramos algumas de suas consequências. Na Seção 4.2, demonstramos a propriedade da aditividade para os canais *entanglement-breaking* e apresentamos as diversas formas da conjectura da aditividade, seguindo [15, 16, 17, 18, 19, 20]. Terminamos com uma breve discussão do contraexemplo de Hastings [8].

O Capítulo 5 se dedica a demonstração do Holevo-Schumacher-Westmoreland, adaptada de [12, 21].

Como referência geral para teoria da informação clássica, veja [9, 22, 23, 24, 25]. Para uma referência em teoria da informação quântica, veja [11, 10, 13, 26].

Capítulo 1

Teoria da Informação Clássica

Neste capítulo, vamos apresentar os conceitos básicos de teoria da informação com o objetivo de formular e demonstrar o célebre Teorema da Codificação de Canais. Na Seção 1.1, seguindo principalmente [9, 23], apresentamos o conceito de entropia de uma variável aleatória e demonstramos a Propriedade da Equipartição Assintótica, uma ferramenta teórica essencial para a análise do comportamento assintótico de sequências de variáveis aleatórias. Na Seção 1.2, apresentamos a noção de canal ruidoso, que é o principal formalismo matemático para o estudo da transmissão de informação, assim como a noção de capacidade de um canal. Na Seção 1.3, demonstramos o Teorema da Codificação de Canais, um dos principais resultados da teoria da informação clássica, seguindo principalmente [10, 21]. A demonstração é similar à demonstração do teorema análogo na teoria da informação quântica, o Teorema de HSW, que será demonstrado no Capítulo 5. Por fim, na Seção 1.4, discutimos a propriedade de aditividade da capacidade clássica, essencial na demonstração do Teorema da Codificação de Canais e cujo análogo na teoria da informação quântica foi alvo de uma das principais conjecturas da área.

1.1 Entropia de Shannon

No que segue, X é uma variável aleatória discreta tomando valores $x \in \mathcal{X}$ com distribuição p_X, onde $\mathcal{X}$ é um conjunto finito de cardinalidade $|\mathcal{X}|$. Similarmente Y toma valores $y \in \mathcal{Y}$ com distribuição p_Y, e assim por diante.

1.1.1 Definição e Exemplos

Definição 1.1.1 (Conteúdo Informacional)**.** Seja X tal que $p_X(x) > 0$ para todo $x \in \mathcal{X}$. Chamamos de **conteúdo informacional** de x a quantidade

$$i(x) = \log\left(\frac{1}{p_X(x)}\right),$$

onde o logaritmo é em base 2 e sua unidade de medida é o bit.

Este valor quantifica a surpresa ao obter o valor x em uma realização de X, ou em outras palavras, a quantidade de informação nova sobre X que obtemos pela realização com resultado x. Podemos estender a i como uma função de eventos do espaço de medida $A = (\mathcal{X}, p_X)$.

O conteúdo informacional é definido de forma a satisfazer as seguintes propriedades:

1. Se X é determinística, ou seja, existe $x_0 \in \mathcal{X}$ com $p_X(x_0) = 1$, então $i(x) = 0$. Uma realização de X não nos fornece informação nova.

2. i é decrescente com a probabilidade. Resultados menos prováveis fornecem mais informação.

3. $i(U \cap V) = i(U) + i(V)$ para $U, V \subset \mathcal{X}$ independentes.

$i(X)$ é em si uma variável aleatória. A **entropia de Shannon** ou **entropia** de X é o valor esperado de $i(X)$ com respeito à p_X:

$$H(X) = \sum_{x \in \mathcal{X}} p_X(x) \log\left(\frac{1}{p_X(x)}\right) = -\sum_{x \in \mathcal{X}} p_X(x) \log(p_X(x)).$$

Expandimos essa definição para incluir quaisquer distribuições de probabilidade (não somente estritamente positivas) pela convenção:

$$0 \cdot \log(0) = 0,$$

que é motivada pelo limite $\lim_{\epsilon \to 0} \epsilon \log(1/\epsilon) = 0$.

A definição acima sugere que $H(X)$ seja interpretada como a quantidade de informação que esperamos adquirir com uma observação de X, ou ainda, como uma medida de incerteza do valor de X antes de observá-la.

Note que a entropia não depende explicitamente de X, apenas de sua distribuição. Dado um conjunto finito $\mathcal{X}$, seja $\mathcal{P}(\mathcal{X})$ o conjunto de medidas de probabilidade sobre $\mathcal{X}$. Seus elementos são funções do tipo $p : \mathcal{X} \to [0, 1]$ tas que $\sum_{x \in \mathcal{X}} p(x) = 1$. Fixada uma enumeração de $\mathcal{X}$, seja $\mathbb{R}^{\mathcal{X}} := \mathbb{R}^{|\mathcal{X}|}$. Temos que $\mathcal{P}(\mathcal{X})$ é um conjunto compacto convexo de $\mathbb{R}^{\mathcal{X}}$. Podemos então definir a entropia como um mapa $H : \mathcal{P}(\mathcal{X}) \to \mathbb{R}$:

Definição 1.1.2 (Entropia)**.** A **entropia** de uma distribuição de probabilidade $p \in \mathcal{P}(\mathcal{X})$ é dada por:

$$H(p) = -\sum_{x \in \mathcal{X}} p(x) \log p(x),$$

com a convenção de que $0 \cdot \log(0) = 0$.

Chamamos de **distribuição de Dirac** ou delta de Dirac a distribuição de probabilidade $\delta_x \in \mathcal{P}(\mathcal{X})$ definida por:

$$\delta_x(y) = \begin{cases} 1, & \text{se} \quad y = x \\ 0, & \text{se} \quad y \neq x \end{cases},$$

que é a distribuição de uma variável aleatória determinística $X = x$. O próximo teorema confirma a intuição de que tais funções são exatamente aquelas que não carregam incerteza quanto a seus valores.

Teorema 1.1.3. *Para todo $p \in \mathcal{P}(\mathcal{X})$, $H(p) \geq 0$. $H(p) = 0$ se e somente se $p = \delta_x$ para algum $x \in \mathcal{X}$.*

Demonstração. Para cada $x \in \mathcal{X}$, se $0 \leq p(x) \leq 1$, então $-\log p(x) \geq 0$, portanto $H(p) \geq 0$. Mas então:

$$H(p) = -\sum_{x\in\mathcal{X}} p(x)\log p(x) = 0 \iff \forall x \in \mathcal{X},\ p(x) = 0 \text{ ou } p(x) = 1.$$

Claro que $p(x) = 1$ pode ocorrer somente para um elemento x. Neste caso $p = \delta_x$. □

Dado um conjunto convexo A de um espaço vetorial $\mathcal{H}$, denotamos por $\mathrm{Ext}(A)$ o conjunto de elementos extremos de A, ou seja, dos elementos de A que não são dados como combinação convexa de outros elementos em A.

Teorema 1.1.4. *Para $\mathcal{X}_1$ e $\mathcal{X}_2$ conjuntos finitos:*

$$\mathrm{Ext}(\mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)) = \mathrm{Ext}(\mathcal{P}(\mathcal{X}_1)) \times \mathrm{Ext}(\mathcal{P}(\mathcal{X}_2)).$$

Demonstração. As distribuições extremais sobre um conjunto finito são as distribuições de Dirac. Logo, os elementos de $\mathrm{Ext}(\mathcal{P}(\mathcal{X}_1))\times\mathrm{Ext}(\mathcal{P}(\mathcal{X}_2))$ são do tipo (δ_x, δ_y) com $x \in \mathcal{X}_1, y \in \mathcal{X}_2$, que são em si distribuições de Dirac $\delta_{x,y} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$, portanto $\mathrm{Ext}(\mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)) \supseteq \mathrm{Ext}(\mathcal{P}(\mathcal{X}_1)) \times \mathrm{Ext}(\mathcal{P}(\mathcal{X}_2))$. Por outro lado, $\delta_{x,y}(x', y') = \delta_x(x')\delta_y(y') = (\delta_x, \delta_y)(x', y')$, logo $\mathrm{Ext}(\mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)) \subseteq \mathrm{Ext}(\mathcal{P}(\mathcal{X}_1)) \times \mathrm{Ext}(\mathcal{P}(\mathcal{X}_2))$. □

Exemplo 1.1.5 (Entropia Binária)**.** A distribuição de Bernoulli $\mu = (p, 1-p)$ tem entropia dada por:

$$H(\mu) = -p\log p - (1-p)\log(1-p).$$

Definimos a **entropia binária** como a função real $H_{bin}(p) = -p\log p - (1-p)\log(1-p)$. Esta é uma função contínua positiva, côncava, com valor máximo $\log 2 = 1$ em $p = 1/2$, que corresponde a distribuição uniforme. Vamos ver que essas são propriedades geralmente satisfeitas pela entropia.

Vejamos que a entropia é uma função côncava em $\mathcal{P}(\mathcal{X})$:

Teorema 1.1.6. *O mapa $H : \mathcal{P}(\mathcal{X}) \to \mathbb{R}$ é contínuo e para $1 \leq i \leq k$, $\lambda_i > 0$ tais que $\sum_i \lambda_i = 1$, e $p_i \in \mathcal{P}(\mathcal{X})$, temos:*

$$H\left(\sum_{i=1}^{k} \lambda_i p_i\right) \geq \sum_{i=1}^{k} \lambda_i H(p_i),$$

com igualdade se, e somente se, $p_1 = p_2 = \cdots = p_k$.

Demonstração. Considere a função:

$$\varphi(t) = \begin{cases} -t\ln t & \text{se} \quad 0 < t \leq 1 \\ 0 & \text{se} \quad t = 0 \end{cases},$$

restrita ao intervalo $[0, 1]$. Como $\varphi' = -(1 + \ln t)$ e $\varphi'' = -t^{-1}$, segue que φ é contínua e estritamente côncava. Podemos escrever:

$$H(p) = \frac{1}{\ln 2}\sum_{x\in\mathcal{X}} \varphi(p(x)).$$

A continuidade de H segue da continuidade de φ. Já pela concavidade estrita de φ, temos que para cada $x \in \mathcal{X}$:

$$\varphi\left(\sum_{i=1}^{k} \lambda_i p_i(x)\right) \geq \sum_{i=1}^{k} \lambda_i \varphi(p_i(x)),$$

com igualdade se e somente se $p_i(x)$ não depender de i. Somando essas desigualdades para cada $x \in \mathcal{X}$ e multiplicando por $-\ln 2$ nos dá o resultado. □

Intuitivamente, o teorema anterior mostra que uma mistura estatística de distribuições pode criar mais incerteza.

Vejamos agora uma desigualdade fundamental em teoria da informação.

Teorema 1.1.7 (Desigualdade de Gibbs)**.** *Para quaisquer $p, q \in \mathcal{P}(\mathcal{X})$:*

$$-\sum_{x\in\mathcal{X}} p(x) \log q(x) \geq -\sum_{x\in\mathcal{X}} p(x) \log p(x),$$

com igualdade se e somente se $p(x) = q(x)$ para todo $x \in \mathcal{X}$.

Demonstração. Seja $I = \{x \in \mathcal{X} \,|\, p(x) \neq 0\}$. Tomando derivadas, é fácil verificar que, para todo $y > 0$, $\ln y \leq y - 1$. Assim, temos:

$$-\sum_{x\in I} p(x) \ln \frac{q(x)}{p(x)} \geq -\sum_{x\in I} p(x) \left(\frac{q(x)}{p(x)} - 1\right) \geq -\sum_{x\in I} q(x) + \sum_{x\in I} p(x) \geq 0,$$

pois $\sum_{x\in I} p(x) = 1$ e $\sum_{x\in I} q(x) \leq 1$. A desigualdade segue das propriedades do logaritmo e de $\ln y = \log y \ln 2$:

$$-\sum_{x\in I} p(x) \log q(x) \geq -\sum_{x\in I} p(x) \log p(x).$$

Os somatórios podem ser estendidos para todo $x \in \mathcal{X}$, já que os termos se anulam caso $x \notin I$. O caso de igualdade implica que para todo $x \in I$, $p(x) \ln \frac{q(x)}{p(x)} = 0$. Como $p(x) > 0$, isso só é possível caso $p(x) = q(x)$ para $x \in I$. Também neste caso temos que $\sum_{x\in I} q(x) = 1$, logo $q(x) = p(x) = 0$ caso $x \notin I$.

□

Teorema 1.1.8. *Para qualquer $p \in \mathcal{P}(\mathcal{X})$:*

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

com igualdade se e somente se p é a distribuição uniforme em $\mathcal{X}$.

Demonstração. Pela desigualdade de Gibbs, com q sendo a distribuição uniforme:

$$H(p) = -\sum_{x\in\mathcal{X}} p(x) \log p(x) \leq -\sum_{x\in\mathcal{X}} p(x) \log \frac{1}{|\mathcal{X}|} = \log(|\mathcal{X}|).$$

O critério de igualdade para a desigualdade de Gibbs implica que $H(p) = \log |\mathcal{X}|$ se e somente se p é uniforme. □

Ver [10] para uma demonstração alternativa do Teorema 1.1.8 utilizando multiplicadores de Lagrange.

1.1.2 Entropia Conjunta e Entropia Condicional

Vamos estender o conceito de entropia para várias variáveis aleatórias.

Definição 1.1.9 (Entropia Conjunta)**.** A **entropia conjunta** de X e Y é dada por:

$$H(X,Y) = -\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} p_{X,Y}(x,y)\log(p_{X,Y}(x,y)).$$

Chamamos de **informação condicional** de y dado x a quantidade $i(y|x) = -\log(p_{Y|X}(y|x))$, onde $x \in \mathcal{X}$ e $y \in \mathcal{Y}$. Temos então a entropia de Y condicionada pela realização $X = x$, dada por:

$$H(Y|X=x) = \mathbb{E}_{Y|X=x}\{i(Y|x)\} = -\sum_{y\in\mathcal{Y}} p_{Y|X}(y|x)\log(p_{Y|X}(y|x)). \tag{1.1}$$

Logo, $H(Y|X=x)$ é a incerteza que permanece sobre Y após observar o evento $\{X=x\}$.

Definição 1.1.10 (Entropia Condicional)**.** A **entropia condicional** de Y dado X é:

$$\begin{aligned} H(Y|X) &= \mathbb{E}_X\{H(Y|X=x)\} \\ &= \mathbb{E}_{X,Y}\{i(Y|X)\} \\ &= -\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} p_{X,Y}(x,y)\log(p_{Y|X}(y|x)) \end{aligned} \tag{1.2}$$

Proposição 1.1.11. *Para qualquer variável aleatória X, $H(X|X) = 0$.*

Interpretamos $H(Y|X)$ como a quantidade média de incerteza que permanece sobre Y tendo observado X. De fato, isso se traduz formalmente na relação $H(Y|X) = H(X,Y) - H(X)$. Vamos demonstrar uma equação equivalente cujo formato nos lembra a regra da cadeia para diferenciação.

Teorema 1.1.12 (Regra da cadeia)**.** $H(X,Y) = H(X) + H(Y|X)$.

Demonstração.

$$\begin{aligned} H(X,Y) &= -\sum_{x,y} p_{X,Y}(x,y)\log(p_{X,Y}(x,y)) \\ &= -\sum_{x,y} p_{X,Y}(x,y)\log(p_X(x)p_{Y|X}(y|x)) \\ &= -\sum_{x,y} p_{X,Y}(x,y)(\log p_X(x) + \log p_{Y|X}(y|x)) \\ &= -\sum_{x,y} p_{X,Y}(x,y)\log p_X(x) - \sum_{x,y} p_{X,Y}(x,y)\log p_{Y|X}(y|x) \\ &= -\sum_{y} p_X(x)\log p_X(x) - \sum_{x,y} p_{X,Y}(x,y)\log p_{Y|X}(y|x) \\ &= H(X) + H(Y|X) \end{aligned}$$

□

Note o uso da Regra de Bayes $p_{X,Y} = p_{X|Y} \cdot p_Y = p_{Y|X} \cdot p_X$. Assim, trocando X e Y acima nos dá $H(Y) + H(X|Y) = H(X) + H(Y|X)$. Pode-se facilmente mostrar por indução uma generalização do resultado anterior para várias variáveis aleatórias:

Corolário 1.1.13 (Regra da Cadeia Generalizada [9]). *Sejam $X_1, X_2, \ldots, X_n$ e Y variáveis aleatórias. Então:*

$$H(X_1, X_2, \ldots, X_n|Y) = \sum_{i=1}^{n} H(X_i|Y, X_1, X_2 \ldots, X_{i-1})$$

Com $X_0 = \emptyset$. Em particular, para $n = 2$:

$$H(X_1, X_2|Y) = H(X_1|Y) + H(X_2|Y, X_1)$$

1.1.3 Informação Mútua

Vejamos agora uma noção de correlação entre duas variáveis aleatórias, que mede a quantidade de informação compartilhada entre elas.

Definição 1.1.14 (Informação Mútua). Chamamos de **informação mútua** entre X e Y a quantidade:

$$I(X : Y) = H(Y) - H(Y|X).$$

Para $p_X \in \mathcal{P}(\mathcal{X})$ e $p_Y \in \mathcal{P}(\mathcal{Y})$, definimos:

$$I(p_X : p_Y) = I(X : Y),$$

com X e Y quaisquer variáveis aleatórias tais que $X \sim p_X$ e $Y \sim p_Y$.

A informação mútua é simétrica, já que

$$I(X : Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = I(Y : X),$$

e quantifica a informação em comum entre as variáveis. Segue facilmente as seguintes equações:

$$I(X : X) = H(X),$$

$$I(X : Y) = H(X) + H(Y) - H(X, Y),$$

e

$$I(X : Y) \leq \min\{H(X), H(Y)\}. \tag{1.3}$$

Teorema 1.1.15. *Em termos das distribuições de X e Y temos:*

$$I(X : Y) = \sum_{x,y} p_{X,Y}(x, y) \log \left(\frac{p_{X,Y}(x, y)}{p_X(x) p_Y(y)} \right).$$

Demonstração.

$$\begin{aligned} I(X:Y) = H(X) - H(X|Y) &= -\sum_{x,y} p_{X,Y}(x,y)\log p_X(x) + \sum_{x,y} p_{X,Y}(x,y)\log(p_{X|Y}(x|y)) \\ &= \sum_{x,y} p_{X,Y}(x,y)\log\left(\frac{p_{X|Y}(x|y)}{p_X(x)}\right) \\ &= \sum_{x,y} p_{X,Y}(x,y)\log\left(\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\right). \end{aligned}$$

□

Teorema 1.1.16. *$I(X:Y) \geq 0$, com igualdade se e somente se X e Y são independentes.*

Demonstração. Segue da desigualdade de Gibbs para as distribuições $p_{X,Y}$ e $p_X \cdot p_Y$ sobre $\mathcal{X} \times \mathcal{Y}$:

$$\begin{aligned} I(X:Y) &= \sum_{x,y} p_{X,Y}(x,y)\log\left(\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\right) \\ &= \sum_{x,y} p_{X,Y}(x,y)\log p_{X,Y}(x,y) - \sum_{x,y} p_{X,Y}(x,y)\log(p_X(x)p_Y(y)) \\ &\geq 0, \end{aligned}$$

com igualdade se, e somente se, $p_{X,Y} = p_X \cdot p_Y$, ou seja, se X e Y são independentes. □

Teorema 1.1.17 (Sub-Aditividade da Entropia)**.** *$H(X,Y) \leq H(X) + H(Y)$, com igualdade se e somente se X e Y forem independentes.*

Demonstração. Segue do Teorema 1.1.16 e de $I(X,Y) = H(X) + H(Y) - H(X,Y) \geq 0$. □

Intuitivamente, esperamos que adquirir informação sobre Y não aumente nossa incerteza sobre X, e que caso X e Y forem independentes, informação sobre Y não diminua nossa incerteza sobre X.

Teorema 1.1.18. *$H(X|Y) \leq H(X)$, com igualdade se e somente se X e Y forem independentes.*

Demonstração. Segue de do Teorema 1.1.17 e da relação $I(X:Y) = H(X) - H(X|Y)$. □

Definição 1.1.19 (Informação Mútua Condicional)**.** Chamamos de informação mútua condicional entre X e Y dado Z a quantidade:

$$I(X:Y|Z) = H(X|Z) - H(X|Y,Z).$$

Se $I(X:Y|Z) = 0$, dizemos que X é condicionalmente independente de Y dado Z, e vise-versa.

Proposição 1.1.20 (Regra da Cadeia para a Informação Mútua [9])**.** *Sempre temos que:*

$$I(X:Y,Z) = I(X:Z) + I(X:Y|Z) = I(X:Y) + I(X:Z|Y).$$

Proposição 1.1.21 (Sub-Aditividade Forte da Entropia de Shannon [27]). *Sempre temos que:*

$$H(X,Y,Z) + H(Z) \leq H(X,Z) + H(Y,Z),$$

com igualdade se e somente se $I(X:Y|Z) = 0$. *Equivalentemente* $I(X:Y|Z) \geq 0$.

Corolário 1.1.22 (Monotonicidade da Informação Mútua). $I(X:Y,Z) \geq I(X:Y)$. *Intuitivamente, aprendemos pelo menos tanto sobre* X *observando* Y *e* Z *quanto observando apenas* Y.

Demonstração. Imediato da Proposição 1.1.20 e de que $I(X:Y|Z) \geq 0$ pela Proposição 1.1.21. □

Dizemos que variáveis aleatórias X, Y e Z formam uma **Cadeia de Markov** (denotamos isso por $X \to Y \to Z$) caso Y depender apenas de X e Z depender apenas de Y, ou equivalentemente, caso a probabilidade conjunta é dada por:

$$p_{X,Y,Z}(x,y,z) = p_X(x) \cdot p_{Y|X}(y|x) \cdot p_{Z|Y}(z|y).$$

Proposição 1.1.23 ([9]). $I(X:Z|Y) = 0$ *se e somente se* $X \to Y \to Z$.

Teorema 1.1.24 (Desigualdade do Processamento de Dados). *Se* $X \to Y \to Z$, *então* $I(X:Y) \geq I(X:Z)$.

Demonstração. Segue da Proposição 1.1.20 que:

$$I(X:Z) + I(X:Y|Z) = I(X:Y) + I(X:Z|Y),$$

já que $I(X:Y|Z) \geq 0$ pelo Teorema 1.1.21, e $I(X:Z|Y) = 0$ pela Proposição 1.1.23. □

1.1.4 Entropia Relativa

A teoria da informação fornece uma maneira de comparar distribuições de probabilidade através de suas entropias, uma ferramenta de grande importância em estatística (ver, por exemplo, [28]).

Definição 1.1.25 (Entropia Relativa). Dadas $p, q \in \mathcal{P}(\mathcal{X})$, a **entropia relativa** ou **KL-divergência** (divergência de Kullback-Leibler) de p para q é dada por:

$$D(p||q) = \sum_{x \in \mathcal{X}} p(x) \log \left(\frac{p(x)}{q(x)} \right),$$

desde que $q(x) \neq 0$ para todo $x \in \mathcal{X}$, caso contrário definimos $D(p||q) = +\infty$.

Note que $D(p||q) = 0$ se, e somente se $\forall x \in \mathcal{X}, p(x) \neq 0$ que implica em $p(x) = q(x)$. A entropia relativa não define formalmente uma métrica, pois não é simétrica (por isso dizemos entropia relativa de p para q, e não entre p e q). Apesar disso, a entropia relativa fornece uma noção de distância que permite o uso de métodos geométricos em estatística [29].

Quando p e q forem as distribuições de variáveis aleatórias X e Y, escrevemos $D(X||Y) = D(p||q)$.

Vamos ver a seguir como as diversas quantidades que definimos na seção anterior podem ser expressas em termos da entropia relativa de distribuições adequadas, novamente tomando distribuições $p, q \in \mathcal{P}(\mathcal{X})$. Primeiro uma proposição fundamental que já demonstramos e utilizamos várias vezes:

Teorema 1.1.26. *$D(p||q) \geq 0$, com igualdade se, e somente se, $p = q$.*

Demonstração. $D(p||q) \geq 0$ nada mais é do que a desigualdade de Gibbs! □

Seja $\mathcal{U}_\mathcal{X} \in \mathcal{P}(\mathcal{X})$ a distribuição uniforme em $\mathcal{X}$.

Teorema 1.1.27. *$H(p) = \log|\mathcal{X}| - D(p\,||\,\mathcal{U}_\mathcal{X})$.*

Demonstração.

$$\begin{aligned} D(p||\mathcal{U}_\mathcal{X}) &= \sum_{x\in\mathcal{X}} p(x)\log\left(\frac{p(x)}{\mathcal{U}_\mathcal{X}}\right) \\ &= \sum_{x\in\mathcal{X}} p(x)\log p(x) + \log|\mathcal{X}| \\ &= -H(p) + \log|\mathcal{X}|\,. \end{aligned}$$

□

Isso confirma a intuição de que a entropia de uma distribuição é uma medida de quanto ela se assemelha à distribuição uniforme.

Teorema 1.1.28. *$I(X:Y) = D(p_{X,Y}\,||\,p_X \cdot p_Y)$.*

Demonstração. Imediato do Teorema 1.1.15. □

Isso mostra como a informação mútua entre duas distribuições mede o quão longe elas estão de serem independentes.

1.1.5 Propriedade da Equipartição Assintótica

Vejamos agora como a entropia surge na descrição da estatística de sequências de variáveis aleatórias i.i.d. (independentes e identicamente distribuídas).

Considere $X^n = (X_1, X_2, \ldots, X_n)$ uma sequência i.i.d. de distribuição p, ou seja, se $x^n = (x_1, \ldots, , x_n) \in \mathcal{X}^n$ então $p_{X^n}(x^n) = p(x_1)p(x_2)\ldots p(x_n)$. O que podemos dizer sobre a distribuição de tais sequências? Para n grande o suficiente, esperamos que x_i apareça em torno de $np(x_i)$ vezes. Considere uma sequência típica que contém cada símbolo x_i aproximadamente $np(x_i)$ vezes:

$$p_{X^n}(x^n) \approx p(x_1)^{np(x_1)}p(x_2)^{np(x_2)}\cdots p(x_n)^{np(x_n)}.$$

Isso implica em:

$$-\frac{1}{n}\log p_{X^n}(x^n) \approx -\sum_{i=1}^{k} p(x_i)\log p(x_i) = H(X),$$

ou seja, a probabilidade de obter tal sequência é $p_{X^n}(x^n) \approx 2^{-nH(X)}$. Sequências de tamanho cada vez maior tendem a seguir uma distribuição uniforme, caso nos restringirmos apenas a tais sequências típicas, e sua probabilidade está relacionada à entropia de X. Vamos ver que tal restrição é razoável para n suficientemente grande. Para isso vamos usar uma das versões da Lei dos Grandes Números:

Proposição 1.1.29 (Lei Fraca dos Grandes Números [30]). *Seja $(Y_i)_{i\in\mathbb{N}}$ uma sequência de variáveis aleatórias i.i.d. com valor esperado $\mathbb{E}(Y)$ finito partindo de um espaço de probabilidade $(\Omega, \mathcal{X}, P)$ e para cada n seja $\overline{Y_n} = \frac{1}{n}\sum_{i=0}^{n} Y_i$, dita n-ésima média amostral. Então temos que $\overline{Y}_n \to \mathbb{E}(Y)$ em probabilidade, ou seja, para todo $\epsilon > 0$:*

$$\lim_{n\to\infty} P\left(|\overline{Y_n} - \mathbb{E}(Y)| < \epsilon\right) = 1$$

Definição 1.1.30 (Entropia Amostral). A entropia amostral $\overline{H}(x^n)$ de uma sequência $x^n \in \mathcal{X}^n$ com respeito à distribuição $p_X \in \mathcal{P}(\mathcal{X})$ é o valor

$$\overline{H}(x^n) = -\frac{1}{n}\log(p_{X^n}(x^n)),$$

onde $p_{X^n}(x^n) = \prod_{i=1}^{n} p_X(x_i)$.

Teorema 1.1.31 (Propriedade da Equipartição Assintótica). *$\overline{H}(X^n) \to H(X)$ em probabilidade.*

Demonstração.

$$\begin{aligned}
\overline{H}(X^n) &= -\frac{1}{n}\log p_{X^n}(X_1, \ldots, X_n) \\
&= -\frac{1}{n}\sum_{i=1}^{n}\log p(X) \\
&\to -\mathbb{E}(\log p(X)) \qquad \text{(em probabilidade)} \\
&= H(X).
\end{aligned}$$

□

O nome Propriedade da Equipartição Assintótica (PEA) merece explicação. Equipartição é um termo que se refere a uma distribuição uniforme e, como veremos, esse teorema implica que com $n \to \infty$ (assintoticamente) a distribuição de sequências se acumula em um conjunto relativamente pequeno, onde a distribuição é bastante uniforme (equipartição).

Definição 1.1.32 (Conjunto Típico). Chamamos $x^n \in \mathcal{X}^n$ de **sequência δ-típica** em relação a X se:

$$\left|\overline{H}(x^n) - H(X)\right| < \delta.$$

O **conjunto δ-típico** em relação a X é o conjunto de sequências δ-típicas em relação a X:

$$T_\delta^{X^n} = \{x^n \mid \left|\overline{H}(x^n) - H(X)\right| < \delta\}.$$

Quando o contexto for claro, chamamos conjuntos da forma $T_\delta^{X^n}$ de conjuntos típicos.

Os conjuntos típicos tendem a ser menores que o conjunto de sequências possíveis, porém acumulam a maior densidade de probabilidade (Figura 1.1.5).

Teorema 1.1.33 (Equipartição). *A probabilidade p_{X^n} é aproximadamente uniforme em T_δ^n. Para qualquer $x^n \in T_\delta^{X^n}$:*

$$2^{-n(H(X)+\delta)} \leq p_{X^n}(x^n) \leq 2^{-n(H(X)-\delta)}.$$

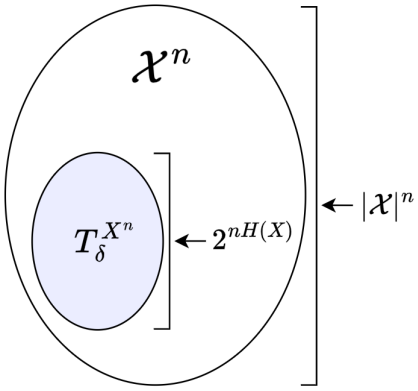

Figura 1.1: Conjunto δ-típico em relação a X. Apesar de menor que o conjunto de possíveis sequências por um fator exponencial, tende a acumular a medida de probabilidade.

Demonstração. Segue de $H(X) - \delta \leq \overline{H}(x^n) \leq H(X) + \delta$ e $\overline{H}(x^n) = -\frac{1}{n} \log p_{X^n}(x^n)$. □

Teorema 1.1.34 (Acúmulo de Probabilidade)**.** *Para todo $\epsilon \in (0,1)$ e $\delta > 0$, existe $n \in \mathbb{N}$ tal que $p_{X^n}\{T_\delta^{X^n}\} \geq 1 - \epsilon$.*

Demonstração. Pela PEA temos que $\overline{H}(X^n) \to H(X)$ em probabilidade, ou seja:

$$\lim_{n\to\infty} p_{X^n}\left\{|\overline{H}(X^n) - H(X)| < \delta\right\} = 1.$$

Logo, existe n suficientemente grande tal que:

$$p_{X^n}\{T_\delta^{X^n}\} = p_{X^n}\left\{|\overline{H}(X^n) - H(X)| < \delta\right\} \geq 1 - \epsilon.$$

□

Teorema 1.1.35 (Cardinalidade)**.** *Para todo $\epsilon \in (0,1)$, existe $n \in \mathbb{N}$ tal que:*

$$(1-\epsilon)2^{n(H(X)-\delta)} \leq \left|T_\delta^{X^n}\right| \leq 2^{n(H(X)+\delta)}.$$

Demonstração. Primeiro vejamos a desigualdade à direita:

$$\begin{aligned} 1 &= \sum_{x^n \in \mathcal{X}^n} p_{X^n}(x^n) \\ &\geq \sum_{x^n \in T_\delta^{X^n}} p_{X^n}(x^n) \\ &\geq \sum_{x^n \in T_\delta^{X^n}} 2^{-n(H(X)+\delta)} \\ &= 2^{-n(H(X)+\delta)} \cdot \left|T_\delta^{X^n}\right|. \end{aligned}$$

Por outro lado, pelo Teorema 1.1.34:

$$\begin{aligned} 1 - \epsilon &\leq p_{X^n}(T_\delta^{X^n}) \\ &\leq \sum_{x^n \in T_\delta^{X^n}} 2^{-n(H(X)-\delta)} \\ &= 2^{-n(H(X)-\delta)} \cdot \left|T_\delta^{X^n}\right|. \end{aligned}$$

□

Note finalmente que, para n suficientemente grande, os conjuntos típicos são menores que o conjunto $\mathcal{X}^n$ de todas as sequências na razão de $2^{nH(X)}$. No caso de entropia máxima $H(X) = \log|\mathcal{X}|$:

$$\left|T_\delta^{X^n}\right| \leq 2^{n(H(X)+\delta)} \approx |\mathcal{X}|^n .$$

1.2 Canais Clássicos

Vamos agora nos dedicar ao estudo da transmissão de informação. Para isso vamos formalizar o conceito de canal ruidoso, que modela um meio de transmissão de informação sujeito a erros aleatórios.

1.2.1 Definições e Exemplos

Definição 1.2.1. Um **canal clássico** é dado por conjuntos $\mathcal{X}$, dito o **alfabeto de entrada**, e $\mathcal{Y}$, dito o **alfabeto de saída**, e um mapa $N : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, dito o **mapa de transição do canal** tal que, para cada $x \in \mathcal{X}$, temos uma distribuição de probabilidade $N(x) = p(\cdot|x) \in \mathcal{P}(\mathcal{Y})$ dita a **probabilidade de transição**. Denotamos este canal por $(\mathcal{X}, N, \mathcal{Y})$ ou $(\mathcal{X}, N \equiv p(y|x), \mathcal{Y})$. Se a emissão de um símbolo de $\mathcal{X}$ é dada pela variável aleatória X e a saída correspondente é dada por $Y \in \mathcal{Y}$, então:

$$p(y|x) = p_{Y|X}(y|x)$$

é a probabilidade da saída ser y dado que a entrada foi x.

Neste capítulo, vamos nos referir a canais clássicos simplesmente por canais.

Dada uma distribuição $p \in \mathcal{P}(\mathcal{X})$, temos, através do canal, uma distribuição $N(p) \in \mathcal{P}(\mathcal{Y})$ associada, dada por:

$$N(p)(y) = \sum_{x\in\mathcal{X}} p(x)p(y|x) \tag{1.4}$$

e uma distribuição conjunta $\omega \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ que satisfaz $\omega(x, y) = p(x)p(y|x)$.

Assim, um canal $(\mathcal{X}, N, \mathcal{Y})$ define um mapa $N : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{Y})$, com $N(x) = N(\delta_x)$, ou equivalentemente, uma matriz estocástica $|\mathcal{X}| \times |\mathcal{Y}|$:

$$N = \begin{bmatrix} p(y_1|x_1) & p(y_2|x_1) & \cdots & p(y_{|\mathcal{Y}|}|x_1) \\ p(y_1|x_2) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ p(y_1|x_{|\mathcal{X}|}) & \cdots & \cdots & p(y_{|\mathcal{Y}|}|x_{|\mathcal{X}|}) \end{bmatrix}$$

de forma que a entrada do canal p_X é um vetor de probabilidade sobre $\mathcal{X}$ e $p_X^T N = p_Y$ é o vetor de probabilidade sobre $\mathcal{Y}$ que representa a saída do canal.

Exemplo 1.2.2 (Canal Ideal)**.** Um canal $(\mathcal{X}, N, \mathcal{Y})$ é dito um **canal ideal** se N for determinística, ou seja, se para cada $x \in \mathcal{X}$, existe $y \in \mathcal{Y}$ tal que $N(\delta_x) = \delta_y$. A matriz estocástica associada a um canal ideal tem entradas igual a 0 ou 1.

Exemplo 1.2.3 (Canal Binário Simétrico)**.** Chamamos de **canal binário simétrico com ruído** α o canal $\mathcal{B}_\alpha : \{0,1\} \to \{0,1\}$, tal que:

$$p(0|0) = p(1|1) = 1 - \alpha$$
$$p(0|1) = p(1|0) = \alpha$$

Se $\alpha = 0$ ou $\alpha = 1$, então $\mathcal{B}_\alpha$ é um canal ideal, caso contrário há uma chance de erro. Se $\alpha = \frac{1}{2}$, então o canal é completamente ruidoso, e nenhuma informação pode ser transmitida por ele.

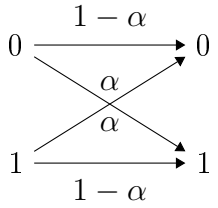

Figura 1.2: Canal binário simétrico.

Note que para uma entrada X e saída $Y = \mathcal{B}_\alpha(X)$, pela Definição 1.1.10 temos:

$$\begin{aligned} H(Y|X) &= -\sum_{x=0,1} p_X(x) \sum_{y=0,1} p_{Y|X}(y|x) \log(p_{Y|X}(y|x)) \\ &= H_{bin}(\alpha). \end{aligned}$$

É possível ver que, se $X \sim \mathcal{U}_2$, então $Y \sim \mathcal{U}_2$.

A proposição a seguir mostra que canais podem somente introduzir ruído, de forma que qualquer divergência entre duas distribuições se torna menos acentuada passando pelo canal.

Proposição 1.2.4 (Monotonicidade da Entropia Relativa [10])**.** *Dadas $p, q \in \mathcal{P}(\mathcal{X})$ e $(\mathcal{X}, N, \mathcal{Y})$ um canal clássico, então:*

$$D(p||q) \geq D(N(p)||N(q)).$$

Esta proposição é equivalente a desigualdade do processamento de dados (Teorema 1.1.24). Para uma demonstração, ver [27]. O análogo quântico da Proposição 1.2.4 é a dita desigualdade de processamento de dados quântica (Proposição 3.2.15).

1.2.2 Códigos de Canal

Vamos formalizar o canal resultante do uso repetido independente do canal para transmitir sequências de símbolos (palavras) de tamanho arbitrário.

Definição 1.2.5. A **n-ésima extensão sem memória do canal** $(\mathcal{X}, N \equiv p(y|x), \mathcal{Y})$ é dada por $(\mathcal{X}^n, N^{\times n} \equiv p(y^n|x^n), \mathcal{Y}^n)$ onde:

$$p(y^n|x^n) = \prod_{i=1}^{n} p(y_i|x_i).$$

Se a entrada for dada por $X^n \sim p_{X^n}$, então a saída é dada por

$$Y^n = N^{\times n}(X^n) = (N(X_1), N(X_2), \ldots, N(X_n)) = (Y_1, Y_2, \ldots, Y_n),$$

e temos que:

$$p(y^n|x^n) = p_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^{n} p_{Y_i|X_i}(y_i|x_i). \tag{1.5}$$

Ou seja, a n-ésima extensão sem memória corresponde a n usos independentes do canal. Note que não temos necessariamente que $X^n = (X_1, X_2, \ldots, X_n)$ é i.i.d. De fato, no geral observamos correlações entre usos seguidos de um canal. Por exemplo, se X^n é uma palavra da língua portuguesa, com cada entrada uma letra do alfabeto latino, esperamos que se X_1 for uma consoante, então a probabilidade de X_2 ser uma vogal é relativamente maior. Na Seção 1.4 vamos demonstrar que correlações em X^n, não podem ser utilizadas para melhorar a performance do canal, sob a hipótese de que seus usos são independentes.

Chamamos de **canal discreto sem memória (canal DSM)** um canal considerado com suas n-ésimas extensões sem memória, para todo $n > 1$.

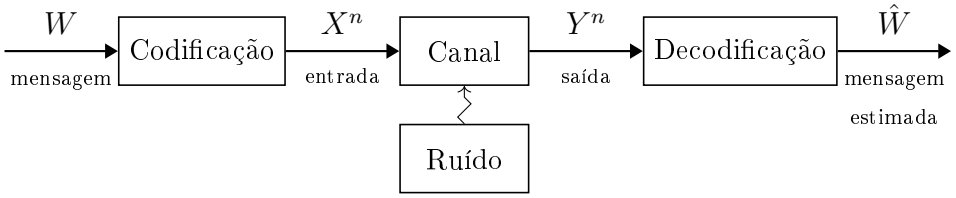

Figura 1.3: Modelo de comunicação por um canal ruidoso.

Vejamos agora o modelo de comunicação por canal ruidoso entre dois sistemas A e B, tradicionalmente referidos como Alice e Bob. Alice deseja enviar uma mensagem m dentre um conjunto de mensagens $\mathcal{M}$. Para isso ela codifica cada mensagem por uma palavra-código $x^n(m) \in \mathcal{X}^n$ de comprimento n, e envia uma letra por vez, ou equivalentemente, envia a palavra pela n-ésima extensão sem memória do canal. Bob recebe a saída $y^n \in \mathcal{Y}^n$ a qual ele atribui a mensagem $\hat{m}$.

Definição 1.2.6 (Código de canal)**.** Um (M, n)**-código** para o canal $(\mathcal{X}, N, \mathcal{Y})$ DSM consiste em:

- Um conjunto $\mathcal{M} = \{1, 2, \ldots, M\}$ de mensagens.
- Uma função codificação $\mathcal{C}^n : \mathcal{M} \to \mathcal{X}^n$ que leva cada mensagem $m \in \mathcal{M}$ à uma palavra $\mathcal{C}^n(m) = x^n(m)$, dita a **palavra-código** de m. Denotamos $X^n = \mathcal{C}^n(W)$, onde W é uniformemente distribuída em $\mathcal{M}$.
- Uma função decodificação $\mathcal{D}^n : \mathcal{Y}^n \to \mathcal{M}$, a partir da qual definimos a variável aleatória $\hat{W} = \mathcal{D}^n(Y^n)$, com $Y^n = N^{\times n}(X^n)$.

Definição 1.2.7. Dado um (M, n)-código para um Canal DSM definimos:

Probabilidade condicional de erro:

$$P_e^{(n)}(m) = p_W(\hat{W} \neq W | W = m);$$

Probabilidade média de erro:

$$P_e^{(n)} = \frac{1}{M} \sum_{m=1}^{M} P_e^{(n)}(m);$$

Probabilidade máxima de erro:

$$P_{e,max}^n = \max_{1 \leq m \leq M} P_e^{(n)}(m).$$

A hipótese de que as mensagens W são uniformemente distribuídas não gera perda de generalidade (Ver [9][24]) e tem a consequência conveniente de que a probabilidade média de erro é de fato igual a probabilidade de um erro ocorrer:

$$p_W(\hat{W} \neq W) = \sum_{m=1}^{M} p_W(\hat{W} \neq W \cap W = m) = \frac{1}{M} \sum_{m=1}^{M} \frac{p_W(\hat{W} \neq W \cap W = m)}{p_W(W = m)} = P_e^{(n)}. \tag{1.6}$$

O próximo teorema será usado de forma essencial na Seção 1.3, no contexto do método de codificação aleatória.

Teorema 1.2.8. *Dado um (M, n)-código para um canal DSM, se $P_e^{(n)} < \epsilon$ para algum $\epsilon \in [0, \frac{1}{2}]$, então para pelo menos $\lceil M/2 \rceil$ mensagens m, temos que:*

$$P_e^{(n)}(m) < 2\epsilon.$$

Demonstração. Suponha que existem $\lceil M/2 \rceil$ mensagens com probabilidade de erro maior que 2ϵ, que podemos assumir, sem perda de generalidade, serem $\{1, 2, \ldots, \lceil M/2 \rceil\}$. Então:

$$P_e^{(n)} = \frac{1}{M} \sum_{m=1}^{M} P_e^{(n)}(m)$$

$$
\begin{aligned}
&= \frac{1}{M}\sum_{m=1}^{\lceil M/2\rceil} P_e^{(n)}(m) + \frac{1}{M}\sum_{m=\lceil M/2\rceil+1}^{M} P_e^{(n)}(m) \\
&\geq \frac{2\epsilon}{M}\lceil M/2\rceil + \frac{1}{M}\sum_{m=\lceil M/2\rceil+1}^{M} P_e^{(n)}(m) \\
&\geq \epsilon + \frac{1}{M}\sum_{m=\lceil M/2\rceil+1}^{M} P_e^{(n)}(m) \\
&\geq \epsilon,
\end{aligned}
$$

que contradiz a hipótese. □

Se Alice usa um canal DSM para transmitir uma entre M mensagens equiprováveis, ela dará como entrada uma palavra-código binária de comprimento $\log M$. A taxa de transmissão de dados é a razão entre o comprimento da mensagem original e o da mensagem codificada. O objetivo da codificação de canal é encontrar um código que maximize a taxa de transmissão pelo canal, transmitindo o máximo de informação com o mínimo de usos do canal, sem que a mensagem seja corrompida pelo ruído.

Definição 1.2.9 (Taxa de Transmissão)**.** A **taxa de transmissão** de um (M, n)-código é dada por:

$$R = \frac{\log M}{n},$$

com unidade bits/transmissão. O número de mensagens M que podemos transmitir com uma taxa R, usando blocos de tamanho n, é $M = \lceil 2^{nR} \rceil$. No que segue, quando escrevemos $(2^{nR}, n)$-códigos, nos referimos a $(\lceil 2^{nR} \rceil, n)$-códigos.

Definição 1.2.10 (Taxa Atingível)**.** Dizemos que uma taxa R é **atingível** para o canal $(\mathcal{X}, N, \mathcal{Y})$ se $R = 0$ ou se existe uma sequência de (M_n, n)-códigos para N, tais que:

$$R = \lim_{n\to\infty} \frac{\log M_n}{n} \quad \text{com} \quad P_{e,max}^{(n)} \to 0.$$

Definição 1.2.11 (Capacidade Operacional)**.** A **capacidade operacional** de um canal $(\mathcal{X}, N, \mathcal{Y})$ é dada por:

$$C_{op}(N) = \sup\{R \geq 0 \,|\, R \text{ é atingível para } (\mathcal{X}, N, \mathcal{Y})\}. \tag{1.7}$$

Ou seja, a capacidade operacional é a máxima taxa de transmissão de bits por uso do canal que pode ser atingida com probabilidade de erro negligente. Transmissões pelo canal com taxa acima de $C_{op}(N)$ vão inevitavelmente conter uma maior parcela de erros, mas ainda assim são consideradas de interesse na teoria da informação (ver, por exemplo, *Rate-distortion Theory* em [9]). Aqui vamos nos concentrar em computar a capacidade operacional de um canal, que a princípio não é uma tarefa trivial, pois se trata de uma quantidade assintótica.

Definição 1.2.12 (Capacidade Informacional)**.** A **capacidade informacional** de um canal $(\mathcal{X}, N, \mathcal{Y})$ é dada por:

$$C(N) = \sup_{p\in\mathcal{P}(\mathcal{X})} I(p : N(p)) = \sup_{p\in\mathcal{P}(\mathcal{X})} \left\{ H(N(p)) - \sum_{x\in\mathcal{X}} p(x) H(N(\delta_x)) \right\}, \tag{1.8}$$

onde $N(p)$ é dada por (1.4). Se a entrada é dada por X e a saída por Y, então:

$$C(N) = \sup_{p_X\in\mathcal{P}(\mathcal{X})} I(X : Y) = \sup_{p_X\in\mathcal{P}(\mathcal{X})} \left\{ H(Y) - \sum_{x\in\mathcal{X}} p_X(x) H(N(x)) \right\}. \tag{1.9}$$

Note que $0 \leq I(X : Y) \leq H(X) \leq \log(|\mathcal{X}|)$, ou seja, $I(X : Y)$ é limitada com respeito a X. Vejamos também que $I(X, Y)$ é côncava com respeito a p_X. Como $\mathcal{P}(\mathcal{X})$ é compacto em $\mathbb{R}^{\mathcal{X}}$, sempre existe uma distribuição p_X que é máximo global de $I(X, Y)$, o que não quer dizer que tal distribuição possa ser caracterizada facilmente.

Teorema 1.2.13. *Dado um canal $(\mathcal{X}, N, \mathcal{Y})$ e $Y = N(X)$, $I(X : Y)$ é côncava com respeito a p_X.*

Demonstração. Por definição, $I(X : Y) = H(Y) - H(Y|X)$. Agora:

$$H(Y) = -\sum_{y\in\mathcal{Y}} p_Y(y) \log p_Y(y) = -\sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} p_{Y|X}(y|x) p_X(x) \log \left\{ \sum_{x\in\mathcal{X}} p_{Y|X}(y|x) p_X(x) \right\}.$$

Logo, para $p_{X|Y}$ fixo, $H(Y)$ é côncava com respeito à p_X. Também temos, pela Definição 1.1.10:

$$H(Y|X) = -\sum_{x\in\mathcal{X}} p_X(x) \sum_{y\in\mathcal{Y}} p_{Y|X}(y|x) \log p_{Y|X}(y|x),$$

que é linear com respeito à p_X para $p_{Y|X}$ fixo. Assim, $I(X : Y)$ é a soma de uma função côncava e uma linear em p_X, logo $I(X : Y)$ deve ser côncava em p_X. □

Como o supremo em (1.8) é de uma função côncava sobre o conjunto fechado convexo $\mathcal{P}(\mathcal{X})$, ele é de fato atingido:

Corolário 1.2.14. *Para todo canal $(\mathcal{X}, N, \mathcal{Y})$, existe $p \in \mathcal{P}(\mathcal{X})$ tal que:*

$$C(N) = \max_{p\in\mathcal{P}(\mathcal{X})} I(p : N(p)).$$

Exceto para certas classes simples de canais [24], a capacidade informacional não admite uma fórmula fechada e deve ser computada numericamente com métodos de programação não-linear, como o método do gradiente [9].

O Teorema da Codificação de Canal de Shannon afirma que:

> *A capacidade operacional de um Canal DSM $(\mathcal{X}, N, \mathcal{Y})$ coincide com sua capacidade informacional:*
>
> $$C_{op}(N) = C(N).$$
>
> *Ou seja, uma taxa de transmissão R é atingível se, e somente se, $R < C(N)$.*

Note que isso mostra que a capacidade operacional, definida em relação a um limite assintótico de múltiplos usos do canal, corresponde à capacidade informacional, que quantifica a máxima correlação entre entrada e saída de um único uso do canal. A hipótese de que o canal é DSM é essencial. O resultado análogo para o caso em que as transmissões do canal não forem independentes, mas definirem um processo ergódico, é dado pelo Teorema Shannon-McMillan-Breiman (ver [23]).

Antes de demonstrar o teorema, vejamos alguns exemplos de cálculo da capacidade de canal.

Exemplo 1.2.15. Vamos calcular a capacidade informacional de $\mathcal{B}_\alpha$ do Exemplo 1.2.3, com entrada X e saída Y:

$$\begin{aligned} I(X:Y) &= H(Y) - H(Y|X) \\ &= H(Y) - H_{bin}(\alpha) \\ &\leq 1 - H_{bin}(\alpha), \end{aligned}$$

onde a última desigualdade parte de $H(Y) \leq \log 2 = 1$. Tomando $X \sim \mathcal{U}_2$, temos que $Y \sim \mathcal{U}_2$ e portanto que $H(Y) = 1$. Logo $C(\mathcal{B}(\alpha)) = 1 - H_{bin}(\alpha)$.

Exemplo 1.2.16 (Máquina de Escrever com Ruído)**.** Se $\mathcal{X} = \mathcal{Y} = \{A, B, C, \ldots, Y, Z\}$, chamamos de **máquina de escrever com ruído** o canal dado pela Figura 1.4, tal que $p(y|x) = \frac{1}{2}$ caso $x = y$ ou x for a próxima letra do alfabeto depois de y (convencionando que A vem depois de Z). A capacidade deste canal é $\log 13$. Vejamos isso de duas formas.

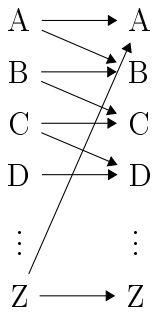

Figura 1.4: Máquina de escrever com ruído.

Primeiro, como no exemplo anterior:

$$C(\mathcal{M}) = \max I(X:Y) = \max(H(Y) - H(Y|X)) = \max(H(Y) - 1) = \log 26 - 1 = \log 13.$$

Mas podemos chegar neste valor de forma que, como veremos a seguir, ilustra um método geral para obter (em teoria) a capacidade de um canal. Basta notar que podemos transmitir informação pelo canal sem erro se enviarmos, por exemplo, apenas as letras $\{A, C, D, F \ldots, Y\}$. De fato podemos particionar $\mathcal{Y}$ em 13 conjuntos disjuntos correspondendo unicamente à 13 letras que podemos enviar fielmente, nos dando a capacidade de $\log 13$ bits. Podemos, por exemplo, considerar um $(13, 1)$-código dado por:

$$\mathcal{C}(A) \in \{A, B\},\ \mathcal{C}(C) \in \{C, D\}, \ldots,\ \mathcal{C}(Y) \in \{Y, Z\}.$$

Exemplo 1.2.17 (Canal de Capacidade Zero)**.** Para que a definição de capacidade seja útil, devemos ter que um canal cuja capacidade é zero não pode ser usado para transmitir informação. A saída de tal canal deve ser aleatória. De fato, um canal $(\mathcal{X}, N \equiv p(y|x), \mathcal{Y})$ tem capacidade zero se e somente se N for constante, ou seja, se existe $\nu \in \mathcal{P}(\mathcal{Y})$ tal que $\forall x \in \mathcal{X}$, $N(x) = \nu$. Vejamos isso. Pelo Teorema da Codificação de Canais temos que, para qualquer $p \in \mathcal{P}(\mathcal{X})$, $I(p : N(p)) = 0$. Pelo Teorema 1.1.16, temos que p e $N(p)$ são independentes. Logo $N(p) = \nu \in \mathcal{P}(\mathcal{Y})$.

Vamos nos dedicar agora a demonstração do Teorema da Codificação de Canais, que mostra que, em média, podemos transmitir fielmente $C(N)$ bits de informação por uso do canal. Intuitivamente, a demonstração mostra que o uso independente do canal várias vezes leva a um comportamento parecido com o da máquina de escrever com ruído (Exemplo 1.2.16), onde o conjunto de possíveis sequências produzidas pelo canal pode ser coberto por conjuntos quase disjuntos, correspondentes a entradas típicas.

O argumento utiliza uma variação da propriedade da equipartição assintótica: um canal DSM leva cada entrada típica X^n a aproximadamente $2^{nH(Y|X)}$ sequências típicas em Y^n, todas aproximadamente equiprováveis. O número total de sequências de saída típicas de comprimento n é aproximadamente $2^{nH(Y)}$, logo podemos particionar $\mathcal{Y}^n$ em no máximo $2^{n(H(Y)-H(Y|X))} = 2^{nI(X:Y)}$ conjuntos disjuntos, que corresponde às possíveis entradas típicas. Assim, podemos fielmente transmitir no máximo $\approx 2^{nI(X:Y)}$ mensagens de comprimento n, ou seja, a uma taxa de informação, no máximo, $I(X : Y)$. O desafio maior é mostrar que esta taxa é assintoticamente atingível. Para formalizar esse argumento, começamos estendendo o conceito de tipicalidade para canais.

1.2.3 Tipicalidade Condicional

Definição 1.2.18 (Entropia Amostral Condicional)**.** Dados $x^n \in \mathcal{X}^n$, $y^n \in \mathcal{Y}^n$, a **entropia amostral condicional** de (x^n, y^n) com respeito a $(X, Y) \sim p_{X,Y}(x, y) = p_X(x)p_{Y|X}(y|x)$ é dada por:

$$\overline{H}(y^n|x^n) = -\frac{1}{n} \log p_{Y^n|X^n}(y^n|x^n),$$

onde

$$p_{Y^n|X^n}(y^n|x^n) = p_{Y|X}(y_1|x_1)p_{Y|X}(y_2|x_2) \cdots p_{Y|X}(y_n|x_n).$$

Definição 1.2.19 (Conjunto Condicionalmente Típico)**.** Dado $x^n \in \mathcal{X}^n$, $y^n \in \mathcal{Y}^n$ é dita **condicionalmente δ-típica** com x^n em relação a (X, Y) se:

$$\left|\overline{H}(y^n|x^n) - H(Y|X)\right| < \delta.$$

Chamamos de **conjunto condicionalmente δ-típico** com x^n o conjunto de sequências condicionalmente δ-típicas com x^n em relação à (X, Y):

$$T_\delta^{Y^n|x^n} = \{y^n \in \mathcal{Y}^n \mid \left|\overline{H}(y^n|x^n) - H(Y|X)\right| < \delta\}.$$

Quando o contexto for claro, chamamos conjuntos da forma $T_\delta^{Y^n|x^n}$ de conjuntos condicionalmente típicos.

Vejamos agora que conjuntos condicionalmente típicos possuem propriedades similares às vistas na Seção 1.1.5 para conjuntos típicos.

Teorema 1.2.20 (Equipartição)**.** *A probabilidade condicional $p_{Y^n|X^n}$ é aproximadamente uniforme em $T_\delta^{Y^n|x^n}$. Para quaisquer $x^n \in \mathcal{X}^n, y^n \in \mathcal{Y}^n$:*

$$2^{-n(H(Y|X)+\delta)} \leq p_{Y^n|X^n}(y^n|x^n) \leq 2^{-n(H(Y|X)-\delta)}. \tag{1.10}$$

Demonstração. Imediato da definição. □

Teorema 1.2.21 (Acúmulo (em Média) de Probabilidade)**.** *Para todo $\epsilon \in (0,1)$ e $\delta > 0$, existe $n \in \mathbb{N}$ tal que, em média com respeito à p_{X^n}, $T_\delta^{Y^n|x^n}$ tem probabilidade maior que $1-\epsilon$:*

$$\mathbb{E}_{X^n}\left\{p_{Y^n|X^n}\left\{Y^n \in T_\delta^{Y^n|X^n}\right\}\right\} \geq 1-\epsilon. \tag{1.11}$$

Note como estamos introduzindo mais uma camada de aleatoriedade, supondo que a sequência x^n é dada pela variável aleatória X^n, e depois eliminando esta aleatoriedade tomando a média. Como veremos, este é um truque essencial na demonstração do Teorema da Codificação de Canais.

Demonstração. O resultado vai seguir de uma aplicação da Propriedade da Equipartição Assintótica (Teorema 1.1.31). Denotando a função indicadora de um conjunto Ω por I_Ω, temos:

$$\begin{aligned}\mathbb{E}_{X^n}\left\{p_{Y^n|X^n}\left\{Y^n \in T_\delta^{Y^n|X^n}\right\}\right\} =& \mathbb{E}_{X^n}\left\{\mathbb{E}_{Y^n|X^n}\left\{I_{T_\delta^{Y^n|X^n}}(Y^n)\right\}\right\} \\ =& \mathbb{E}_{X^n,Y^n}\left\{I_{T_\delta^{Y^n|X^n}}(Y^n)\right\} \\ =& p_{X^n,Y^n}\left\{Y^n \in T_\delta^{Y^n|X^n}\right\}. \end{aligned} \tag{1.12}$$

Lembrando da noção de informação condicional, podemos definir a variável aleatória $i(Y|X) = -\log p_{Y|X}(Y|X)$, que usamos para expressar a entropia amostral condicional $H(Y^n|X^n)$ das sequências i.i.d. X^n e Y^n:

$$\begin{aligned}\overline{H}(Y^n|X^n) &= -\frac{1}{n}\log p_{Y^n|X^n}(Y^n|X^n) \\ &= -\frac{1}{n}\log\left(\prod_{k=1}^n p_{Y|X}(Y_k|X_k)\right) \\ &= -\frac{1}{n}\sum_{k=1}^n \log p_{Y|X}(Y_k|X_k) \\ &= \frac{1}{n}\sum_{k=1}^n i(Y_k|X_k).\end{aligned}$$

Com isso em mãos, e lembrando da equação (1.2) para $H(Y|X)$, podemos reescrever (1.12) como:

$$p_{X^n,Y^n}\left\{\left|\overline{H}(Y^n|X^n) - H(Y|X)\right| < \delta\right\} = p_{X^n,Y^n}\left\{\left|\frac{1}{n}\sum_{k=1}^n i(Y_k|X_k) - \mathbb{E}_{X,Y} i(Y|X)\right| < \delta\right\},$$

que, pela Lei dos Grandes Números, tende a 1 e portanto existe n grande o suficiente tal que (1.11) é satisfeita. □

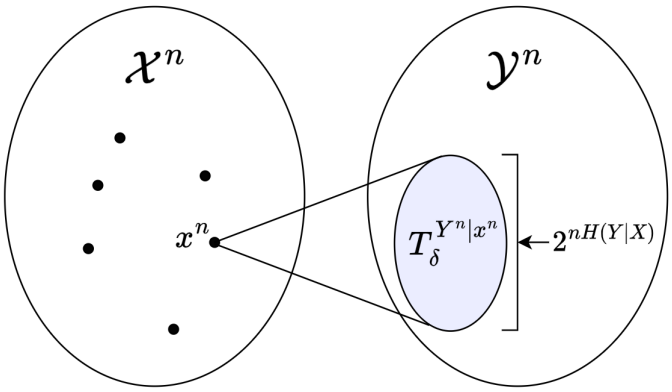

Figura 1.5: Conjunto condicionalmente δ-típico com x^n em relação a (X, Y).

Teorema 1.2.22 (Cardinalidade)**.** *Para qualquer $\delta > 0$, existe $n \in \mathbb{N}$ tal que, para todo $x^n \in \mathcal{X}^n$:*

$$\left|T_\delta^{Y^n|x^n}\right| < 2^{n(H(Y|X)+\delta)}. \tag{1.13}$$

Além disso, temos que, para todo $\epsilon \in (0,1)$, existe $n \in \mathbb{N}$ tal que:

$$\mathbb{E}_{X^n}\left\{\left|T_\delta^{Y^n|X^n}\right|\right\} \geq (1-\epsilon)2^{n(H(Y|X)-\delta)}. \tag{1.14}$$

Demonstração. A demonstração de (1.13) é idêntica à do Teorema 1.1.35. Para ver que vale (1.14), note que, pelo Teorema 1.2.21, existe $n > 0$ com:

$$\begin{aligned}
1-\epsilon &\leq \mathbb{E}_{X^n}\left\{p_{Y^n|X^n}\left\{Y^n \in T_\delta^{Y^n|X^n}\right\}\right\} \\
&\leq \mathbb{E}_{X^n}\left\{\sum_{y^n \in T_\delta^{Y^n|X^n}} p_{Y^n|X^n}(y^n)\right\} \\
&\leq \mathbb{E}_{X^n}\left\{2^{-n(H(Y|X)-\delta)}\left|T_\delta^{Y^n|X^n}\right|\right\} \\
&= 2^{-n(H(Y|X)-\delta)}\mathbb{E}_{X^n}\left\{\left|T_\delta^{Y^n|X^n}\right|\right\}.
\end{aligned}$$

□

1.3 Teorema da Codificação de Canais

Estamos prontos para demonstrar o Teorema da Codificação de Canais.

Teorema 1.3.1 (Teorema da Codificação de Canais)**.** *A capacidade operacional de um canal DSM $(\mathcal{X}, N \equiv p(y|x), \mathcal{Y})$ é dada por sua capacidade informacional:*

$$C_{op}(N) = C(N)$$

ou seja, uma taxa de transmissão R é atingível para $(\mathcal{X}, N, \mathcal{Y})$ se e somente se $R \leq C(N)$.

É comum na literatura dividir a demonstração deste teorema em duas partes:

Parte Direta: Para $R < C(N)$, usamos o *método probabilístico* para mostrar a existência de uma sequência de códigos cujas taxas tendem à R mas com probabilidade máxima de erros tendendo a zero.

Parte Recíproca: Para $R > C(N)$, usamos a *aditividade da informação mútua* para mostrar que não pode existir uma sequência de códigos com taxas tendendo a R mas cuja probabilidade máxima de erro tende a zero.

Por método probabilístico nos referimos ao método de demonstração não-construtivo que garante a existência de um objeto em um espaço de probabilidade satisfazendo alguma propriedade mostrando que a probabilidade de tal objeto existir é não-nula. No nosso contexto, o método é chamado de **codificação aleatória**, onde o espaço de probabilidade será o de possíveis códigos para um canal. Em [31] é demonstrada uma versão mais forte da parte recíproca, que dá uma cota inferior explícita para a probabilidade máxima de erro e mostra que para códigos com taxa assintótica maior que $C(N)$, $\liminf_{n\to\infty} P_e^{(n)} = 1$ (ver também [24, 25]).

1.3.1 Parte Direta

Demonstração. Lembre que um $(2^{nR}, n)$-código determina $M = 2^{nR}$ palavras-código de comprimento n, uma para cada mensagem do conjunto $\mathcal{M} = \{1, 2, \ldots, M\}$:

$$x^n(1), x^n(2), \ldots, x^n(M) \in \mathcal{X}^n,$$

dadas pelo esquema de codificação $\mathcal{C}^n : \mathcal{M} \to \mathcal{X}^n$ que, pensado como a sequência $(x^n(1), x^n(2), \ldots, x^n(M))$, chamamos de **livro-código**.

Fixada uma taxa de transmissão $R < C(N)$, vamos primeiro mostrar a existência de uma sequência de (M, n)-códigos, com $M = 2^{nR+1}$ (ou seja, com taxas de transmissão $R + \frac{1}{n}$), tais que $P_e^{(n)} \to 0$. Depois argumentamos, pelo Teorema 1.2.8, que existe uma sequência de $(2^{nR}, n)$-códigos tais que $P_{e,max}^{(n)} \to 0$. No contexto do método de codificação aleatória, este último passo é as vezes chamado de *expurgação* [10].

Vejamos este problema pela perspectiva de Alice e Bob, que começam fixando $\delta > 0$ e uma distribuição $p_X \in \mathcal{P}(\mathcal{X})$ que maximiza a informação mútua entre a entrada e saída do canal, ou seja, tal que $I(X, Y) = C(N)$ com $Y = N(X)$ (que existe, pelo Corolário 1.2.14). Para todo $n > 0$, eles combinam o seguinte (M, n)-código:

Codificação: Alice e Bob escolhem um livro-código $\mathcal{C}$ (omitimos por enquanto o índice n) de forma aleatória: $\mathcal{C} = (x^n(1), x^n(2), \ldots, x^n(M))$, onde $x^n(m) = (x_1(m), x_2(m), \ldots, x_n(m))$ é uma amostra de $\mathcal{X}^n$ retirada i.i.d de acordo com a probabilidade:

$$p_{X^n}(x^n(m)) = \prod_{i=1}^{n} p_X(x_i(m)) = \prod_{i=1}^{n} p_X(x_i). \tag{1.15}$$

Ou seja, a probabilidade de uma palavra-código ser selecionada é independente de escolha das outras palavras-código e também da mensagem m. Logo, os livros-código são elementos do espaço de probabilidade $(\mathcal{X}^n)^M$ e são distribuídos por:

$$p_{\mathcal{C}}(x^n(1), x^n(2), \ldots, x^n(M)) = \prod_{m=1}^{M} \prod_{i=1}^{n} p_X(x_i(m)). \tag{1.16}$$

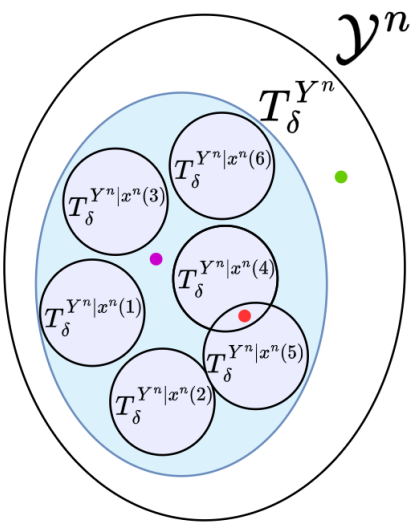

Figura 1.6: Possíveis erros de decodificação da mensagem 4. O ponto verde pertence à $\mathcal{E}_0(4)$, o ponto rosa pertence à $\mathcal{E}_1(4)$ e o ponto vermelho pertence à $\mathcal{E}_2(4)$.

> **Decodificação:** Ao receber a saída y^n do canal, Bob determina, baseado em certos critérios, se vai decodificar y^n ou declarar que houve um erro. No caso de erro, ele simplesmente decodifica y^n como $\mathcal{D}(y^n) = m'$, com $m' \in \mathcal{M}$ aleatória, escolhida uniformemente. Primeiro Bob testa se y^n pertence ao conjunto δ-típico $T_\delta^{Y^n}$ referente a distribuição $Y^n = N^{\times n}(X^n)$ dada pela equação (1.5) na página 16, declarando um erro caso não pertença. Ele então testa se existe alguma mensagem m tal que y^n está no conjunto condicionalmente δ-típico $T_\delta^{Y^n|x^n(m)}$ (lembre que Bob conhece o livro-código). Caso não exista tal mensagem, Bob declara um erro. Caso existam duas ou mais tais mensagens, Bob também declara um erro. E caso exista uma única tal mensagem m, Bob toma $\mathcal{D}(y^n) = m$.

Podemos particionar em três eventos (Figura 1.6) o caso em que ocorre um erro de decodificação quando Alice envia a palavra-código $x^n(m)$:

$\mathcal{E}_0(m)$: Bob recebe y^n que não é δ-típica.

$\mathcal{E}_1(m)$: Bob recebe y^n δ-típica, mas não condicionalmente δ-típica com $x^n(m)$. Ou seja, $y^n \notin T_\delta^{Y^n|x^n(m)}$.

$\mathcal{E}_2(m)$: Bob recebe y^n δ-típica, condicionalmente δ-típica com $x^n(m)$ e condicionalmente δ-típica com $x^n(m')$, com $m' \neq m$. Ou seja, $y^n \in T_\delta^{Y^n|x^n(m)} \cap T_\delta^{Y^n|x^n(m')}$.

Vamos estimar a probabilidade de ocorrência de cada tipo de erro. Para isso, considere $\mathcal{C}$ como vetor aleatório com valor no conjunto de possíveis livros-código $(\mathcal{X}^n)^M$ e com distribuição dada por (1.16). Tomando o valor esperado da probabilidade média de erro (Definição 1.2.7) sobre a distribuição de $\mathcal{C}$, temos:

$$\mathbb{E}_{\mathcal{C}}\left\{P_e^{(n)}\right\} = \mathbb{E}_{\mathcal{C}}\left\{\frac{1}{M}\sum_{m=1}^{M} P_e^{(n)}(m)\right\} = \mathbb{E}_{\mathcal{C}}\left\{\frac{1}{M}\sum_{m=1}^{M} p_{Y^n|\mathcal{C}}\left\{\mathcal{E}_0(m) \cup \mathcal{E}_1(m) \cup \mathcal{E}_2(m)\right\}\right\}.$$

Por linearidade:

$$\mathbb{E}_{\mathcal{C}}\left\{P_e^{(n)}\right\} = \frac{1}{M}\sum_{m=1}^{M} \mathbb{E}_{\mathcal{C}}\left\{p_{Y^n|\mathcal{C}}\left\{\mathcal{E}_0(m) \cup \mathcal{E}_1(m) \cup \mathcal{E}_2(m)\right\}\right\}.$$

Note que, como a escolha de $x^n(m)$ é independente de m, a probabilidade de erro não depende da mensagem m, já que toda mensagem tem a mesma probabilidade de ser emitida. Podemos então assumir, sem perda de generalidade, que a palavra-código enviada por Alice se refere à uma mensagem $m \in \mathcal{M}$ fixada:

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\left\{P_e^{(n)}\right\} &= \frac{1}{M}\sum_{m'=1}^{M}\mathbb{E}_{\mathcal{C}}\left\{p_{Y^n|\mathcal{C}}\left\{\mathcal{E}_0(m)\cup\mathcal{E}_1(m)\cup\mathcal{E}_2(m)\right\}\right\}\\ &= \mathbb{E}_{\mathcal{C}}\left\{p_{Y^n|\mathcal{C}}\left\{\mathcal{E}_0(m)\cup\mathcal{E}_1(m)\cup\mathcal{E}_2(m)\right\}\right\}.\end{aligned}$$

Pela subaditividade da probabilidade:

$$\mathbb{E}_{\mathcal{C}}\left\{P_e^{(n)}\right\} \leq \mathbb{E}_{\mathcal{C}}\left\{p_{Y^n|\mathcal{C}}\left\{\mathcal{E}_0(m)\right\}\right\} + \mathbb{E}_{\mathcal{C}}\left\{p_{Y^n|\mathcal{C}}\left\{\mathcal{E}_1(m)\right\}\right\} + \mathbb{E}_{\mathcal{C}}\left\{p_{Y^n|\mathcal{C}}\left\{\mathcal{E}_2(m)\right\}\right\}. \quad (1.17)$$

Podemos agora analisar cada tipo de erro individualmente. Note que as funções indicadoras dos eventos $\mathcal{E}_0(m), \mathcal{E}_1(m)$ e $\mathcal{E}_2(m)$ são dadas, em termos da saída y^n, por:

$$I_{\mathcal{E}_0(m)} = 1 - I_{T_\delta^{Y^n}}(y^n), \quad (1.18)$$

$$I_{\mathcal{E}_1(m)} = I_{T_\delta^{Y^n}}(y^n)(1 - I_{T_\delta^{Y^n|X^n(m)}}(y^n)), \quad (1.19)$$

$$I_{\mathcal{E}_2(m)} = \sum_{m'\neq m} I_{T_\delta^{Y^n}}(y^n) I_{T_\delta^{Y^n|X^n(m)}}(y^n) I_{T_\delta^{Y^n|X^n(m')}}(y^n). \quad (1.20)$$

Como as coordenadas de $\mathcal{C} = \{X^n(1), X^n(2), \ldots, X^n(m)\}$ são i.i.d., podemos levar em conta somente a distribuição marginal $X^n(m)$ e substituir $\mathbb{E}_{\mathcal{C}}$ por $\mathbb{E}_{X^n(m)}$ e $p_{Y^n|\mathcal{C}}$ por $p_{Y^n|X^n(m)}$ em (1.18) e (1.19). Assim, utilizando (1.18), e que para uma variável aleatória Z e um evento Ω, $\mathbb{E}_Z\{I_\Omega\} = p_Z\{\Omega\}$, temos:

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\{p_{Y^n|X^n(m)}\{\mathcal{E}_0(m)\}\} &= \mathbb{E}_{X^n(m)}\left\{p_{Y^n|X^n(m)}\left\{\mathcal{E}_0(m)\right\}\right\}\\ &= \mathbb{E}_{X^n(m)}\left\{\mathbb{E}_{Y^n|X^n(m)}\left\{1 - I_{T_\delta^{Y^n}}(y^n)\right\}\right\}\\ &= 1 - \mathbb{E}_{X^n(m),Y^n}\left\{I_{T_\delta^{Y^n}}(y^n)\right\}\\ &= 1 - \mathbb{E}_{Y^n}\left\{I_{T_\delta^{Y^n}}(y^n)\right\}\\ &= p_{Y^n}\{Y^n \notin T_\delta^{Y^n}\}.\end{aligned}$$

Por (1.19), temos:

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\{p_{Y^n|X^n(m)}\{\mathcal{E}_1(m)\}\} &= \mathbb{E}_{X^n(m)}\left\{\mathbb{E}_{Y^n|X^n(m)}\left\{I_{T_\delta^{Y^n}}(y^n)(1 - I_{T_\delta^{Y^n|X^n(m)}}(y^n))\right\}\right\}\\ &\leq \mathbb{E}_{X^n(m)}\left\{\mathbb{E}_{Y^n|X^n(m)}\left\{1 - I_{T_\delta^{Y^n|X^n(m)}}(y^n)\right\}\right\}\\ &= \mathbb{E}_{X^n(m)}\left\{1 - \mathbb{E}_{Y^n|X^n(m)}\left\{I_{T_\delta^{Y^n|X^n(m)}}(y^n)\right\}\right\}\\ &= \mathbb{E}_{X^n(m)}\left\{p_{Y^n|X^n(m)}\left\{Y^n \notin T_\delta^{Y^n|X^n(m)}\right\}\right\}.\end{aligned}$$

Para analisar a probabilidade de erro do evento $\mathcal{E}_2(m)$, tomamos $\tilde{X}^n = (X^n(m'))_{m'\neq m}$. Note que $p_{\mathcal{C}} = p_{X^n(m)} \cdot p_{\tilde{X}}$. Por (1.20), temos:

$$\mathbb{E}_{\mathcal{C}}\{p_{Y^n|\mathcal{C}}\{\mathcal{E}_2(m)\}\} = \mathbb{E}_{\mathcal{C}}\left\{\mathbb{E}_{Y^n|\mathcal{C}}\left\{\sum_{m'\neq m} I_{T_\delta^{Y^n}}(y^n) I_{T_\delta^{Y^n|X^n(m)}}(y^n) I_{T_\delta^{Y^n|X^n(m')}}(y^n)\right\}\right\}$$

$$
\begin{aligned}
&\leq \mathbb{E}_{\mathcal{C}} \left\{ \mathbb{E}_{Y^n|\mathcal{C}} \left\{ \sum_{m' \neq m} I_{T_\delta^{Y^n}}(y^n) I_{T_\delta^{Y^n|X^n(m')}}(y^n) \right\} \right\} \\
&\leq \sum_{m' \neq m} \mathbb{E}_{X^n(m),\tilde{X}^n} \left\{ \mathbb{E}_{Y^n|\tilde{X}^n} \left\{ I_{T_\delta^{Y^n}}(y^n) I_{T_\delta^{Y^n|X^n(m')}}(y^n) \right\} \right\} \\
&= \sum_{m' \neq m} \mathbb{E}_{X^n(m),X^n(m'),Y^n} \left\{ I_{T_\delta^{Y^n}}(y^n) I_{T_\delta^{Y^n|X^n(m')}}(y^n) \right\} \\
&= \sum_{x^n(m)} \sum_{m' \neq m} \sum_{x^n(m')} \sum_{y^n} p_{X^n}(x^n(m)) p_{X^n}(x^n(m')) \\
&\times p_{Y^n|X^n}(y^n|x^n(m)) I_{T_\delta^{Y^n}}(y^n) I_{T_\delta^{Y^n|X^n(m')}}(y^n) \\
&= \sum_{m' \neq m} \sum_{x^n(m')} \sum_{y^n} p_{X^n}(x^n(m')) \left(p_{Y^n}(y^n) I_{T_\delta^{Y^n}}(y^n) \right) I_{T_\delta^{Y^n|X^n(m')}}(y^n) \\
&\leq 2^{-n(H(Y)-\delta)} \sum_{m' \neq m} \sum_{x^n(m')} \sum_{y^n} p_{X^n}(x^n(m')) I_{T_\delta^{Y^n|X^n(m')}}(y^n) \\
&\leq 2^{-n(H(Y)-\delta)} 2^{n(H(Y|X)+\delta)} \sum_{m' \neq m} \sum_{x^n(m')} p_{X^n}(x^n(m')) \\
&\leq M 2^{-n(I(X:Y)-2\delta)} \\
&= 2^{nR+1}\, 2^{-n(C(N)-2\delta)} \\
&= 2^{-n(C(N)-R-2\delta-\frac{1}{n})}.
\end{aligned}
$$

A cota acima é chamada de **Lema do Empacotamento Clássico** por algumas referências [32][21]. Finalmente, pelos Teoremas 1.1.34 e 1.2.21, para todo $\epsilon \in (0,1)$, existe $N > 0$ suficientemente grande tal que, para todo $n > N$:

$$
\begin{aligned}
\mathbb{E}_{\mathcal{C}} \left\{ P_e^{(n)} \right\} &\leq p_{Y^n}\{Y^n \notin T_\delta^{Y^n}\} + \mathbb{E}_{X^n(m)} \left\{ p_{Y^n|X^n(m)} \left\{ Y^n \notin T_\delta^{Y^n|X^n(m)} \right\} \right\} + 2^{-n(C(N)-R-2\delta-\frac{1}{n})} \\
&\leq \epsilon + \epsilon + 2^{-n(C(N)-R-2\delta-\frac{1}{n})}.
\end{aligned}
$$

Como $R < C(N)$, podemos tomar $0 < \delta < \frac{C(N)-R}{2}$, de forma que $\mathbb{E}_{\mathcal{C}^n} \left\{ P_e^{(n)} \right\} \to 0$. Ou seja, em média, códigos aleatoriamente gerados a partir de nosso esquema tem probabilidade média de erro tendendo à zero. Em particular a probabilidade de que exista uma sequência de códigos $(\mathcal{C}^n)_{n\in\mathbb{N}}$ com esta propriedade é não-nula.

Pelo método probabilístico, concluímos que deve existir tal sequência (esse passo é às vezes chamado *derandomização*). Lembre porém que nosso objetivo é mostrar a existência de $(2^{nR}, n)$-códigos cuja probabilidade máxima de erro tende a zero. Mas isso parte do Teorema 1.2.8, já que podemos *expurgar* os códigos obtidos: excluímos a metade das mensagens com maior probabilidade de erro, finalmente obtendo uma sequência de $(M/2, n)$-códigos com taxa $\log(2^{nR+1}/2)/n = R$ tais que $P_{n,max}^{(n)} \to 0$. □

1.3.2 Parte Recíproca

Na demonstração que segue, vamos utilizar a hipótese da aditividade da informação mútua:

Se N é um canal DSM, então para todo $n > 0$ e X^n uma variável aleatória com valores em $\mathcal{X}^n$ temos que :

$$I(X^n : N^{\times n}(X^n)) \leq nC(N).$$

Note que aqui X^n não precisa ser i.i.d. Esse resultado vai seguir da aditividade da capacidade de canais, que vamos enunciar e demonstrar na próxima seção. Enfatizamos esse resultado pois, como veremos no capítulo 4, vamos ver que o análogo quântico desta propriedade não é geralmente válida.

Dado $R > C(N)$, vamos mostrar que para qualquer sequência de (M, n)-códigos com taxa tendendo a R, temos que:

$$\liminf_{n\to\infty} P_e^{(n)} > 0,$$

que implica que a probabilidade de erro máximo não tende a zero (pois $P_e^{(n)} \leq P_{e,max}^{(n)}$). Para isso vamos precisar de um resultado auxiliar. Lembre que um $(2^{nR}, n)$-código para o canal DSM N consiste em funções codificação $\mathcal{C}^n$ e decodificação $\mathcal{D}^n$, onde as mensagens são dadas pela variável aleatória W uniformemente distribuída e um erro do código ocorre quando $W \neq \hat{W}$, com $\hat{W} = \mathcal{D}_n(Y^n)$, $Y^n = N^{\times n}(X^n)$ e $X^n = \mathcal{C}^n(W)$.

Lema 1.3.2 (Desigualdade de Fano)**.** *Dado um $(2^{nR}, n)$-código para o canal DSM N:*

$$H(W|\hat{W}) \leq 1 + P_e^{(n)} \cdot nR.$$

Demonstração. Seja $e = \{W \neq \hat{W}\}$ o evento em que ocorre um erro e I_e sua função indicadora. Note que por (1.6), $\mathbb{P}\{I_e = 1\} = P_e^{(n)}$. Pelo Corolário 1.1.13, podemos expandir $H(I_e, W|\hat{W})$ de duas formas:

$$H(I_e, W|\hat{W}) = H(W|\hat{W}) + \underbrace{H(I_e|W, \hat{W})}_{=0}$$

$$H(I_e, W|\hat{W}) = \underbrace{H(I_e|\hat{W})}_{\leq 1} + \underbrace{H(W|I_e, \hat{W})}_{P_e^{(n)} \cdot nR},$$

onde $H(I_e|W, \hat{W}) = 0$ pois temos conhecimento da entrada e da saída. $H(I_e|\hat{W}) \leq H(I_e) \leq 1$ pelo Teorema 1.1.18 pois $H(I_e) = H_{bin}(P_e^{(n)}) \leq 1$. Pela definição de entropia condicional, temos que:

$$\begin{aligned} H(W|I_e, \hat{W}) &= \mathbb{P}\{I_e = 0\}H(W|\hat{W}, I_e = 0) + \mathbb{P}\{I_e = 1\}H(W|\hat{W}, I_e = 1) \\ &\leq P_e^{(n)} \log |M| \\ &= P_e^{(n)} \cdot nR, \end{aligned}$$

pois $H(W|\hat{W}, I_e = 1) \leq H(W) \leq \log M$ e $H(W|\hat{W}, I_e = 0) = 0$, já que $\{W = m\} = \{\hat{W} = m\} \cap \{I_e(m) = 0\}$ e portanto que $H(W|\hat{W}, I_e = 0) = \mathbb{E}_{\hat{W}} \left\{ H(W|\hat{W} = m, I_e(m) = 0) \right\} = \mathbb{E}_{\hat{W}} \{H(W|W = m\}) = 0$.

□

Demonstração da Parte Recíproca. Para $n > 0$ e um $(2^{nR}, n)$-código, note que $W \to X^n \to Y^n \to \hat{W}$ forma uma cadeia de Markov. Assim podemos usar a Desigualdade do Processamento de Dados (1.1.24) duas vezes para concluir que:

$$I(W : \hat{W}) \leq I(W : Y^n) \leq I(X^n : Y^n).$$

Pela Desigualdade de Fano, e pela hipótese de que $I(X^n : Y^n) \geq nC(N) \geq nI(X_1 : Y_1)$:

$$\begin{aligned} P_e^{(n)} \geq \frac{H(W|\hat{W}) - 1}{nR} &= \frac{H(W) - I(W : \hat{W}) - 1}{nR} \\ &\geq \frac{H(W) - I(X^n : Y^n) - 1}{nR} \\ &= \frac{nR - I(X^n : Y^n) - 1}{nR} \\ &= 1 - \frac{I(X^n : Y^n)}{nR} - \frac{1}{nR} \\ &\geq 1 - \frac{nI(X_1 : Y_1)}{nR} - \frac{1}{nR} \\ &\geq 1 - \frac{C(N)}{R} - \frac{1}{nR}. \end{aligned}$$

Assim, se $R > C(N)$, para qualquer sequência de códigos com taxa R, temos:

$$\liminf_{n\to\infty} P_e^{(n)} \geq 1 - \frac{C(N)}{R} > 0.$$

□

1.4 Aditividade da Informação Mútua

Na demonstração da parte recíproca do Teorema da Codificação de Canais, utilizamos o fato de que, para todo $n > 1$, $I(X^n : N^{\times n}(X^n)) \leq nC(N)$. Em particular temos:

$$C(N^{\times n}) \leq nC(N). \tag{1.21}$$

Nessa seção, vejamos que essa desigualdade atesta que usos repetidos de um canal DMS não permite comunicação com uma taxa maior que $C(N)$, que se refere a um único uso do canal. Na parte direta do Teorema da Codificação de Canais utilizamos as n-ésimas extensões sem memória de N, mas apenas consideramos o caso em que as entradas X^n são i.i.d. Vamos considerar agora entradas X^n possivelmente correlacionadas.

Definição 1.4.1 (Canal Produto)**.** Dados canais $(\mathcal{X}_1, N_1 \equiv p_1(y|x), \mathcal{Y}_1)$ e $(\mathcal{X}_2, N_2 \equiv p_2(y|x), \mathcal{Y}_2)$, chamamos de **canal produto** o canal $(\mathcal{X}_1 \times \mathcal{X}_1, N_1 \times N_2 \equiv p(y_1, y_2|x_1, x_2), \mathcal{Y}_1 \times \mathcal{Y}_2)$ tal que:

$$p(y_1, y_2|x_1, x_2) = p_1(y_1|x_1)p_2(y_2|x_2).$$

Se a entrada é dada por (X_1, X_2), então:

$$p(y_1, y_2|x_1, x_2) = p_{Y_1,Y_2|X_1,X_2}(y_1, y_2|x_1, x_2) = p_{Y_1|X_1}(y_1|x_1) \cdot p_{Y_2|X_2}(y_2|x_2). \tag{1.22}$$

Iterando, podemos definir o produto de qualquer número finito de canais como o canal resultante de uma sequência de usos independentes de cada fator. Note que as n-ésimas extensões sem memória são canais produto. É razoável questionar se correlações entre as entradas de um canal produto pode resultar em uma capacidade superior à soma das capacidades dos fatores. Isto é, se $C(N_1 \times N_2) > C(N_1) + C(N_2)$, onde

$$C(N_1 \times N_2) = \sup_{p_{X_1,X_2}} I(X_1, X_2 : Y_1, Y_2).$$

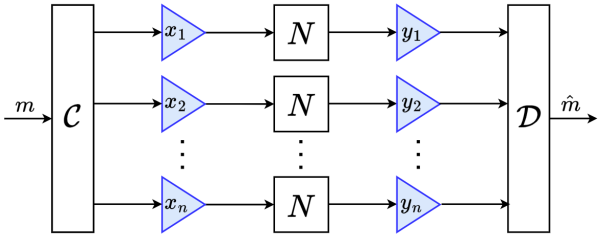

Figura 1.7: Usos independentes do canal para transmitir entradas correlacionadas. Triângulos representam informação clássica.

Teorema 1.4.2. *Dados canais* $(\mathcal{X}_1, N_1, \mathcal{Y}_1)$ *e* $(\mathcal{X}_2, N_2, \mathcal{Y}_2)$*, temos:*

$$C(N_1 \times N_2) \geq C(N_1) + C(N_2).$$

Demonstração. Sejam $p_{X_1} \in \mathcal{P}(\mathcal{X}_1)$ e $p_{X_2} \in \mathcal{P}(\mathcal{X}_2)$ tais que $C(N_1) = I(X_1 : Y_1)$, $C(N_2) = I(X_2 : Y_2)$ e $p_{X_1,X_2} = p_{X_1} \cdot p_{X_2}$, com $Y_1 = N_1(X_1)$ e $Y_2 = N_2(X_2)$. Nesse caso, as únicas dependências são entre X_1 e Y_1, e entre X_2 e Y_2.

$$\begin{aligned}
C(N_1 \times N_2) &\geq I(X_1, X_2 : Y_1, Y_2) && (1.23)\\
&= H(Y_1, Y_2) - H(Y_1, Y_2|X_1, X_2) \\
&= H(Y_1, Y_2) - H(Y_1|X_1, X_2) - H(Y_2|X_1, X_2, Y_1) && (1.24)\\
&= H(Y_1) - H(Y_1|X_1) + H(Y_2) - H(Y_2|X_2) && (1.25)\\
&= I(X_1 : Y_1) + I(X_2 : Y_2) \\
&= C(N_1) + C(N_2),
\end{aligned}$$

onde (1.23) segue da definição de capacidade informacional como supremo, (1.24) segue da regra da cadeia generalizada e (1.25) da escolha de X_1 e X_2 (e portanto Y_1 e Y_2) independentes. □

Em particular, dado N, para qualquer $n > 0$:

$$\frac{1}{n} C(N^{\times n}) \geq C(N).$$

Como a capacidade operacional é definida em relação a múltiplos usos independentes de um canal, a capacidade informacional relevante, levando em consideração possíveis correlações entre as entradas, é dada pela seguinte definição:

Definição 1.4.3 (Capacidade Regularizada)**.** Dado um canal $(\mathcal{X}, N, \mathcal{Y})$, sua capacidade regularizada é dada por:

$$C_{reg}(N) = \sup_n \frac{1}{n} C(N^{\times n}) = \lim_{n\to\infty} \frac{1}{n} C(N^{\times n}).$$

Note que $C(N^{\times n}) \leq \log(|\mathcal{X}|^n) = n \log(|\mathcal{X}|)$, logo o limite acima está bem definido.

Se $C_{reg}(N) > C(N)$, isso significa que é possível tomar vantagem de correlações entre as entradas para transmitir informação com uma taxa maior que a capacidade de um único uso do canal. De fato, como veremos agora, isso nunca acontece. Por consequência, o

Teorema da Codificação de Canais mostra que a melhor taxa de transmissão assintótica, a capacidade operacional, é idêntica à maior quantidade de informação que podemos transmitir (em média) por um único uso do canal.

Teorema 1.4.4 (Aditividade [9]). *Dados canais* $(\mathcal{X}_1, N_1, \mathcal{Y}_1)$ *e* $(\mathcal{X}_2, N_2, \mathcal{Y}_2)$*, temos:*

$$C(N_1 \times N_2) = C(N_1) + C(N_2). \tag{1.26}$$

Em particular, por indução:

$$C_{reg}(N) = C(N). \tag{1.27}$$

Demonstração. Pelo Teorema (1.4.2), basta mostrar que $C(N_1 \times N_2) \leq C(N_1) + C(N_2)$. Considere (X_1, X_2) tal que $C(N_1 \times N_2) = I(X_1, X_2 : Y_1, Y_2)$, com $(Y_1, Y_2) = N^{\times 2}(X_1, X_2)$. Então:

$$\begin{aligned} C(N_1 \times N_2) &= I(X_1, X_2 : Y_1, Y_2) \\ &= H(Y_1, Y_2) - H(Y_1, Y_2|X_1, X_2), \end{aligned}$$

que pelo Corolário 1.1.13:

$$= H(Y_1, Y_2) - H(Y_1|X_1, X_2) - H(Y_2|X_1, X_2, Y_1),$$

e como, para $i = 1, 2$, Y_i depende somente de X_i e é condicionalmente independente de das outras variáveis aleatórias:

$$\begin{aligned} &= H(Y_1, Y_2) - H(Y_1|X_1) - H(Y_2|X_2) \\ &\leq H(Y_1) + H(Y_2) - H(Y_1|X_1) - H(Y_2|X_2) \\ &= I(X_1 : Y_1) + I(X_2 : Y_2) \\ &\leq C(N_1) + C(N_2). \end{aligned}$$

Portanto $C(N_1 \times N_2) = C(N_1) + C(N_2)$. Em particular, para qualquer canal DSM N, $C(N^{\times 2}) = 2C(N)$. Por indução, $C(N^{\times n}) = nC(N)$ para todo $n \geq 1$, logo:

$$C_{reg}(N) = \limsup_{n \to \infty} \frac{1}{n} C(N^{\times n}) = C(N).$$

□

Vamos ver no capítulo (4) que uma noção razoável de capacidade para canais quânticos não satisfaz a propriedade de aditividade. Portanto é interessante refletir sobre a demonstração do Teorema 1.4.4. Note que utilizamos a Regra da Cadeia Generalizada (Corolário 1.1.13), que não tem análogo no caso quântico (o análogo da entropia condicional na teoria quântica se comporta de forma distinta do caso clássico, devido à ausência de uma noção satisfatória de probabilidade condicional quântica). Para enfatizar a natureza clássica deste resultado, vejamos uma demonstração alternativa do Teorema 1.4.4 que não utiliza a Regra da Cadeira Generalizada mas sim outra propriedade característica da teoria clássica, o Teorema 1.1.4. A seguinte demonstração foi inspirada por um comentário em [19].

Demonstração alternativa para o Teorema 1.4.4. Sejam $p_{X_1,X_2} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$ e $(N_1 \times N_2)(p_{X_1,X_2}) = p_{Y_1,Y_2} \in \mathcal{P}(\mathcal{Y}_1 \times \mathcal{Y}_2)$. Pela expressão (1.8) para a capacidade de canais clássicos, temos que:

$$\begin{aligned} C(N_1 \times N_2) &\leq H((N_1 \times N_2)(p_{X_1,X_2})) - \sum_{x \in \mathcal{X}_1 \times \mathcal{X}_2} p_{X_1,X_2}(x) H((N_1 \times N_2)(\delta_x)) \\ &= H(p_{Y_1,Y_2}) - \sum_{x \in \mathcal{X}_1 \times \mathcal{X}_2} p_{X_1,X_2}(x) H((N_1 \times N_2)(\delta_x)) \end{aligned}$$

e como, para cada $x \in \mathcal{X}_1 \times \mathcal{X}_2$, δ_x é um ponto extremal de $\mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_1)$, pelo Teorema 1.1.4, existem $x_1 \in \mathcal{X}_1$ e $x_2 \in \mathcal{X}_2$ tais que $\delta_x = (\delta_{x_1}, \delta_{x_2})$, logo:

$$\begin{aligned} &= H(p_{Y_1,Y_2}) - \sum_{x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2} p_{X_1,X_2}(x_1, x_2) H((N_1 \times N_2)(\delta_{x_1}, \delta_{x_2})) \\ &= H(p_{Y_1,Y_2}) - \sum_{x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2} p_{X_1,X_2}(x_1, x_2) H(N_1(\delta_{x_1}), N_2(\delta_{x_2})) \\ &= H(p_{Y_1,Y_2}) - \sum_{x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2} p_{X_1,X_2}(x_1, x_2) H(N_1(x_1), N_2(x_2)) \end{aligned}$$

Por (1.22), $N_1(x_1)$ e $N_1(x_2)$ são independentes, logo:

$$\begin{aligned} &= H(p_{Y_1,Y_2}) - \sum_{x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2} p_{X_1,X_2}(x_1, x_2)(H(N_1(x_1)) + H(N_2(x_2))) \\ &= H(p_{Y_1,Y_2}) - \sum_{x_1 \in \mathcal{X}_1} p_{X_1}(x_1) H(N_1(x_1)) - \sum_{x_2 \in \mathcal{X}_2} p_{X_2}(x_2) H(N_2(x_2)) \end{aligned}$$

e pela subaditividade da entropia:

$$\begin{aligned} &\leq H(p_{Y_1}) + H(p_{Y_2}) - \sum_{x_1 \in \mathcal{X}} p_{X_1}(x_1) H(N_1(x_1)) - \sum_{x_2 \in \mathcal{X}_2} p_{X_2}(x_2) H(N_2(x_2)) \\ &= H(N_1(p_{X_1})) - \sum_{x_1 \in \mathcal{X}} p_{X_1}(x_1) H(N_1(x_1)) + H(N_2(p_{X_2})) - \sum_{x_2 \in \mathcal{X}_2} p_{X_2}(x_2) H(N_2(x_2)) \\ &= I(p_{X_1} : N_1(p_{X_1})) + I(p_{X_2} : N_2(p_{X_2})) \\ &\leq C(N_1) + C(N_2). \end{aligned}$$

□

Vamos ver no capítulo seguinte que o análogo quântico do Teorema 1.1.4 não é válido no geral devido ao fenômeno de emaranhamento na teoria quântica, que oferece uma das principais diferenças entre a teoria da informação clássica e da teoria da informação quântica.

Capítulo 2

A Teoria Quântica

Neste capítulo, desenvolvemos o mínimo necessário do formalismo da teoria quântica para compreender os conceitos de canal e entropia quânticas, seguindo principalmente [33, 11, 10, 34]. Na Seção 2.1, introduzimos o vocabulário básico e os postulados da teoria quântica. Na Seção 2.2 apresentamos sistemas quânticos compostos e emaranhamento quântico. A Seção 2.3 se dedica a canais quânticos, análogos quânticos dos canais ruidosos vistos no Capítulo 1, suas diversas representações e as relações entre elas, terminando com a apresentação da importante classe de canais *entanglement-breaking*.

2.1 Sistemas, Estados e Medições

Vamos apresentar os conceitos básicos da teoria quântica, enunciados como postulados que dão significado às noções de sistema quântico, estados, medições, etc. Começamos apresentando a notação de Dirac, particularmente conveniente em teoria da informação quântica. Muito deste capítulo serve para manter esse trabalho auto-contido, sem entrar em detalhes sobre as motivações físicas da teoria. Para uma introdução mais completa à notação de Dirac, ver por exemplo [35, 36].

2.1.1 Sistemas quânticos

Por **sistema quântico**, queremos dizer qualquer objeto de estudo da física quântica. Matematicamente, cada sistema quântico físico está associado a um espaço vetorial complexo, dito seu **espaço de estados** (essa associação não é óbvia, e é objeto central no estudo dos fundamentos da física quântica), que vamos assumir como dado para cada sistema. Na prática, o sistema servirá de índice para seu espaço de estados. Neste trabalho, vamos nos limitar a sistemas cujo espaço de estados tem dimensão finita.

Dado um sistema quântico A, denotamos por $\mathcal{H}_A$ o espaço de estados do sistema A. $\mathcal{H}_A$ é um **espaço de Hilbert**. Neste trabalho, veremos exclusivamente sistemas quânticos, e portanto espaços de Hilbert, de dimensão d_A finita, que são simplesmente espaços vetoriais complexos com produto interno. Similarmente, para um sistema quântico B, temos o espaço de estados $\mathcal{H}_B$ de dimensão d_B, etc.

Um vetor em $\mathcal{H}_A$ é escrito como um *ket* $|\psi\rangle_A$, omitindo o subscrito caso o espaço referente esteja claro pelo contexto. O produto interno de $\mathcal{H}_A$ é dado por $\langle \cdot \,|\, \cdot \rangle_A : \mathcal{H}_A \times \mathcal{H}_A \to \mathbb{C}$, linear na entrada à direita e conjugado-linear na entrada à esquerda, notação usual

em teoria quântica . Cada vetor $|\psi\rangle \in \mathcal{H}_A$ define um único funcional linear dado por $|\phi\rangle \mapsto \langle\psi|\phi\rangle$, que denotamos por um *bra* $\langle\psi| \in \mathcal{H}_A^*$. Aqui $\mathcal{H}_A^*$ é o dito **espaço vetorial dual** de $\mathcal{H}_A$.

Pelo teorema da representação de Riesz [37], podemos definir um isomorfismo antilinear isométrico $\cdot^\dagger : \mathcal{H}_A \to \mathcal{H}_A^*$, dita *adaga*, tal que:

$$(\alpha |\psi\rangle + \beta |\phi\rangle)^\dagger = \alpha^* \langle\psi| + \beta^* \langle\phi| ,$$

onde $\cdot^*$ denota conjugação complexa. Vamos utilizar o mesmo símbolo para sua inversa $\cdot^\dagger : \mathcal{H}_A^* \to \mathcal{H}_A$, que age por:

$$(\alpha \langle\tau| + \beta \langle\omega|)^\dagger = \alpha^* |\tau\rangle + \beta^* |\omega\rangle .$$

Identificando $\mathcal{H}_A \simeq \mathcal{H}_A^*$, tratamos a adaga como uma operação involutória em $\mathcal{H}_A$.

Fixamos para $\mathcal{H}_A$ (e para o espaço de estados de qualquer sistema quântico) uma base ortonormal genérica $\{|i\rangle_A\}_{i=1}^{d_A} = \{|1\rangle_A, |2\rangle_A, \ldots, |d_A\rangle_A\}$, que chamamos de **base computacional**. A ortonormalidade de $\{|i\rangle_A\}_{i=1}^{d_A}$ é expressa por $\langle i|j\rangle = \delta_{ij}$. Pela base canônica, tomamos o isomorfismo $\mathcal{H}_A \simeq \mathbb{C}^{d_A}$, identificando $\{|i\rangle\}_{i=1}^{d_A}$ com a base canônica e um vetor $|\psi\rangle \in \mathcal{H}_A$ pela matriz coluna:

$$|\psi\rangle = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_{d_A} \end{bmatrix},$$

onde $\psi_i = \langle i|\psi\rangle$, de forma que $|\psi\rangle = \sum_{i=1}^{d_A} \langle i|\psi\rangle |i\rangle$. Os coeficientes ψ_i são chamados as vezes de amplitudes de probabilidade de $|\psi\rangle$. Considerando a base dual $\{\langle i|\}_{i=1}^{d_A} \subset \mathcal{H}_A^*$, escrevemos:

$$|\psi\rangle^\dagger = \langle\psi| = \begin{bmatrix} \psi_1^* & \psi_2^* & \cdots & \psi_{d_A}^* \end{bmatrix}.$$

Assim, em coordenadas, a adaga corresponde à conjugação hermitiana. Com essas notações, temos que o produto interno corresponde, convenientemente, ao *bra-ket*:

$$\langle\psi|\phi\rangle = \begin{bmatrix} \psi_1^* & \psi_2^* & \cdots & \psi_{d_A}^* \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{d_A} \end{bmatrix} = \sum_{i=1}^{d_A} \psi_i^* \phi_i.$$

A condição de normalização de um vetor $|\psi\rangle \in \mathcal{H}_A$ pode então ser escrita como $\||\psi\rangle\|^2 = \sum_{i=1}^{d_A} |\psi_i|^2 = 1$.

Dados espaços de Hilbert $\mathcal{H}_A, \mathcal{H}_B$, denotamos por $\mathcal{B}(\mathcal{H}_A, \mathcal{H}_B)$ o espaço vetorial de operadores lineares $X : \mathcal{H}_A \to \mathcal{H}_B$, com a convenção de que escrevemos $\mathcal{B}(\mathcal{H}_A)$ para $\mathcal{B}(\mathcal{H}_A, \mathcal{H}_A)$. Denotamos por $\mathbb{1}_A$ o operador identidade de $\mathcal{H}_A$. Dado um operador $X \in \mathcal{B}(\mathcal{H}_A, \mathcal{H}_B)$, escrevemos $X(|\psi\rangle)$ como $X|\psi\rangle$ e o produto interno entre $|\phi\rangle$ e $X|\psi\rangle$ como

$$|\phi\rangle^\dagger (X |\psi\rangle) = \langle\phi|X|\psi\rangle_B ,$$

com o sanduíche no lado direito da igualdade sendo a notação de Dirac, que podemos equivalentemente interpretar como $\langle\phi| X(|\psi\rangle)$, com X agindo no *bra* pela direita. Para ver isso, note que $|\psi\rangle \mapsto \langle\phi| X |\psi\rangle_B$ define um funcional linear para cada $\langle\phi| \in \mathcal{H}_A^*$, que denotamos por $\langle\phi| X$. Assim, dado $X \in \mathcal{B}(\mathcal{H}_A, \mathcal{H}_B)$ definimos a **adjunta** de X como o operador $X^\dagger \in \mathcal{B}(\mathcal{H}_B^*, \mathcal{H}_A^*)$ tal que, para todo $|\phi\rangle \in \mathcal{H}_B$:

$$X^\dagger(\langle\phi|) = \langle X^\dagger\phi| = \langle\phi| X,$$

que nos permite expressar a propriedade familiar:

$$\langle X^\dagger\phi|\psi\rangle_A = \langle\phi|X\psi\rangle_B ,$$

para todo $|\psi\rangle \in \mathcal{H}_A$. Para $X, Y \in \mathcal{B}(\mathcal{H}_A)$, temos que $(XY)^\dagger = Y^\dagger X^\dagger$. Note a consistência na notação que nos permite pensar na adjunta de operadores como extensão da operação adaga:

$$(X |\psi\rangle)^\dagger = \langle\psi| X^\dagger.$$

De fato $|\psi\rangle^\dagger = \langle\psi|$ é o adjunto hermitiano de $|\psi\rangle$, se visto como um funcional linear em $\mathcal{H}_A^*$.

Por dualidade, podemos, equivalentemente, definir a adjunta de $X \in \mathcal{B}(\mathcal{H}_A, \mathcal{H}_B)$ como o operador $X^\dagger \in \mathcal{B}(\mathcal{H}_B, \mathcal{H}_A)$ tal que, para todo $|\psi\rangle \in \mathcal{H}_A, |\phi\rangle \in \mathcal{H}_B$:

$$\langle\phi|X|\psi\rangle_B = \langle\psi|X^\dagger|\phi\rangle_A^* ,$$

que expressa o fato de que, em termos de matrizes, a adjunta corresponde à conjugada complexa da transposta.

Exemplo 2.1.1. Para $X \in \mathcal{B}(\mathcal{H}_A)$ auto-adjunto, temos para todo $|\psi\rangle, |\phi\rangle \in \mathcal{H}_A$:

$$\langle\phi|X|\psi\rangle_B = \langle\psi|X|\phi\rangle_A^* .$$

Em particular, se $|\psi\rangle$ é um autovetor normalizado de X, com autovalor λ, temos:

$$\lambda = \langle\psi|X|\psi\rangle_B = \langle\psi|X|\psi\rangle_A^* = \lambda^*$$

Demonstrando que os autovalores de um operador auto-adjuntos são todos reais.

Denotamos por $M_{d_A \times d_B}(\mathbb{C})$ o espaço vetorial de matrizes d_A por d_B com entradas complexas. Pelas bases computacionais, podemos munir $\mathcal{B}(\mathcal{H}_A, \mathcal{H}_B) \simeq M_{d_A \times d_B}(\mathbb{C})$ de uma estrutura de espaço de Hilbert com o **produto interno de Hilbert-Schmidt** $\langle\cdot,\cdot\rangle_{HS}$ dado por:

$$\langle X, Y\rangle_{HS} = \mathrm{Tr}\left\{X^\dagger Y\right\} .$$

A notação de Dirac também nos permite expressar operadores lineares. Dados $|x\rangle \in \mathcal{H}_A$ e $|y\rangle \in \mathcal{H}_B$, definimos o operador linear *ket-bra* $|y\rangle\langle x| : \mathcal{H}_A \to \mathcal{H}_B$ por:

$$|\psi\rangle \mapsto |y\rangle \langle x|\psi\rangle = \langle x|\psi\rangle |y\rangle .$$

Podemos pensar nesse operador como aquele que leva $|x\rangle$ a $|y\rangle$. Pela notação de Dirac, é imediato que se $X = |y\rangle\langle x|$, então:

$$\langle\phi|X|\psi\rangle = \langle\phi |y\rangle\langle x| \psi\rangle = \langle x |\psi\rangle\langle\phi| y\rangle .$$

A adjunta de tal operador é dada por:

$$(|y\rangle\langle x|)^\dagger = |x\rangle\langle y|\,,$$

e a composição de tais operadores é dada por:

$$(|\phi\rangle\langle\psi|)(|\tau\rangle\langle\omega|) = |\phi\rangle\,\langle\psi|\tau\rangle\,\langle\omega| = \langle\psi|\tau\rangle\,|\phi\rangle\langle\omega|\,.$$

Sejam $\{|j\rangle_A\}_{j=1}^{d_A}, \{|i\rangle_B\}_{i=1}^{d_B}$ bases computacionais de $\mathcal{H}_A$ e $\mathcal{H}_B$ respectivamente. Então os operadores $\{|i\rangle_B\langle j|_A\}_{j=1,i=1}^{d_A,d_B}$ formam uma base ortonormal de $\mathcal{B}(\mathcal{H}_A,\mathcal{H}_B)$ sob o produto interno de Hilbert-Schmidt. Em termos de matrizes:

$$|\phi\rangle\langle\psi| = \begin{bmatrix}\phi_1\\ \phi_2\\ \vdots\\ \phi_{d_A}\end{bmatrix}\begin{bmatrix}\psi_1 & \psi_2 & \cdots & \psi_{d_A}\end{bmatrix} = \begin{bmatrix}\phi_1\psi_1 & \phi_1\psi_2 & \cdots & \phi_1\psi_{d_A}\\ \phi_2\psi_1 & \ddots & & \vdots\\ \vdots & & \ddots & \vdots\\ \phi_{d_A}\psi_1 & \cdots & \cdots & \phi_{d_A}\psi_1\end{bmatrix}.$$

A ortonormalidade da base resulta na expressão, dita *resolução da identidade*:

$$\sum_{j=1}^{d_A} |j\rangle\langle j| = \mathbb{1}_A. \tag{2.1}$$

Como um exemplo da conveniência da expressão acima, podemos decompor $X \in \mathcal{B}(\mathcal{H}_A,\mathcal{H}_B)$ como:

$$X = \sum_{i=1}^{d_B}\sum_{j=1}^{d_A} |i\rangle\langle i|\,X\,|j\rangle\langle j| = \sum_{i=1}^{d_B}\sum_{j=1}^{d_A} \langle i|X|j\rangle\,|i\rangle\langle j|\,.$$

Ou seja, $X_{ij} = \langle i|X|j\rangle$ nos dá a representação matricial de X nas bases $\{|j\rangle_A\}_{j=1}^{d_A}, \{|i\rangle_B\}_{i=1}^{d_B}$. Por exemplo, podemos facilmente verificar que $X^\dagger = \overline{X}^T$:

$$X^\dagger_{ij} = \langle i|X^\dagger|j\rangle = \langle j|X|i\rangle^* = X^*_{ji}.$$

A representação matricial de um produto XY, para $X, Y \in \mathcal{B}(\mathcal{H}_A)$ é obtida aplicando 2.1:

$$\langle i|XY|j\rangle = \sum_{k=1}^{d_A} \langle i|X|k\rangle\,\langle k|Y|j\rangle = \sum_{k=1}^{d_A} X_{ik}Y_{kj}.$$

Exemplo 2.1.2. Seja $U \in \mathcal{B}(\mathcal{H}_A)$ unitário, ou seja, vale que $UU^\dagger = U^\dagger U = \mathbb{1}_A$. Lembrando que $U\,\langle\phi| = \langle\phi|\,U^\dagger$, vemos que U preserva o produto interno:

$$\langle U\phi|U\psi\rangle = \langle\phi|U^\dagger U|\psi\rangle = \langle\phi|\psi\rangle$$

e age em operadores de posto um (que são exatamente os *ket-bra*s) por:

$$|\psi\rangle\langle\phi| \mapsto U\,|\psi\rangle\langle\phi|\,U^\dagger.$$

Se $|\psi\rangle$ é um autovetor normalizado de U para o autovalor λ, então:

$$\|\lambda\|^2 = \lambda^*\lambda = \langle\psi|U^\dagger U|\psi\rangle = \langle\psi|\psi\rangle = 1.$$

Dadas duas bases $\{|\phi_j\rangle\}_{j=1}^{d_A}, \{|\psi_j\rangle\}_{j=1}^{d_A}$ de $\mathcal{H}_A$, existe um operador unitário U tal que $|\phi_j\rangle \mapsto |\psi_j\rangle$. De fato:

$$U = \sum_{j=1}^{d_A} |\psi_j\rangle\langle\phi_j|\,.$$

Seja $X \in \mathcal{B}(\mathcal{H}_A)$ um operador normal (ou seja, $X^\dagger X = XX^\dagger$). Pelo Teorema Espectral para Matrizes, existe uma base ortonormal de $\{|\alpha_i\rangle\}_{i=1}^{d_A}$ de autovetores de X, de forma que:

$$X = \sum_{i=1}^{d_A} a_i \, |\alpha_i\rangle\langle\alpha_i| \, ,$$

onde a_i são autovalores de X. Note que operadores do tipo $|x\rangle\langle x|$ são projeções ortogonais no espaço unidimensional gerado pelo vetor $|x\rangle$.

2.1.2 Estados quânticos

Por estado de um sistema quântico, nos referimos a um objeto que descreva o sistema em um dado momento. A noção de estado quântico vai exercer um papel na teoria quântica análogo às variáveis aleatórias na teoria da informação clássica, descrevendo não só uma configuração do sistema, mas uma densidade de possíveis configurações.

O traço de um operador X é dado, em notação de Dirac, por:

$$\mathrm{Tr}\{X\} = \sum_{j=1}^{d_A} \langle j|X|j\rangle \, .$$

Notando que valem:

$$\mathrm{Tr}\left\{X^\dagger\right\} = \mathrm{Tr}\{X\}^* \, ,$$
$$\mathrm{Tr}\{|\phi\rangle\langle\psi|\} = \langle\psi|\phi\rangle \, ,$$

e

$$\mathrm{Tr}\{|\phi\rangle\langle\psi| \, X\} = \langle\psi|X|\phi\rangle \, .$$

Um operador $X \in \mathcal{B}(\mathcal{H}_A)$ é dito positivo se, para todo $|\psi\rangle \in \mathcal{H}_A$, $\langle\psi|X|\psi\rangle \geq 0$. Todo operador positivo é hermitiano e seus autovalores são todos não-negativos.

> **Primeiro Postulado:** Um estado de um sistema quântico A corresponde a um operador $\rho \in \mathcal{B}(\mathcal{H}_A)$ tal que $\rho \geq 0$ e $\mathrm{Tr}\{\rho\} = 1$. Uma matriz para o estado em alguma base de $\mathcal{H}_A$ é dita uma **matriz densidade** para o estado.

Não devemos confundir essa noção de estado com a expressão *espaço de estados* para $\mathcal{H}_A$. Nem todo vetor em $\mathcal{H}_A$ corresponde a um estado. Denotamos por $\mathcal{D}(\mathcal{H}_A)$ o conjunto de estados de $\mathcal{H}_A$, ou seja:

$$\mathcal{D}(\mathcal{H}_A) = \{\rho \in \mathcal{B}(\mathcal{H}_A) \,|\, \rho \geq 0 \text{ e } \mathrm{Tr}\{\rho\} = 1\}.$$

Chamamos de **estado puro** qualquer estado que pode ser expresso por uma matriz densidade do tipo $\rho = |\psi\rangle\langle\psi|$, onde $|\psi\rangle$ é um vetor normalizado de $\mathcal{H}_A$. Note que vetores normalizados que diferem apenas por uma fase, como $|\psi\rangle$ e $e^{i\theta}|\psi\rangle$, para algum $\theta \in \mathbb{R}$, correspondem ao mesmo estado:

$$e^{i\theta} \, |\psi\rangle\langle\psi| \, e^{-i\theta} = |\psi\rangle\langle\psi| \, .$$

Quando conveniente, vamos identificar o estado puro $|\psi\rangle\langle\psi| \in \mathcal{B}(\mathcal{H}_A)$ com o vetor normalizado $|\psi\rangle \in \mathcal{H}_A$.

Chamamos de **ensemble** uma mistura estatística de estados puros $\{p_X(x), |\psi_x\rangle\}_{x\in\mathcal{X}}$, onde X é uma variável aleatória com valores em conjunto de índices $\mathcal{X}$ com distribuição p_X e $|\psi_x\rangle$ são vetores normalizados associados ao índice $x \in \mathcal{X}$. Como veremos, ensembles surgem naturalmente na teoria quântica quando temos somente conhecimento parcial sobre um sistema. Associamos ao ensemble $\{p_X(x), |\psi_x\rangle\}_{x\in\mathcal{X}}$ a matriz densidade:

$$\rho = \mathbb{E}_X\{|\psi_X\rangle\langle\psi_X|\} = \sum_{x\in\mathcal{X}} p_x \, |\psi_x\rangle \, \langle\psi_x| \, . \tag{2.2}$$

Estados descritos por uma matriz densidade da forma (2.2) são ditos **estados mistos** (estado puros sendo um caso particular, quando ρ tem posto um). Um mesmo estado misto pode corresponder a infinitos ensembles distintos, correspondentes às distintas maneiras de escrever o operador densidade do estado como uma combinação convexa de estados puros [11]. Reciprocamente, é fácil ver que para qualquer matriz densidade ρ existe um ensemble de estados puros $\{p_X(x), |\psi_x\rangle\}_{x\in\mathcal{X}}$ tal que vale (2.2) (basta considerar a decomposição espectral).

Proposição 2.1.3. *O conjunto $\mathcal{D}(\mathcal{H}_A)$ é convexo e seus elementos extremais são os estados puros.*

Para uma demonstração, ver [13]. No que segue, também vamos considerar ensembles de estados mistos (ensembles de ensembles), já que a matriz densidade do estado correspondente à tal ensemble é simplesmente uma combinação convexa de estados mistos, que também é um estado.

Definição 2.1.4 (Pureza). A pureza de uma matriz densidade $\rho \in \mathcal{D}(\mathcal{H}_A)$ é dada por:

$$P(\rho) = \mathrm{Tr}\left\{\rho^2\right\}.$$

Exemplo 2.1.5 (Estado Maximamente Misto)**.** O estado de A cuja matriz densidade é dada por $\frac{1}{d}\mathbb{1}_A$ é dito o **estado maximamente misto**. A pureza deste estado é $\frac{1}{d}$.

Proposição 2.1.6. *Dado um estado $\rho \in \mathcal{D}(\mathcal{H}_A)$, $P(\rho) = 1$ se, e somente se, ρ é um estado puro.*

Exemplo 2.1.7 (qubit)**.** Neste caso $\dim \mathcal{H}_A = 2$, logo $\mathcal{B}(\mathcal{H}_A) \simeq M_{2\times 2}(\mathbb{C})$ e uma base (ortogonal, pelo produto interno de Hilbert-Schmidt) conveniente é dada por:

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma_X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

onde σ_X, σ_Y e σ_Z são ditas *matrizes de Pauli*, todas unitárias e hermitianas, logo qualquer operador hermitiano é uma combinação linear nesta base com coeficientes reais. Em particular, como $\mathrm{Tr}\{\mathbb{1}_A\} = 2$ e as matrizes da Pauli tem traço nulo, uma matriz densidade $\rho \in \mathcal{B}(\mathcal{H}_A)$ pode ser escrita como:

$$\rho = \frac{1}{2}(\mathbb{1}_A + x\sigma_X + y\sigma_Y + z\sigma_Z) = \frac{1}{2}\begin{bmatrix} 1+z & x-iy \\ x+iy & 1-z \end{bmatrix},$$

onde $1/2$ é o fator normalizador e o vetor $\sigma = (x, y, z) \in \mathbb{R}^3$ é dito *vetor de Bloch*. A condição $\det(\rho) \geq 0$ implica que

$$\|\sigma\|^2 = x^2 + y^2 + z^2 \leq 1.$$

Logo, existe uma correspondência bijetiva entre $\mathcal{D}(\mathcal{H}_A)$ e a bola unitária do $\mathbb{R}^3$, que neste contexto chamamos de *esfera de Bloch*. Temos que:

$$\mathrm{Tr}\left\{\rho^2\right\} = \frac{1}{2}(1 + x^2 + y^2 + z^2),$$

de forma que os estados puros correspondem, pelo Teorema 2.1.6, à fronteira da esfera de Bloch, enquanto que os estados mistos formam seu interior. O estado correspondendo ao centro da esfera é o estado maximamente misto:

$$\frac{1}{2}\mathbb{1}_A = \frac{1}{2}\left|0\right\rangle\!\left\langle 0\right| + \frac{1}{2}\left|1\right\rangle\!\left\langle 1\right|.$$

2.1.3 Evolução Unitária

Vamos agora descrever a dinâmica de sistemas quânticos, que podem evoluir de forma determinística, familiar da física clássica, como também de uma forma não-determinística e fundamentalmente quântica, através de medições. Mais adiante vamos considerar evoluções estocásticas, que incluem evoluções determinísticas e medições como casos particulares.

> **Segundo Postulado:** A evolução determinística de um sistema quântico é descrita por operadores unitários.

Pelo Exemplo 2.1.2, um operador unitário U age em um estado puro $|\psi\rangle\langle\psi| \in \mathcal{D}(\mathcal{H}_A)$ por:

$$|\psi\rangle\langle\psi| \mapsto U\,|\psi\rangle\langle\psi|\,U^\dagger.$$

Similarmente, um estado $\rho \in \mathcal{D}(\mathcal{H}_A)$ evolui por U como $\rho \mapsto U\rho U^\dagger$.

No contexto de computação quântica, evoluções unitárias são chamadas de portas lógicas quânticas, e constituem as operações elementares de um computador quântico [33].

Exemplo 2.1.8. É fácil ver que as matrizes de Pauli (Exemplo 2.1.7) são operadores unitários em $\mathcal{B}(\mathbb{C}^2)$. Em computação quântica, correspondem às portas X, Y e Z.

2.1.4 Medições

Estados quânticos são caracterizados pelas informações que carregam sobre o sistema. Características mensuráveis de um sistema são chamadas de **observáveis**, cujos valores são obtidos por um processo de medição. Diferentemente da física clássica, onde o valor de um observável é completamente determinado pelo estado do sistema, na física quântica o estado do sistema caracteriza apenas a probabilidade de obter um determinado resultado ao realizar uma medição. Além disso, nem toda medição preserva o sistema, no sentido de que possamos falar do estado do sistema após a medição. Vejamos o formalismo adequado para lidar com medições de sistemas quânticos.

Definição 2.1.9 (PVM)**.** Chamamos de **medição projetiva** ou **PVM** (*Projective Valued Measure*) sobre A uma função $\Pi : \mathcal{X} \to \mathcal{B}(\mathcal{H}_A)^+$, para um conjunto finito $\mathcal{X}$, cuja imagem consiste em projeções ($\Pi_x^2 = \Pi_x$) ortogonais entre si ($\Pi_x\Pi_y = 0$ para $x \neq y$), tais que:

$$\sum_{x\in\mathcal{X}} \Pi_x = \mathbb{1}_A.$$

Também descrevemos Π como um conjunto de projeções $\{\Pi_x\}_{x\in\mathcal{X}} \subset \mathcal{B}(\mathcal{H}_A)$.

Terceiro Postulado: Se um sistema quântico A se encontra no estado puro $|\psi\rangle\langle\psi| \in \mathcal{D}(\mathcal{H}_A)$, a probabilidade associada ao resultado x ao realizar a medição $\{\Pi_x\}_{x\in\mathcal{X}}$ é dada por:

$$p_x = \langle\psi|\Pi_x|\psi\rangle = \mathrm{Tr}\left\{|\psi\rangle\langle\psi|\,\Pi_x\right\}.$$

Se o resultado x foi obtido, então o estado pós-medição do sistema é:

$$|\psi_x\rangle\langle\psi_x| = \frac{\Pi_x\,|\psi\rangle\langle\psi|\,\Pi_x}{\langle\psi|\Pi_x|\psi\rangle}.$$

Se $|\mathcal{X}| = d_A$, então $\Pi_x = |\psi_x\rangle\langle\psi_x|$ para alguma base $\{|\psi_x\rangle\}_{x\in\mathcal{X}}$ de $\mathcal{H}_A$. Neste caso dizemos que o PVM é uma **medição com respeito à base** $\{|\psi_x\rangle\}_{x\in\mathcal{X}}$.

Se o estado de A é dado por $\rho \in \mathcal{D}(\mathcal{H}_A)$, a medição $\{\Pi_x\}_{x\in\mathcal{X}}$ nos dá o resultado x com probabilidade:

$$p_x = \langle\rho, \Pi_x\rangle_{HS} = \mathrm{Tr}\left\{\rho\Pi_x\right\},$$

e neste caso o estado pós-medição do sistema é:

$$\rho_x = \frac{\Pi_x\rho\Pi_x}{\mathrm{Tr}\left\{\rho\Pi_x\right\}}.$$

Na física quântica, é comum considerar o PVM associado a um observável X dado por $\{|x_n\rangle\langle x_n|\}_{n=1}^{d_A}$, onde, para cada $1 \leq n \leq d_A$, $|x_n\rangle$ é um autovetor normalizado com autovalor $\lambda_n \in \mathbb{R}^+$.

PVMs não são a noção mais geral de medição na teoria quântica. Na prática, vamos considerar medições como dadas pela seguinte definição, que inclui medições que oferecem somente informação parcial do sistema:

Definição 2.1.10 (POVM)**.** Chamamos de **medição generalizada** ou **POVM** (*Positive Operator Valued Measure*) sobre A uma função $\Lambda : \mathcal{X} \to \mathcal{B}(\mathcal{H}_A)^+$, com $\mathcal{X}$ um conjunto finito e tal que:

$$\sum_{x\in\mathcal{X}} \Lambda_x = \mathbb{1}_A.$$

No que segue vamos denotar um POVM Λ por sua imagem $\{\Lambda_x\}_{x\in\mathcal{X}}$. Se o estado de A é dado por $\rho \in \mathcal{D}(\mathcal{H}_A)$, o POVM $\{\Lambda_x\}_{x\in\mathcal{X}}$ nos dá o resultado x com probabilidade

$$p(x) = \langle\rho, \Lambda_x\rangle_{HS} = \mathrm{Tr}\left\{\rho\Lambda_x\right\},$$

e o estado pós-medição é considerado indeterminado, o que leva alguns autores a chamar POVMs de medições destrutivas.

No que segue, sempre que falarmos de medição, nos referimos a medições generalizadas. Medições generalizadas em um sistema quântico sempre correspondem a uma medição projetiva em um sistema maior (ver *Naimark's theorem* em [12]).

2.2 Sistemas Compostos

Outro aspecto em que a física quântica difere da física clássica é na descrição de sistemas compostos. No geral, dois sistemas quânticos que entram em contato não podem mais ser

descritos individualmente. Dados sistemas quânticos A, B e C, podemos considerá-los em conjunto como um único sistema quântico, dito o **sistema composto** ABC, e chamamos os sistemas constituintes de subsistemas. Um sistema composto de dois subsistemas é dito um **sistema bipartite**. No caso do sistema composto ABC podemos identificar os subsistemas bipartites $A|BC, AB|C$ e $AC|B$, ditas bipartições de ABC.

2.2.1 Produto Tensorial

> **Quarto postulado**: Dados sistemas quânticos A e B, o espaço de estados do sistema bipartite AB é dado pelo produto tensorial $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$.

Para mais detalhes sobre o produto tensorial de espaços vetoriais, ver, por exemplo, [38].

Os estados puros de $\mathcal{H}_A \otimes \mathcal{H}_B$ são do tipo:

$$|\psi\rangle\langle\psi|_{AB} = |\varphi\rangle\langle\varphi|_A \otimes |\phi\rangle\langle\phi|_B .$$

Para as bases computacionais $\{|i\rangle_A\}_{i=1}^{d_A}, \{|j\rangle_B\}_{j=1}^{d_B}$ de $\mathcal{H}_A$ e $\mathcal{H}_B$ respectivamente, temos que $\{|i\rangle_A \otimes |j\rangle_B\}_{i=1,j=1}^{d_A,d_B}$ é base ortonormal de $\mathcal{H}_{AB}$ (com $\langle\cdot|\cdot\rangle_{AB} = \langle\cdot|\cdot\rangle_A \langle\cdot|\cdot\rangle_B$). Também usamos a notação $|i,j\rangle_{AB} := |i\rangle_A \otimes |j\rangle_B$ quando conveniente. Definimos a base computacional de $\mathcal{H}_A \otimes \mathcal{H}_B$ como $\{|i,j\rangle_{AB}\}_{i=1,k=1}^{d_A,d_B}$, ordenada pela ordem lexicográfica, de forma que o produto tensorial de vetores corresponde ao produto de Kronecker de suas representações matriciais:

$$|\psi\rangle_A \otimes |\phi\rangle_B = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_{d_A} \end{bmatrix} \otimes \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{d_B} \end{bmatrix} = \begin{bmatrix} \psi_1\phi_1 \\ \psi_1\phi_2 \\ \vdots \\ \psi_1\phi_{d_B} \\ \psi_2\phi_1 \\ \vdots \\ \psi_{d_A}\phi_{d_B} \end{bmatrix} .$$

Analogamente, para $X \in \mathcal{B}(\mathcal{H}_A)$ e $Y \in \mathcal{B}(\mathcal{H}_B)$, o produto tensorial de operadores $X \otimes Y \in \mathcal{B}(\mathcal{H}_A) \otimes \mathcal{B}(\mathcal{H}_B) \simeq \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ é dado por $(X \otimes Y)(|\psi\rangle \otimes |\phi\rangle) := X|\psi\rangle \otimes Y|\phi\rangle$, e matricialmente pelo produto de Kronecker:

$$X \otimes Y = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1d_A} \\ a_{21} & a_{22} & \dots & a_{2d_A} \\ \vdots & \vdots & \ddots & \vdots \\ a_{d_A1} & a_{d_A2} & \dots & a_{d_Ad_A} \end{bmatrix} \otimes Y = \begin{bmatrix} a_{11}Y & a_{12}Y & \dots & a_{1d_A}Y \\ a_{21}Y & a_{22}Y & \dots & a_{2d_A}Y \\ \vdots & \vdots & \ddots & \vdots \\ a_{d_A1}Y & a_{d_A2}Y & \dots & a_{d_Ad_A}Y \end{bmatrix} .$$

De fato, considerando $\mathcal{H}_A \simeq \mathbb{C}^{d_A}$, temos para cada elemento $z \in \mathbb{C}^{d_A} \otimes \mathcal{H}_B$ uma única representação $z = \sum_{i=1}^{d_A} |i\rangle \otimes b_i$, com $b_i \in \mathcal{H}_B$. Dessa forma:

$$\|z\|^2 = \left\langle \sum_{i=1}^{d_A} |i\rangle \otimes b_i \middle| \sum_{j=1}^{d_A} |j\rangle \otimes b_j \right\rangle = \sum_{i=1}^{d_A} \langle i|i\rangle \langle b_i|b_i\rangle = \sum_{i=1}^{d_A} \|b_i\|^2 ,$$

que evidencia um isomorfismo de espaços de Hilbert:

$$\mathbb{C}^{d_A} \otimes \mathcal{H}_B \simeq \underbrace{\mathcal{H}_B \oplus \mathcal{H}_B \cdots \oplus \mathcal{H}_B}_{n \text{ vezes}} = \mathcal{H}_B^{(d_A)}.$$

De forma que a representação matricial de z, como vetor em $\mathcal{H}_B^{(d_A)}$ é:

$$z = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{d_A} \end{bmatrix}.$$

Podemos similarmente identificar $\mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \simeq \mathcal{B}(\mathcal{H}_A) \otimes \mathcal{B}(\mathcal{H}_B) \simeq M_{d_A}(\mathbb{C}) \otimes \mathcal{B}(\mathcal{H}_B) \simeq M_{d_A}(\mathcal{B}(\mathcal{H}_B)) \simeq \mathcal{B}(\mathcal{H}_B^{(d_A)})$, onde $M_n(\mathcal{B}(\mathcal{H}_B))$ denota, para cada $n \geq 0$, a álgebra de matrizes $n \times n$ com entradas em $\mathcal{B}(\mathcal{H}_B)$, com norma sendo a correspondente norma de operadores em $\mathcal{B}(\mathcal{H}_B^{(d_A)})$. Assim, $Z = (Z_{ij}) \in M_{d_A}(\mathcal{B}(\mathcal{H}_B))$ é identificado com

$$Z = \sum_{i,j=1}^{d_A} |i\rangle\langle j| \otimes Z_{ij} \in M_{d_A} \otimes \mathcal{B}(\mathcal{H}_B).$$

Agora, para $X \in \mathcal{B}(\mathcal{H}_A)$, $Y \in \mathcal{B}(\mathcal{H}_B)$ temos:

$$\begin{aligned} X \otimes Y &= \left(\sum_{i,j=1}^{d_A} x_{ij} |i\rangle\langle j| \right) \otimes \left(\sum_{k,l=1}^{d_B} y_{kl} |k\rangle\langle l| \right) \\ &= \sum_{i,j=1}^{d_A} |i\rangle\langle j| \otimes \left(x_{ij} \sum_{k,l=1}^{d_B} y_{kl} |k\rangle\langle l| \right), \end{aligned}$$

que é exatamente o produto de Kronecker entre X e Y. Para mais detalhes sobre álgebras de matrizes, ver [39].

Definição 2.2.1 (Operador Local)**.** Sejam A, B, C e D sistemas quânticos. Chamamos de **operador local** um elemento de $\mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B, \mathcal{H}_C \otimes \mathcal{H}_D)$ do tipo

$$X^A = X \otimes \mathbb{1}_B \qquad \text{ou} \qquad Y^B = \mathbb{1}_A \otimes Y,$$

para $X \in \mathcal{B}(\mathcal{H}_A, \mathcal{H}_C), Y \in \mathcal{B}(\mathcal{H}_B, \mathcal{H}_D)$. Dizemos que X^A é um operador local de A e que Y^B é um operador local de B.

Operadores locais afetam somente um subsistema de um sistema composto. Note que operadores como X^A e Y^B comutam ($X^A Y^B = Y^B X^A = X \otimes Y$).

O isomorfismo $\mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \simeq M_{d_A}(\mathcal{H}_B^{(d_A)})$ nos permite identificar uma relação de ordem em $\mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$, e determinar seus operadores positivos $\mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)^+$:

Proposição 2.2.2. *$X = (X_{ij}) \in M_{d_A}(\mathcal{H}_B^{(d_A)})$ é positiva se e somente se para quaisquer d_A elementos $x_1, x_2, \ldots, x_{d_A} \in \mathcal{H}_B$, a matriz $(\langle x_i | X | x_j \rangle)_{i,j}$ é positiva.*

Proposição 2.2.3. *Dados $X \in \mathcal{B}(\mathcal{H}_\mathcal{A}), Y \in \mathcal{B}(\mathcal{H}_\mathcal{B})$:*

- $\operatorname{Tr}\{X \otimes Y\} = \operatorname{Tr}\{X\} \operatorname{Tr}\{Y\}$.
- *Se $\{x_i\}_{i=1}^{d_A}$ são os autovalores de X e $\{y_j\}_{j=1}^{d_B}$ são os autovalores de Y, então os autovalores de $X \otimes Y$ são $\{x_i y_j\}_{i,j=1}^{d_A, d_B}$.*

2.2.2 Decomposição de Schmidt

Nesta seção introduzimos uma maneira de identificar estados de um sistema bipartite com operadores lineares, permitindo a transferência de propriedades e decomposições de operadores para vetores no produto tensorial.

Definição 2.2.4 (Vetorização)**.** A **vetorização** é o mapa $\mathrm{Vec} : \mathcal{B}(\mathcal{H}_A, \mathcal{H}_B) \to \mathcal{H}_A \otimes \mathcal{H}_B$, dado pela extensão linear de

$$\mathrm{Vec}(|\phi\rangle\langle\psi|) = \overline{|\psi\rangle} \otimes |\phi\rangle\,,$$

Para $|\psi\rangle \in \mathcal{H}_A$, $|\phi\rangle \in \mathcal{H}_B$.

Proposição 2.2.5. $\mathrm{Vec} : \mathcal{B}(\mathcal{H}_A, \mathcal{H}_\mathcal{B}) \to \mathcal{H}_A \otimes \mathcal{H}_B$ *é um isomorfismo de espaços de Hilbert, considerando* $\mathcal{B}(\mathcal{H}_A, \mathcal{H}_\mathcal{B})$ *com o produto interno de Hilbert-Schmidt. Ou seja, para* $X, Y \in \mathcal{B}(\mathcal{H}_A)$*:*

$$\langle X, Y\rangle_{HS} = \mathrm{Tr}\left\{X^\dagger Y\right\} = \overline{\mathrm{Vec}(X)}^T \mathrm{Vec}(Y) = \langle \mathrm{Vec}(X), \mathrm{Vec}(Y)\rangle_{AB}.$$

Pela identificação $\mathbb{C}^{d_A} \otimes \mathcal{H}_B \simeq \mathcal{H}_B^{(d_A)}$, dada pelo produto de Kronecker, temos que:

$$\mathrm{Vec}(X) = \sum_{j=1}^{d_A} |j\rangle_A \otimes X\,|j\rangle_A = \begin{bmatrix} X\,|1\rangle \\ X\,|2\rangle \\ \vdots \\ X\,|d_A\rangle \end{bmatrix}, \tag{2.3}$$

ou seja, a vetorização representa um operador $X : \mathbb{C}^{d_A} \to \mathbb{C}^{d_B}$ como um vetor $\mathrm{Vec}(X) \in \mathbb{C}^{d_A d_B}$.

Vamos agora usar a vetorização para transferir para estados bipartites a noção de posto e a decomposição do valor singular.

Proposição 2.2.6 (Decomposição de Schmidt)**.** *Para qualquer vetor* $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$*, existem bases ortonormais* $\{|a_i\rangle\}_{i=1}^{d_A}$ *e* $\{|b_j\rangle\}_{j=1}^{d_B}$ *de* $\mathcal{H}_A$ *e* $\mathcal{H}_B$ *respectivamente, tais que*

$$|\psi\rangle_{AB} = \sum_{k=1}^{\min(d_A, d_B)} \sqrt{\lambda_k}\,|a_k\rangle \otimes |b_k\rangle\,, \tag{2.4}$$

para $\sqrt{\lambda_k} \in \mathbb{R}^+$ *ditos* ***coeficientes de Schmidt****. A expressão* (2.4) *é dita a* ***decomposição de Schmidt*** *de* $|\psi\rangle_{AB}$*, e o número de coeficientes de Schmidt não-nulos em* (2.4) *é dito o* ***rank de Schmidt*** *de* $|\psi\rangle_{AB}$*, denotado por* $\mathrm{SR}(|\psi\rangle_{AB})$*. Se* $|\psi\rangle_{AB}$ *representa um estado puro, então* $\sum_{k=1} \lambda_k = 1$*.*

A decomposição de Schmidt de $|\psi\rangle_{AB}$ é a vetorização da decomposição do valor singular de $X_{AB} = \mathrm{Vec}^{-1}(|\psi\rangle_{AB}) \in \mathcal{B}(\mathcal{H}_A, \mathcal{H}_\mathcal{B})$ [10]. Desta forma obtemos que, para qualquer $X \in \mathcal{B}(\mathcal{H}_A, \mathcal{H}_\mathcal{B})$:

$$\mathrm{rk}(X) = \mathrm{SR}(\mathrm{Vec}(X)). \tag{2.5}$$

2.2.3 Emaranhamento quântico

Vejamos que existem estados puros de sistemas compostos que não podem ser expressos como produto tensorial de estados dos subespaços constituintes.

Definição 2.2.7 (Estado Produto)**.** Chamamos de **estados produto** os estados de $\mathcal{H}_A \otimes \mathcal{H}_B$ que podem ser escritos como $\rho^A \otimes \sigma^B$, com $\rho^A \in \mathcal{D}(\mathcal{H}_A)$, $\sigma^B \in \mathcal{D}(\mathcal{H}_B)$.

O estado $\rho^A \otimes \sigma^B \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ é a matriz densidade do ensemble

$$\{p_X(x) \cdot p_Y(y), |\psi_x\rangle \otimes |\varphi_y\rangle\}_{x\in\mathcal{X}, y\in\mathcal{Y}},$$

onde $\{p_X(x), |\psi_x\rangle\}_{x\in\mathcal{X}}$ tem matriz densidade ρ^A e $\{p_Y(y), |\varphi_y\rangle\}_{y\in\mathcal{Y}}$ tem matriz densidade σ^B. Ou seja, estados produto correspondem ao produto tensorial de estados mistos estatisticamente independentes.

Definição 2.2.8 (Estado Separável)**.** Chamamos $\tau^{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ de **estado separável** se pode ser escrito como combinação convexa de estados produto:

$$\tau^{AB} = \sum_{i=1}^{K} p_i \, \rho_i^A \otimes \sigma_i^B, \quad \text{tal que } \sum_{i=1}^{K} p_i = 1, \tag{2.6}$$

para algum $K \geq 1$, $\rho_i^A \in \mathcal{D}(\mathcal{H}_A), \sigma_i^B \in \mathcal{D}(\mathcal{H}_B)$ para cada $i \leq K$.

Proposição 2.2.9 ([10])**.** *Um estado separável $\tau^{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ sempre pode ser escrito na forma* (2.6) *tomando ρ_i^A e σ_i^B estados puros.*

Assim, estados separáveis correspondem às matrizes densidade de ensembles do tipo $\{p_X(x), |\psi_x\rangle \otimes |\varphi_x\rangle\}_{x\in\Sigma}$, implicando que o estado dos sistemas A e B estão correlacionados classicamente.

A existência de estados do sistema composto AB que não são separáveis é necessária para a representação de correlações entre os sistemas A e B que são essencialmente não-clássicas, como é o caso no fenômeno de emaranhamento quântico. É o que nos leva a definir o espaço de estados de AB como o produto tensorial, e não a soma direta, de $\mathcal{H}_A$ e $\mathcal{H}_B$.

Definição 2.2.10 (Estado Emaranhado)**.** Um estado de um sistema composto AB que não é separável é chamado de **estado emaranhado**. A correlação não-clássica resultante entre os sistemas é o que nos referimos como **emaranhamento quântico**.

Exemplo 2.2.11 (Estado Maximamente Emaranhado Canônico)**.** Seja $d_A \geq 2$. Considere o vetor $|\Omega_{d_A}\rangle \in \mathcal{H}_A$, dado por:

$$|\Omega_{d_A}\rangle = \mathrm{Vec}(\mathbb{1}_{\mathbb{C}^{d_A}}) = \sum_{i=1}^{d_A} |i\rangle \otimes |i\rangle \in \mathbb{C}^{d_A} \otimes \mathbb{C}^{d_A}. \tag{2.7}$$

Chamamos de *estado maximamente emaranhado canônico* o estado puro:

$$\omega_{d_A} = \frac{1}{d_A} |\Omega_{d_A}\rangle\langle\Omega_{d_A}| \in \mathcal{D}(\mathcal{H}_A).$$

Vejamos que ω_{d_A} é emaranhado. De fato, suponha que ω_{d_A} é separável. Então existem estados puros $\sigma_i, \tau_i \in \mathcal{D}(\mathcal{H}_A)$, $0 < i \leq d_A$, tais que:

$$\omega_{d_A} = \sum_{i=1}^{d_A} p_i \sigma_i \otimes \tau_i,$$

com $p_i \geq 0$, $\sum p_i = 1$. Como ω_{d_A} é puro, pela Proposição 2.1.3, podemos assumir que $\omega_{d_A} = \sigma_1 \otimes \tau_1$, e portanto que $\mathrm{SR}(\omega_{d_A}) = 1$. Mas por (2.5) temos que $\mathrm{SR}(\omega_{d_A}) = \mathrm{rk}(\mathbb{1}_A) = d_A$. Logo, se $d_A \geq 2$, ω_{d_A} é emaranhado.

Seguindo o exemplo acima, é fácil ver que vale o seguinte:

Proposição 2.2.12. *Um estado puro $\rho^{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_A)$ é separável se e somente se $\mathrm{SR}(\rho) = 1$.*

Ou seja, os coeficientes de Schmidt de um estado bipartite determinam se o estado é emaranhado. Como os coeficientes de Schmidt são invariantes perante operadores unitários locais, obtemos:

Proposição 2.2.13. *$\rho^{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_A)$ é separável se e somente se para quaisquer $U \in \mathcal{B}(\mathcal{H}_A)$ e $V \in \mathcal{B}(\mathcal{H}_B)$ unitários, $(U \otimes V)\rho^{AB}$ é separável.*

A existência de estados puros emaranhados de sistemas quânticos bipartites é uma das principais características da teoria quântica. Em contraste com o Teorema 1.1.4, obtemos:

Corolário 2.2.14. *Para sistemas quânticos A e B, se $d_A > 1$ ou $d_B > 1$:*

$$\mathrm{Ext}(\mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)) \supsetneq \mathrm{conv}(\mathrm{Ext}(\mathcal{D}(\mathcal{H}_A)), \mathrm{Ext}(\mathcal{D}(\mathcal{H}_B))),$$

onde conv *denota o fecho convexo.*

2.2.4 Traço Parcial

Vejamos como obter o estado de um subsistema a partir do estado do sistema quântico composto.

Definição 2.2.15 (Traço Parcial)**.** Dado um sistema bipartite AB, definimos os **traços parciais** $\mathrm{Tr}_A : \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_B)$ e $\mathrm{Tr}_B : \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A)$ como as extensões lineares das funções satisfazendo:

$$\mathrm{Tr}_A \{X \otimes Y\} = \mathrm{Tr}\{X\} Y, \qquad \mathrm{Tr}_B \{X \otimes Y\} = \mathrm{Tr}\{Y\} X. \tag{2.8}$$

Para $\rho^{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$, chamamos de **estados locais** de ρ^{AB} os estados:

$$\rho^A = \mathrm{Tr}_B \left\{\rho^{AB}\right\} \in \mathcal{D}(\mathcal{H}_A) \text{ e } \rho^B = \mathrm{Tr}_A \left\{\rho^{AB}\right\} \in \mathcal{D}(\mathcal{H}_B).$$

Em particular, para $|\phi\rangle, |\psi\rangle \in \mathcal{H}_A, |\varphi\rangle, |\omega\rangle \in \mathcal{H}_B$, temos:

$$\mathrm{Tr}_B \{|\phi\rangle\langle\psi| \otimes |\omega\rangle\langle\varphi|\} = |\phi\rangle\langle\psi| \langle\varphi|\omega\rangle, \quad \mathrm{Tr}_A \{|\phi\rangle\langle\psi| \otimes |\omega\rangle\langle\varphi|\} = \langle\psi|\phi\rangle |\omega\rangle\langle\varphi|.$$

A partir de (2.8), podemos verificar que, para $\rho^{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$:

$$\mathrm{Tr}_A \left\{\rho^{AB}\right\} = \sum_{i=1}^{d_A} (\langle i| \otimes \mathrm{id}_B)\rho^{AB}(|i\rangle \otimes \mathrm{id}_B), \quad \mathrm{Tr}_B \left\{\rho^{AB}\right\} = \sum_{j=1}^{d_B} (\mathrm{id}_A \otimes \langle j|)\rho^{AB}(\mathrm{id}_A \otimes |j\rangle).$$

Matricialmente, interpretando $Z_{AB} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ como a matriz em blocos $Z_{AB} = (Z_{ij}) \in M_{d_A}(M_{d_B}(\mathbb{C}))$:

$$Z_{AB} = \begin{bmatrix} Z_{11} & Z_{12} & \dots & Z_{1d_A} \\ Z_{21} & Z_{22} & \dots & Z_{2d_A} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{d_A1} & Z_{d_A2} & \dots & Z_{d_Ad_A} \end{bmatrix} = \sum_{i,j}^{d_A} |i\rangle\langle j| \otimes Z_{ij},$$

então temos:

$$\mathrm{Tr}_A \{Z_{AB}\} = Z_{11} + Z_{22} + \cdots + Z_{d_Ad_A} \in \mathcal{B}(\mathcal{H}_B) \tag{2.9}$$

e

$$\mathrm{Tr}_B \{Z_{AB}\} = \begin{bmatrix} \mathrm{Tr}\{Z_{11}\} & \mathrm{Tr}\{Z_{12}\} & \dots & \mathrm{Tr}\{Z_{1d_A}\} \\ \mathrm{Tr}\{Z_{21}\} & \mathrm{Tr}\{Z_{22}\} & \dots & \mathrm{Tr}\{Z_{2d_A}\} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Tr}\{Z_{d_A1}\} & \mathrm{Tr}\{Z_{d_A2}\} & \dots & \mathrm{Tr}\{Z_{d_Ad_A}\} \end{bmatrix} \in \mathcal{B}(\mathcal{H}_A).$$

Definição 2.2.16 (Estado Maximamente Emaranhado)**.** Se $d_A = d_B$, um estado $\rho^{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ é dito **maximamente emaranhado** caso $\mathrm{Tr}_A\{\rho^{AB}\} = \mathrm{Tr}_B\{\rho^{AB}\} = \mathbb{1}_A/d_A$, ou seja, caso seus estados locais são o estado maximamente misto (Exemplo 2.1.5).

O estado ω_{d_A} do Exemplo 2.2.11 é maximamente emaranhado, e qualquer outro estado maximamente emaranhado em $\mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ pode ser obtido como $(U \otimes V)\omega_{d_A}$, para $U \in \mathcal{B}(\mathcal{H}_A)$ e $V \in \mathcal{B}(\mathcal{H}_B)$ unitários [41] .

Uma propriedade útil do traço parcial é a seguinte:

Proposição 2.2.17 ([42])**.** *Para todo $\rho^{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$:*

$$\mathrm{Tr}\{\rho^{AB}\} = \mathrm{Tr}_B\{\mathrm{Tr}_A\{\rho^{AB}\}\} = \mathrm{Tr}_A\{\mathrm{Tr}_B\{\rho^{AB}\}\}. \tag{2.10}$$

Além disso, $\mathrm{Tr}_A\{\cdot\}$ *e* $\mathrm{Tr}_B\{\cdot\}$ *são as única funções que satisfazem* (2.10)*.*

2.2.5 Medições Locais

O Exemplo 2.2.11 mostra que não é sempre possível atribuir um estado puro a um subsistema de um sistema bipartite. Vejamos agora que dado $\rho^{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$, ρ^A e ρ^B são os únicos estados dos subsistemas A e B que são consistentes com o estado do sistema composto.

Chamamos de **medição local** em A uma PVM $\{P_x\}_{x\in\mathcal{X}}$ sobre $\mathcal{H}_A \otimes \mathcal{H}_B$ tal que, para todo $x \in \mathcal{X}$:

$$P_x = \Pi_x \otimes \mathbb{1}_B, \tag{2.11}$$

onde $\{\Pi_x\}_{x\in\mathcal{X}}$ é um PVM sobre A. Vejamos que estes operadores de fato se restringem a medições no subsistema.

Suponha que o sistema composto AB se encontra no estado puro $\rho^{AB} = |\psi\rangle\langle\psi| \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$. Ao realizar a medição local $\{P_x\}_{x\in\mathcal{X}}$, a probabilidade de obter x é dada por:

$$\begin{aligned} p_x &= \mathrm{Tr}\left\{|\psi\rangle\langle\psi|\, P_x\right\} \\ &= \mathrm{Tr}\left\{|\psi\rangle\langle\psi|\, (\Pi_x \otimes \mathbb{1}_B)\right\} \\ &= \sum_{i,j} (\langle i|_A \otimes \langle j|_B)(|\psi\rangle\langle\psi|\, (\Pi_x \otimes \mathbb{1}_B)(|i\rangle_A \otimes |j\rangle_B) \\ &= \sum_i (\langle i|_A \otimes \mathbb{1}_B)\left(\sum_j (\mathbb{1}_A \otimes \langle j|_B)\rho^{AB}(\mathbb{1}_A \otimes |j\rangle_B)\right)(\Pi_x\, |i\rangle_A \otimes \mathbb{1}_B) \\ &= \mathrm{Tr}\left\{\rho^A \Pi_x\right\}. \end{aligned}$$

Isso mostra que se o sistema total se encontra no estado ρ^{AB}, então o subsistema A é corretamente descrito pelos operador local ρ^A. Usando medições locais em B, o mesmo argumento mostra o resultado completamente análogo, que o estado do subsistema B é dado por ρ^B.

Isso porém, não significa que $\rho^{AB} = \rho^A \otimes \rho^B$ (já que nem todo estado é separável), apesar do fato de que ρ^{AB} e $\rho^A \otimes \rho^B$ têm as mesmas distribuições de resultados para medições locais.

Note que uma operação local em A (evolução unitária ou medição) não deve afetar o estado do subsistema B. Caso contrário, operações locais poderiam ser usadas para transmitir informação instantaneamente, contrariando o princípio da relatividade especial. De fato, como operações locais em A comutam com operações locais em B, a ordem em que essas operações são realizadas não importa. Na verdade, assumindo que os sistemas A e B possam estar separados na espaço por qualquer distância, a noção de simultaneidade não está bem definida!

Teorema 2.2.18 (Teorema da Não-Comunicação)**.** *Seja AB um sistema quântico bipartite e $\rho \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$. O estado local ρ^B do subsistema B é invariante a evoluções unitárias locais e medições locais em A.*

2.2.6 Purificação

Todo estado não-puro pode ser considerado um estado local de um estado puro de um sistema composto maior.

Definição 2.2.19 (Purificação)**.** Uma purificação de um estado $\rho^A \in \mathcal{D}(\mathcal{H}_A)$ é um estado puro dado por $|\Psi\rangle_{AR} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_R)$ para algum sistema quântico de referência R, de forma que:

$$\rho^A = \mathrm{Tr}_R\left\{|\Psi\rangle\langle\Psi|_{AR}\right\}, \tag{2.12}$$

Se a decomposição espectral de ρ^A é dada por:

$$\rho^A = \sum_{k=1}^{d_A} \lambda_k\, |a_k\rangle\langle a_k|\,, \tag{2.13}$$

então, tomando $\mathcal{H}_R \simeq \mathcal{H}_A$, chamamos de **purificação canônica** de ρ^A o estado $|\Psi\rangle\langle\Psi|_{AR} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_R)$ onde:

$$|\Psi\rangle_{AR} = \sum_{k=1}^{d_A} \sqrt{\lambda_k}\, |a_k\rangle_A \otimes |a_k\rangle_R \,. \tag{2.14}$$

Proposição 2.2.20. *Dadas $|\Psi_1\rangle_{AR_1}$ e $|\Psi_2\rangle_{AR_2}$ duas purificações de ρ^A, com $d_{R_1} \leq d_{R_2}$, existe uma isometria $U : R_1 \to R_2$ tal que:*

$$|\Psi_2\rangle_{AR_2} = (\mathbb{1}_A \otimes U\, |\Psi_1\rangle_{AR_1}). \tag{2.15}$$

Esssa proposição é consequência de que, dada a purificação $|\Psi\rangle_{AR}$, (2.14) é sua decomposição de Schmidt.

2.3 Canais Quânticos

Nesta seção, vamos introduzir a noção de canal quântico que, além de surgir como um modelo quântico de canal de comunicação ruidoso, formaliza as evoluções estocásticas de sistemas quânticos, que incluem evoluções unitárias e medições como casos particulares.

2.3.1 Definições e Exemplos

Como vimos na Seção 1.2, um canal clássico $(\mathcal{X}, N, \mathcal{Y})$ pode ser interpretado como um mapa $N : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{Y})$. Similarmente vamos definir um canal quântico $(A, \mathcal{N}, B)$ entre sistemas quânticos A e B, com $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ um mapa linear que leva estados de um sistema para estados de outro. Para isso, é necessário que, para qualquer $\rho^A \in \mathcal{D}(\mathcal{H}_\mathcal{A})$, $\mathcal{N}(\rho^A)$ seja positivo e de traço um. Além disso, para que o canal esteja bem definido quando o sistema A é visto como subsistema de um sistema composto, exigimos que $\mathcal{N}$ seja completamente positivo, como definimos a seguir:

Definição 2.3.1 (Mapa Positivo, Mapa Completamente Positivo)**.** Um mapa linear $\Phi : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ é dito **positivo** se para todo $X \in \mathcal{B}(\mathcal{H}_A)$ positivo, $\Phi(X)$ é positivo. Φ é dito **completamente positivo** se, para qualquer espaço de Hilbert $\mathcal{H}_C$, o mapa $\mathrm{id}_C \otimes \Phi$ é positivo.

Para entender melhor esta definição, considere $\mathcal{H}_C = \mathbb{C}^n$. Um operador $X_{CA} \in \mathcal{B}(\mathcal{H}_C \otimes \mathcal{H}_A) \cong M_n(\mathcal{B}(\mathcal{H}_A))$ pode ser escrito como:

$$X_{CA} = \sum_{i,j} |i\rangle\langle j|_C \otimes X_{ij},$$

então:

$$(\mathrm{id}_C \otimes \Phi)(X_{CA}) = \sum_{i,j} |i\rangle\langle j|_C \otimes \Phi(X_{ij}) = \begin{bmatrix} \Phi(X_{11}) & \Phi(X_{12}) & \cdots & \Phi(X_{1n}) \\ \Phi(X_{21}) & \Phi(X_{22}) & \cdots & \Phi(X_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi(X_{n1}) & \Phi(X_{n2}) & \cdots & \Phi(X_{nn}) \end{bmatrix}. \tag{2.16}$$

Definindo $\Phi^{(n)} : M_n(\mathcal{B}(\mathcal{H}_A)) \to M_n(\mathcal{B}(\mathcal{H}_B))$ por $\Phi^{(n)}(X_{CA}) = (\mathrm{id}_C \otimes \Phi)(X_{CA})$, obtemos que Φ é completamente positiva se e somente se cada $\Phi^{(n)}$ é positiva.

Definição 2.3.2 (Canal Quântico)**.** Um **canal quântico** entre sistemas quânticos A e B é um mapa linear $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ completamente positivo que preserva traço, ou seja, tal que para todo $X \in \mathcal{B}(\mathcal{H}_A)$:

$$\mathrm{Tr}\left\{\mathcal{N}(X)\right\} = \mathrm{Tr}\left\{X\right\}. \tag{2.17}$$

Também denotamos o canal por $(A, \mathcal{N}, B)$.

Note que bastava definir $\mathcal{N}$ com domínio $\mathcal{D}(\mathcal{H}_A)$ e contradomínio $\mathcal{D}(\mathcal{H}_B)$ de forma que $\mathcal{N}$ preserve combinações convexas (de forma que $\mathcal{N}$ seja determinada pela imagem dos estados puros de A). Mas como sempre que $\mathcal{H}_A$ tem dimensão finita podemos obter uma base de $\mathcal{B}(\mathcal{H}_A)$ formada por elementos de $\mathcal{D}(\mathcal{H}_A)$ (ver, por exemplo [10]), tal mapa tem uma única extensão linear $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ de forma que vale (2.17).

Exemplo 2.3.3 (Canal Constante)**.** O canal $(A, \mathcal{N}, B)$ é dito um canal constante se existe $\sigma^B \in \mathcal{D}(\mathcal{H}_B)$ tal que para todo $\rho \in \mathcal{D}(\mathcal{H}_A)$:

$$\mathcal{N}(\rho) = \mathrm{Tr}\left\{\rho\right\}\sigma^B.$$

A presença do traço na fórmula é necessária para que o domínio de $\mathcal{N}$ se estenda para $\mathcal{B}(\mathcal{H}_A)$ de forma a satisfazer (2.17).

Exemplo 2.3.4 (Canais Unitários)**.** Dizemos que $(A, \mathcal{N}, A)$ é um **canal unitário** se:

$$\mathcal{N}(\rho) = U\rho U^\dagger,$$

para algum $U \in \mathcal{B}(\mathcal{H}_A)$ unitário. Em particular, para $U = \mathbb{1}_A$ obtemos o **canal ideal** id_A que é a função identidade em $\mathcal{B}(\mathcal{H}_A, \mathcal{H}_A)$.

Exemplo 2.3.5 (Canal Depolarizador)**.** Para $p \in [0, 1]$ chamamos de **canal depolarizador** o canal quântico (A, D_p, A) dado por:

$$D_p(\rho) = (1-p)\frac{\mathbb{1}_A}{d_A}\mathrm{Tr}\left\{\rho\right\} + p\rho.$$

Note que D_1 é o canal ideal e D_0 é o canal constante que retorna o estado maximamente misto.

Uma ferramenta para demonstrar que os exemplos anteriores realmente são canais quânticos é o Teorema de Choi (2.3.9) que veremos adiante.

2.3.2 Isomorfismo de Choi-Jamiołkowski

Dados espaços de Hilbert $\mathcal{H}_A$ e $\mathcal{H}_B$, vimos que a vetorização (Definição 2.2.4) é um exemplo concreto do isomorfismo:

$$\mathcal{B}(\mathcal{H}_A, \mathcal{H}_\mathcal{B}) \simeq \mathcal{H}_A \otimes \mathcal{H}_B,$$

que nos permite tratar operadores em $\mathcal{B}(\mathcal{H}_A, \mathcal{H}_\mathcal{B})$ como vetores em $\mathcal{H}_B^{(d_A)}$. Vejamos agora um isomorfismo:

$$\mathcal{B}(\mathcal{B}(\mathcal{H}_A), \mathcal{B}(\mathcal{H}_\mathcal{B})) \simeq \mathcal{B}(\mathcal{H}_A) \otimes \mathcal{B}(\mathcal{H}_\mathcal{B}) \simeq \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B),$$

que nos permite tratar mapas lineares em $\Phi \in \mathcal{B}(\mathcal{B}(\mathcal{H}_A), \mathcal{B}(\mathcal{H}_\mathcal{B}))$ como operadores em $C_\Phi \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_A)$ ou, em termos de bases, como matrizes em $M_{d_A}(\mathcal{B}(\mathcal{H}_B))$.

Definição 2.3.6. Para $\mathcal{H}_A \simeq \mathbb{C}^{d_A}$ e $\mathcal{H}_B$ espaços de Hilbert, o **Isomorfismo de Choi-Jamiołkowski** é o mapa $C : \mathcal{B}(\mathcal{B}(\mathcal{H}_A), \mathcal{B}(\mathcal{H}_\mathcal{B})) \to \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ dado por

$$\Phi \mapsto C_\Phi := \sum_{i,j=1}^{d_A} |i\rangle\langle j| \otimes \Phi(|i\rangle\langle j|).$$

Identificando $\mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \simeq \mathcal{B}(\mathcal{H}_A) \otimes \mathcal{B}(\mathcal{H}_\mathcal{B}) \simeq M_{d_A} \otimes \mathcal{B}(\mathcal{H}_\mathcal{B}) \simeq M_{d_A}(\mathcal{B}(\mathcal{H}_\mathcal{B}))$, temos que C_Φ é dada por uma matriz com entradas em $\mathcal{B}(\mathcal{H}_\mathcal{B})$, dita a **matriz de Choi** de Φ:

$$C_\Phi = \begin{bmatrix} \Phi(|1\rangle\langle 1|) & \Phi(|1\rangle\langle 2|) & \cdots & \Phi(|1\rangle\langle d_A|) \\ \Phi(|2\rangle\langle 1|) & \Phi(|2\rangle\langle 2|) & \cdots & \Phi(|2\rangle\langle d_A|) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi(|d_A\rangle\langle 1|) & \Phi(|d_A\rangle\langle 2|) & \cdots & \Phi(|d_A\rangle\langle d_A|) \end{bmatrix}.$$

O posto de C_Φ é chamado de **rank de Choi** de Φ.

Proposição 2.3.7 ([41, 43, 44])**.** *O Isomorfismo de Choi-Jamiołkowski $C : \mathcal{B}(\mathcal{B}(\mathcal{H}_A), \mathcal{B}(\mathcal{H}_\mathcal{B})) \to \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ é um isomorfismo linear. Sua inversa é dada por:*

$$C^{-1}(\rho)(X) := \mathrm{Tr}_A \left\{ \rho(X^T \otimes \mathbb{1}_B) \right\},$$

onde $\rho \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ e $X \in \mathcal{B}(\mathcal{H}_A)$.

Para uma demonstração, ver [45]. Existe uma outra forma de expressar a matriz de Choi que será conveniente:

Proposição 2.3.8 ([34])**.** *Para um mapa linear $\Phi : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ e $X \in \mathcal{B}(\mathcal{H}_A, \mathcal{H}_B)$:*

$$\mathrm{Vec}(X) = (\mathbb{1}_{\mathcal{H}_A} \otimes X) |\Omega_{d_A}\rangle,$$

$$C_\Phi = d_A(\mathrm{id}_A \otimes \Phi)(\omega_{d_A}),$$

onde $|\Omega_{d_A}\rangle$ e ω_{d_A} são dados no Exemplo 2.2.11.

2.3.3 Representação de Choi-Kraus de mapas completamente positivos

Para qualquer $K \in \mathcal{B}(\mathcal{H}_A, \mathcal{H}_B)$, considere o mapa de $\mathrm{Ad}_K : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ dado por $\mathrm{Ad}_K(X) = KXK^\dagger$. $\mathrm{Ad}_K(X)$ é positivo, já que para todo $Y \in \mathcal{B}(\mathcal{H}_A, \mathcal{H}_B)$:

$$\mathrm{Ad}_K(Y^\dagger Y) = KY^\dagger YK^\dagger = (YK^\dagger)^\dagger (YK^\dagger) \geq 0.$$

Além disso, Ad_K é completamente positivo, pois, $\mathrm{id}_C \otimes \mathrm{Ad}_K = \mathrm{Ad}_{\mathbb{1}_C \otimes K}$ para qualquer espaço de Hilbert $\mathcal{H}_C$. Assim, a soma de mapas desta forma também é completamente positivo. Vejamos que de fato todo mapa completamente positivo é desta forma.

Teorema 2.3.9 (Teorema de Choi [46]). *Para $\Phi : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ um mapa linear, as seguintes afirmações são equivalentes:*

1. *Φ é completamente positivo.*
2. *C_Φ é positiva semidefinida.*
3. *Existem operadores $\{K_i\}_{i=1}^R \subset \mathcal{B}(\mathcal{H}_A, \mathcal{H}_B)$ com $R = \mathrm{rk}(C_\Phi)$ o rank de Choi de Φ, tais que:*

$$\Phi = \sum_{i=1}^{R} \mathrm{Ad}_{K_i}. \tag{2.18}$$

Demonstração. Já vimos acima que *3* $\to$ *1*. *1* $\to$ *2* segue da Proposição 2.3.8 e de que $\omega_{d_A} \in \mathcal{B}(\mathbb{C}^{d_A})^+$, já que então $(\mathrm{id}_A \otimes \Phi)(\omega_{d_A})$ é positivo para Φ completamente positivo. Para ver *2* $\to$ *3*, considere a decomposição espectral de C_Φ:

$$C_\Phi = \sum_{i=1}^{R} |\psi_i\rangle\langle\psi_i| ,$$

onde $R = \mathrm{rk}(C_\Phi)$ e $|\psi_i\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ não-normalizados. Pela Proposição 2.2.5, a vetorização é inversível, logo existem $K_i \in \mathcal{B}(\mathcal{H}_A, \mathcal{H}_B)$ tais que:

$$|\psi_i\rangle = \mathrm{Vec}(K_i) = \sqrt{d_A}(\mathbb{1}_{\mathcal{H}_A} \otimes K_i) |\Omega_{d_A}\rangle .$$

Assim obtemos:

$$\begin{aligned} C_\Phi &= \sum_{i=1}^{R} d_A(\mathbb{1}_{\mathcal{H}_A} \otimes K_i) |\Omega_{d_A}\rangle \langle\Omega_{d_A}| (\mathbb{1}_{\mathcal{H}_A} \otimes K_i)^\dagger \\ &= \sum_{i=1}^{R} d_A \mathrm{Ad}_{\mathbb{1}_A \otimes K_i}(\omega_{d_A}) \\ &= \sum_{i=1}^{R} d_A(\mathrm{id}_A \otimes \mathrm{Ad}_{K_i})(\omega_{d_A}) \\ &= \sum_{i=1}^{R} C_{\mathrm{Ad}_{K_i}} \\ &= C_{\sum_{i=1}^{R} \mathrm{Ad}_{K_i}}. \end{aligned}$$

Pelo Isomorfismo de Choi-Jamiołkowski, obtemos (2.18). □

Uma decomposição da forma $\Phi = \sum_{i=1}^{R} \mathrm{Ad}_{K_i}$ é chamada de **representação de Choi-Kraus** de Φ. Apesar de não ser única, distintas representações são relacionadas por transformações unitárias:

Proposição 2.3.10. *As famílias de operadores $\{K_i\}_{i=1}^N, \{V_i\}_{i=1}^M \subset \mathcal{B}(\mathcal{H}_A, \mathcal{H}_B)$ formam a representação de Choi-Kraus de um mesmo mapa linear Φ se, e somente se, existe matriz unitária $U \in M_n(\mathbb{C})$, com $n = \max\{N, M\} \leq d_A d_B$, tal que $K_i = \sum_{j=1}^{n} U_{ij} V_j$, onde a família com menos operadores é estendida pelo operador nulo de forma que ambas famílias tenham o mesmo tamanho.*

Para uma prova, ver [11]. Podemos agora obter uma caracterização dos canais quânticos em termos de sua representação de Choi-Kraus:

Teorema 2.3.11. *Um mapa linear $\Phi : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ é um canal quântico se e somente se existem $\{K_i\}_{i=1}^N \subset \mathcal{B}(\mathcal{H}_A, \mathcal{H}_B)$ tal que:*

$$\Phi = \sum_{i=1}^{N} \mathrm{Ad}_{K_i},$$

garantindo que Φ é completamente positiva, e ainda com:

$$\sum_{i=1}^{N} K_i^\dagger K_i = \mathbb{1}_{\mathcal{H}_\mathcal{A}}, \tag{2.19}$$

que garante a preservação do traço.

Demonstração. Já sabemos que Φ é completamente positiva se e somente se vale a primeira equação. Agora, para todo $X \in \mathcal{B}(\mathcal{H}_A)$:

$$\begin{aligned}
\mathrm{Tr}\left\{\Phi(X)\right\} &= \mathrm{Tr}\left\{\sum_{i=1}^{N} K_i X K_i^\dagger\right\} \\
&= \sum_{i=1}^{N} \mathrm{Tr}\left\{K_i X K_i^\dagger\right\} \\
&= \sum_{i=1}^{N} \mathrm{Tr}\left\{K_i^\dagger K_i X\right\} \\
&= \mathrm{Tr}\left\{\left(\sum_{i=1}^{N} K_i^\dagger K_i\right) X\right\}.
\end{aligned}$$

Portanto $\mathrm{Tr}\left\{\Phi(X)\right\} = \mathrm{Tr}\left\{X\right\}$ se e somente se vale (2.19). □

Com o Teorema de Choi em mãos, vejamos alguns exemplos importantes de canais quânticos.

Exemplo 2.3.12 (Canal Produto)**.** Dados canais quânticos $(A_1, \mathcal{N}_1, B_1)$ e $(A_2, \mathcal{N}_2, B_2)$, com representações de Choi-Kraus $\{K_i\}$ e $\{L_j\}$ respectivamente, obtemos o **canal produto** $(A_1A_2, \mathcal{N}_1 \otimes \mathcal{N}_2, B_1B_2)$ dado por:

$$\mathcal{N}_1 \otimes \mathcal{N}_2 = \sum_{i,j} \mathrm{Ad}_{K_i \otimes L_j}.$$

Em particular, para qualquer canal quântico $\mathcal{N}$, temos que $\mathcal{N}^{\otimes n} = \underbrace{\mathcal{N} \otimes \mathcal{N} \cdots \otimes \mathcal{N}}_{n \text{ vezes}}$ é o canal quântico que resulta de n usos independentes de $\mathcal{N}$.

Exemplo 2.3.13 (Composição de Canais)**.** Dados canais quânticos $(A, \mathcal{N}_1, B)$ e $(B, \mathcal{N}_2, C)$, com representações de Choi-Kraus $\{K_i\}$ e $\{L_j\}$ respectivamente, o canal composição $(A, \mathcal{N}_2 \circ \mathcal{N}_1, C)$ é dado pelos operadores de Choi-Kraus $\{L_jK_i\}$, de fato:

$$(\mathcal{N}_2 \circ \mathcal{N}_1)(\rho^A) = \sum_j L_j \left(\sum_i K_i \rho^A K_i^\dagger\right) L_j^\dagger = \sum_{i,j} (L_jK_i)\rho^A(L_jK_i)^\dagger.$$

Exemplo 2.3.14 (Traço e Traço Parcial como Canais Quânticos)**.** O canal $(A, \mathrm{Tr}_A\{\cdot\}, \mathbb{C})$ com representação de Choi-Kraus $\{K_i\} = \{\langle i|\}$, onde $\{|i\rangle\}$ é a base computacional de $\mathcal{H}_A$, age por:

$$\mathrm{Tr}_A\left\{\rho^A\right\} = \sum_{i=1}^{d_A} \langle i|\,\rho^A(\langle i|)^\dagger = \sum_{i=1}^{d_A} \langle i|\rho^A|i\rangle = \mathrm{Tr}\left\{\rho^A\right\},$$

ou seja, o traço pode ser visto como um canal quântico. Similarmente, o canal produto $\mathrm{Tr}_A\{\cdot\} \otimes \mathrm{id}_B$ tem representação de Choi-Kraus:

$$\mathrm{Tr}_A\{\cdot\} \otimes \mathrm{id}_B(\rho^{AB}) = \sum_{i=1}^{d_A} (\langle i| \otimes \mathbb{1}_B)\rho^{AB}(\langle i| \otimes \mathbb{1}_B)^\dagger = \mathrm{Tr}_A\left\{\rho^{AB}\right\}.$$

Veremos no Capítulo 3 que o processo de preparo de um sistema quântico em um dado estado, assim como a realização de uma medição sobre o sistema, podem ser vistos como canais quânticos.

2.3.4 Representação de Stinespring

A representação de Choi-Kraus é mais frequentemente encontrada na teoria da informação quântica, já que nos permite computar diretamente a ação do canal. Mas sua interpretação física não é imediata. Vejamos agora uma representação de canais quânticos que é conceitualmente mais próxima da noção de canal clássico, ou seja, de um processo com ruído, com perda de informação para um ambiente maior.

Teorema 2.3.15 (Dilatação de Stinespring)**.** *Dado um canal quântico* $\Phi : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$*, existe um espaço de Hilbert* $\mathcal{H}_E$ *e uma isometria* $V : \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_E$*, dita uma* ***isometria de Stinespring****, tal que*

$$\Phi(X) = \mathrm{Tr}_E\left\{VXV^\dagger\right\}, \tag{2.20}$$

para todo $X \in \mathcal{B}(\mathcal{H}_A)$*.*

Demonstração. Seja $\{K_i\}_{i=1}^R \subset \mathcal{B}(\mathcal{H}_A, \mathcal{H}_B)$ uma representação de Choi-Kraus de Φ. Tomando $\mathcal{H}_E = \mathbb{C}^R$ e definindo V por:

$$V = \sum_{i=1}^{R} K_i \otimes |i\rangle$$

com $\{|i\rangle\}_{i=1}^R$ a base computacional de $\mathbb{C}^R$. V é uma isometria pois,

$$V^\dagger V = \left(\sum_{j=1}^{R} K_j^\dagger \otimes \langle j|\right)\left(\sum_{i=1}^{R} K_i \otimes |i\rangle\right) = \sum_{i=1}^{R} K_i^\dagger K_i = \mathbb{1}_{\mathcal{H}_A},$$

e temos que, para qualquer $X \in \mathcal{B}(\mathcal{H}_A)$,

$$\mathrm{Tr}_E\left\{VXV^\dagger\right\} = \sum_{i,j=1}^{R} \mathrm{Tr}_E\left\{K_i X K_j^\dagger \otimes |i\rangle\langle j|\right\} = \sum_{i=1}^{R} K_i X K_i^\dagger = \Phi(X).$$

□

No contexto de C^*-álgebras, a dilatação de Stinespring aparece em outra forma [39][26], e considera mapas completamente positivos unitais. Podemos recuperar esta versão considerando o mapa dual Φ^* de um canal quântico.

Definição 2.3.16. Dado um mapa linear $\Phi : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$, chamamos de **mapa dual** o único mapa linear $\Phi^* : \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A)$ tal que:

$$\langle \Phi(X), Y\rangle_{HS} = \mathrm{Tr}\left\{\Phi(X)^\dagger Y\right\} = \mathrm{Tr}\left\{X^\dagger \Phi^*(Y)\right\} = \langle X, \Phi^*(Y)\rangle_{HS},$$

para todo $X \in \mathcal{B}(\mathcal{H}_A)$,$Y \in \mathcal{B}(\mathcal{H}_B)$.

É fácil ver que para qualquer mapa linear $(\Phi^*)^* = \Phi$, e que Φ é completamente positiva se, e somente se, Φ^* também for. Chamamos um mapa linear de **unital** se preserva os operadores identidade.

Lema 2.3.17. *Um mapa linear $\Phi : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ preserva traço se, e somente se, Φ^* é unital, ou seja, se $\Phi^*(\mathbb{1}_{\mathcal{H}_B}) = \mathbb{1}_{\mathcal{H}_A}$.*

Para uma prova, ver [12].

Lema 2.3.18. *Dado $K \in \mathcal{B}(\mathcal{H}_A, \mathcal{H}_B)$, $\mathrm{Ad}^*_K(Y) = \mathrm{Ad}_{K^\dagger}(Y)$, para todo $Y \in \mathcal{B}(\mathcal{H}_B)$.*

Demonstração.

$$\begin{aligned}\langle \mathrm{Ad}_K(X), Y\rangle_{HS} &= \langle KXK^\dagger, Y\rangle_{HS}\\ &= \mathrm{Tr}\left\{(KXK^\dagger)^\dagger Y\right\}\\ &= \mathrm{Tr}\left\{KX^\dagger K^\dagger Y\right\}\\ &= \mathrm{Tr}\left\{X^\dagger K^\dagger Y K\right\}\\ &= \langle X, \mathrm{Ad}_{K^\dagger}(Y)\rangle_{HS}.\end{aligned}$$

□

Lema 2.3.19. *Sejam $\mathcal{H}_A, \mathcal{H}_E$ espaços de Hilbert de dimensão finita, $\mathrm{Tr}_E : \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_E) \to \mathcal{B}(\mathcal{H}_A)$ o traço parcial. Então para todo $Y \in \mathcal{B}(\mathcal{H}_A)$, $\mathrm{Tr}^*_E(Y) = (Y \otimes \mathbb{1}_{\mathcal{H}_E})$.*

Corolário 2.3.20 (Versão dual da dilatação de Stinespring)**.** *Para qualquer mapa linear $\Phi : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ completamente positivo e unital, existe um espaço de Hilbert $\mathcal{H}_E$ e uma isometria $V : \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_E$ tal que:*

$$\Phi(X) = V^\dagger (X \otimes \mathbb{1}_{\mathcal{H}_E}) V,$$

para todo $X \in \mathcal{B}(\mathcal{H}_A)$.

Demonstração. Como Φ é unital, o Lema 2.3.17 implica que Φ^* é um canal quântico. Pelo teorema 2.3.15, existe $\mathcal{H}_E$ e uma isometria de Stinespring V tal que para todo $Y \in \mathcal{B}(\mathcal{H}_B)$:

$$\Phi^*(Y) = \mathrm{Tr}_E\left\{VYV^\dagger\right\} = \mathrm{Tr}_E\left\{\mathrm{Ad}_V(Y)\right\}.$$

Temos que, para todo $X \in \mathcal{B}(\mathcal{H}_A)$:

$$\begin{aligned}\langle X, \mathrm{Tr}_E\left\{\mathrm{Ad}_V(Y)\right\}\rangle_{HS} &= \langle X \otimes \mathbb{1}_{\mathcal{H}_E}, \mathrm{Ad}_V(Y)\rangle_{HS}\\ &= \langle \mathrm{Ad}_{V^\dagger}(X \otimes \mathbb{1}_{\mathcal{H}_E}), Y\rangle_{HS}\\ &= \langle V^\dagger(X \otimes \mathbb{1}_{\mathcal{H}_E})V, Y\rangle_{HS},\end{aligned}$$

que implica em $\Phi(X) = (\Phi^*)^*(X) = V^\dagger(X \otimes \mathbb{1}_{\mathcal{H}_E})V$. □

Teorema 2.3.21. *Seja $(A, \mathcal{N}, B)$ um canal quântico e $\mathcal{H}_E$ o espaço de Hilbert que surge de uma dilatação de Stinespring. Para qualquer estado puro $|\psi\rangle\langle\psi|_{BE} \in \mathcal{B}(\mathcal{H}_B \otimes \mathcal{H}_E)$, existe um operador unitário $U \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E)$ tal que*

$$\mathcal{N}(X) = \mathrm{Tr}_{AE}\left\{U(X \otimes |\psi\rangle\langle\psi|_{BE})U^\dagger\right\}, \tag{2.21}$$

para todo $X \in \mathcal{B}(\mathcal{H}_A)$.

Demonstração. Seja $V : \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_E$ uma isometria de Stinespring, $|\psi\rangle\langle\psi|_{BE} \in \mathcal{B}(\mathcal{H}_B \otimes \mathcal{H}_E)$ um estado puro e $|\phi\rangle\langle\phi|_A \in \mathcal{B}(\mathcal{H}_A)$ um estado puro. Se $\{|i\rangle_A\}_{i=1}^{d_A}$ é uma base ortonormal de $\mathcal{H}_A$, então temos os conjuntos $\{|x_i\rangle\}_{i=1}^{d_A}, \{|y_i\rangle\}_{i=1}^{d_A} \subset \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E$, ambos ortogonais, definidos por:

$$|x_i\rangle = |\phi\rangle_A \otimes V|i\rangle_A \text{ e } |y_i\rangle = |i\rangle_A \otimes |\psi\rangle_{BE}.$$

Como ambos tem o mesmo tamanho e são ortogonais, então existe $U \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E)$ tal que

$$U|y_i\rangle = |x_i\rangle, \quad i \in \{1, 2, \ldots, d_A\}.$$

Para $|z\rangle \in \mathcal{H}_A$ qualquer, temos que:

$$\begin{aligned} U(|z\rangle \otimes |\psi\rangle_{BE}) &= \sum_{i=1}^{d_A} \langle i|z\rangle\, U(|i\rangle_A \otimes |\psi\rangle_{BE}) \\ &= \sum_{i=1}^{d_A} \langle i|z\rangle\, (|\phi\rangle_A \otimes V|i\rangle_A) \\ &= |\phi\rangle_A \otimes V|z\rangle_A. \end{aligned}$$

Com isso em mão, vejamos que 2.21 vale para $X = |i\rangle\langle j|_A$, para quaisquer $i, j \in \{1, 2, \ldots, d_A\}$, e o teorema segue por linearidade.

$$\begin{aligned} \mathrm{Tr}_{AE}\left\{U(|i\rangle\langle j|_A \otimes |\psi\rangle\langle\psi|_{BE})U^\dagger\right\} &= \mathrm{Tr}_{AE}\left\{U(|i\rangle_A \otimes |\psi\rangle_{BE})(\langle j|_A \otimes \langle\psi|_{BE})U^\dagger\right\} \\ &= \mathrm{Tr}_{AE}\left\{(|\phi\rangle_A \otimes V|i\rangle_A)(\langle\phi|_A \otimes V\langle j|_A)\right\} \\ &= \mathrm{Tr}_{AE}\left\{|\phi\rangle\langle\phi|_A \otimes V|i\rangle\langle j|_A V^\dagger\right\} \\ &= \mathrm{Tr}_E\left\{V|i\rangle\langle j|_A V^\dagger\right\} \\ &= \mathcal{N}(|i\rangle\langle j|_A), \end{aligned}$$

onde usamos que $\mathrm{Tr}_A\{|\phi\rangle\langle\phi|_A\} = 1$, sendo um estado puro. □

2.3.5 Relação entre as representações

Cada forma de representar de canais quânticos que vimos nesse capítulo está relacionada, e podemos obter cada uma dada outra:

Proposição 2.3.22. *Seja $\Phi : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ um canal quântico e $\{K_i\}_{i=1}^R \subset \mathcal{B}(\mathcal{H}_A, \mathcal{H}_B)$ uma família de operadores. As seguintes afirmações são equivalentes:*

1. *(Choi-Kraus) Temos que*

$$\Phi = \sum_{i=1}^R \mathrm{Ad}_{K_i}, \quad \text{e} \quad \sum_{i=1}^R K_i^\dagger K_i = \mathbb{1}_A.$$

2. *(Choi-Jamiołkowski)*

$$C_\Phi = \sum_{i=1}^{R} \mathrm{Vec}(K_i)\mathrm{Vec}^\dagger(K_i), \qquad e \qquad R = \mathrm{rk}(C_\Phi).$$

3. *(Stinespring) Para* $\mathcal{H}_E = \mathbb{C}^R$ *e* $V : \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_E$ *dado por*

$$V = \sum_{i=1}^{R} K_i \otimes |i\rangle\,,$$

temos que, para todo $X \in \mathcal{B}(\mathcal{H}_A)$*:*

$$\Phi(X) = \mathrm{Tr}_E\left\{VXV^\dagger\right\}.$$

Para uma demonstração, ver [12].

2.3.6 Canais *Entanglement-Breaking*

Vamos ver agora uma classe de canais quânticos, introduzidos em [47], de grande importância na teoria da informação quântica, como veremos no próximo capítulo.

Definição 2.3.23. Um canal quântico $(A, \mathcal{N}^{EB}, B)$ é dito um canal ***entanglement-breaking*** (ou canal EB) se, para qualquer sistema quântico C e estado $\rho^{CA} \in \mathcal{D}(\mathcal{H}_C \otimes \mathcal{H}_A)$, o estado $(\mathrm{id}_C \otimes \mathcal{N}^{EB})(\rho^{CA})$ é separável.

Teorema 2.3.24 ([10, 47])**.** *Para um canal quântico* $(A, \mathcal{N}^{EB}, B)$*, as seguintes afirmações são equivalentes:*

1. $\mathcal{N}^{EB}$ *é entanglement-breaking.*
2. *A matriz de Choi* $C_{\mathcal{N}^{EB}}$ *é separável.*
3. $\mathcal{N}^{EB}$ *possui representação de Choi-Kraus com operadores de posto um.*

Demonstração. *1* $\to$ *2* segue da Proposição 2.3.8, que nos dá, para $\mathcal{H}_C \cong \mathcal{H}_A$, que $C_{\mathcal{N}^{EB}} = d_A(\mathrm{id}_C \otimes \mathcal{N}^{EB})(\omega_{d_A})$, que é separável caso vale *1*.

Para ver que *2* $\to$ *3*, suponha que $C_{\mathcal{N}^{EB}}$ é separável. Então pela Proposição 2.2.9, para algum conjunto de índices $\mathcal{X}$ e $p \in \mathcal{P}(\mathcal{X})$, existem vetores normalizados $|\psi_x\rangle \in \mathcal{H}_A, |\varphi_x\rangle \in \mathcal{H}_B$ tais que:

$$(\mathrm{id}_A \otimes \mathcal{N}^{EB})(\omega_{d_A}) = \frac{1}{d_A}\sum_{ij} |i\rangle\langle j|_A \otimes \mathcal{N}^{EB}(|i\rangle\langle j|_A) = \sum_{x\in\mathcal{X}} p(x)\,|\psi_x\rangle\langle\psi_x|_A \otimes |\varphi_x\rangle\langle\varphi_x|_B\,. \tag{2.22}$$

Considere o mapa $\Phi : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ dado por:

$$\Phi(\rho) = d_A \sum_{x\in\mathcal{X}} \mathrm{Tr}\left\{p(x)\,|\psi_x\rangle\langle\psi_x|_A\,\rho\right\}|\varphi_x\rangle\langle\varphi_x|_B\,.$$

Vejamos que Φ é um canal quântico com representação de Kraus dada por:

$$\left\{K_x \equiv \sqrt{d_A p(x)}\,|\varphi_x\rangle_B\,\langle\psi_x|_A\right\}_{x\in\mathcal{X}}.$$

Note que cada K_x tem posto um. É imediato que $\Phi = \sum_x \mathrm{Ad}_{K_x}$, basta então verificar que $\sum_x K_x^\dagger K_x = \mathbb{1}_A$.

$$\begin{aligned}\sum_{x\in\mathcal{X}} K_x^\dagger K_x &= \sum_{x\in\mathcal{X}} d_A p(x)\,\langle\varphi_x|\varphi_x\rangle\,|\psi_x\rangle\langle\psi_x|_A \\ &= d_A \sum_{x\in\mathcal{X}} p(x)\,|\psi_x\rangle\langle\psi_x|_A\,,\end{aligned}$$

note porém que:

$$\begin{aligned}\sum_{x\in\mathcal{X}} p(x)\,|\psi_x\rangle\langle\psi_x|_A &= \mathrm{Tr}_B\left\{\sum_{x\in\mathcal{X}} p(x)\,|\psi_x\rangle\langle\psi_x|_A \otimes |\varphi_x\rangle\langle\varphi_x|_B\right\} \\ &= \frac{1}{d_A}\mathrm{Tr}_B\left\{\sum_{ij}|i\rangle\langle j|_A \otimes \mathcal{N}^{EB}(|i\rangle\langle j|_A)\right\} \qquad \text{(por (2.22))} \\ &= \frac{1}{d_A}\sum_{ij}|i\rangle\langle j|_A \otimes \mathrm{Tr}\left\{\mathcal{N}^{EB}(|i\rangle\langle j|_A)\right\} \\ &= \frac{1}{d_A}\sum_{ij}|i\rangle\langle j|_A \otimes \mathrm{Tr}\left\{|i\rangle\langle j|_A\right\} \\ &= \frac{1}{d_A}\sum_{i}|i\rangle\langle i|_A \\ &= \frac{\mathbb{1}_A}{d_A},\end{aligned}$$

que nos dá $\sum_x K_x^\dagger K_x = \mathbb{1}_A$. Por fim vejamos que $\Phi = \mathcal{N}^{EB}$. Para isso, note que:

$$\begin{aligned}(\mathrm{id}_A \otimes \Phi)(\omega_{d_A}) &= \frac{1}{d_A}\sum_{ij}|i\rangle\langle j|_A \otimes \left(d_A \sum_{x\in\mathcal{X}} \mathrm{Tr}\left\{p(x)\,|\psi_x\rangle\langle\psi_x|_A\,|i\rangle\langle j|\right\}|\varphi_x\rangle\langle\varphi_x|_B\right) \\ &= \sum_{x\in\mathcal{X}} p(x)\sum_{ij}\langle\psi_x|i\rangle\,|i\rangle\langle j|_A\,\langle j|\psi_x\rangle \otimes |\varphi_x\rangle\langle\varphi_x|_B \\ &= \sum_{x\in\mathcal{X}} p(x)\,|\psi_x\rangle\langle\psi_x|_A \otimes |\varphi_x\rangle\langle\varphi_x|_B \\ &= (\mathrm{id}_A \otimes \mathcal{N}^{EB})(\omega_{d_A}),\end{aligned}$$

logo $C_{\mathcal{N}^{EB}} = C_\Phi$, e pelo isomorfismo de Choi-Jamiołkowski temos $\mathcal{N}^{EB} = \Phi$.

Vejamos agora que *3* $\to$ *1*. Suponha que $\mathcal{N}^{EB}$ tem representação de Choi-Kraus $\{K_x \equiv |\varphi_x\rangle_B\,\langle\psi_x|_A\}_{x=1}^R$, onde podemos assumir que $|\varphi_x\rangle_B$ é normalizado. Então:

$$\sum_{x=1}^R K_x^\dagger K_x = \sum_{x=1}^R |\psi_x\rangle\langle\psi_x|_A = \mathbb{1}_A. \tag{2.23}$$

Seja $\mathcal{H}_C \cong \mathcal{H}_A$ e $\rho^{CA} \in \mathcal{D}(\mathcal{H}_C \otimes \mathcal{H}_A)$:

$$(\mathrm{id}_C \otimes \mathcal{N}^{EB})(\rho^{CA}) = \sum_{x=1}^R (\mathbb{1}_C \otimes |\varphi_x\rangle_B\,\langle\psi_x|_A)\rho^{CA}(\mathbb{1}_C \otimes |\psi_x\rangle_A\,\langle\varphi_x|_B)$$

$$= \sum_{x=1}^{R} (\mathbb{1}_C \otimes |\varphi_x\rangle_B)(\mathbb{1}_C \otimes \langle\psi_x|_A)\rho^{CA}(\mathbb{1}_C \otimes |\psi_x\rangle_A)(\mathbb{1}_C \otimes \langle\varphi_x|_B)$$
$$= \sum_{x=1}^{R} \mathrm{Tr}_A \left\{\rho^{CA}(\mathbb{1}_C \otimes |\psi_x\rangle\langle\psi_x|_A)\right\} \otimes |\varphi_x\rangle\langle\varphi_x|_B .$$

Definindo

$$\rho_x^C = \frac{\mathrm{Tr}_A \left\{\rho^{CA}(\mathbb{1}_C \otimes |\psi_x\rangle\langle\psi_x|_A)\right\}}{\mathrm{Tr}\left\{\rho^{CA}(\mathbb{1}_C \otimes |\psi_x\rangle\langle\psi_x|_A)\right\}}, \qquad p(x) = \mathrm{Tr}\left\{\rho^{CA}(\mathbb{1}_C \otimes |\psi_x\rangle\langle\psi_x|_A)\right\},$$

temos que:

$$(\mathrm{id}_C \otimes \mathcal{N}^{EB})(\rho^{CA}) = \sum_{x=1}^{R} p(x)\rho_x^C \otimes |\varphi_x\rangle\langle\varphi_x|_B$$

é separável se p for uma distribuição de probabilidade. Do Lema 3.2.12:

$$p(x) = \mathrm{Tr}\left\{\rho^{CA}(\mathbb{1}_C \otimes |\psi_x\rangle\langle\psi_x|_A)\right\} = \mathrm{Tr}\left\{\mathrm{Tr}_C\left\{\rho^{CA}\right\}|\psi_x\rangle\langle\psi_x|_A\right\},$$

e por (2.23) isso define uma probabilidade (pois as projeções $|\psi_x\rangle\langle\psi_x|_A$ definem um POVM). Como ρ^{CA} é arbitrário, obtemos *1*. □

Capítulo 3

Teoria da Informação Quântica

Na Seção 3.1, que segue principalmente [10], vamos ver como informação clássica pode ser representada por estados quânticos na forma de ensembles de estados distinguíveis, e como informação clássica pode ser armazenada em sistemas quânticos e transmitida por canais quânticos. Para isso, definimos classes de canais quânticos cujas entradas ou saídas são estados distinguíveis, fornecendo uma caracterização dos canais *entanglement-breaking* em termos dessas classes. Na Seção 3.2, seguindo [10, 27], introduzimos a entropia quântica, assim como outras versões quânticas das quantidades entrópicas vistas no Capítulo 1. Veremos também a dita quantidade de Holevo de um ensemble, uma cota superior para a quantidade máxima de informação clássica que pode ser armazenada fielmente no ensemble.

3.1 Informação Clássica em Sistemas Quânticos

Nesta seção, vamos ver que sistemas quânticos podem armazenar informação clássica fielmente caso seja preparado em um estado misto cujas componentes puras são ortogonais entre si.

3.1.1 Estados Distinguíveis

Considere o seguinte cenário: Alice quer armazenar símbolos de $\mathcal{X}$, gerados pela variável aleatória X, em um sistema quântico A. Para isso, ela escolhe um estado $\rho_x \in \mathcal{D}(\mathcal{H}_A)$ para cada $x \in \mathcal{X}$, e prepara o sistema no estado:

$$\rho^A = \sum_{x\in\mathcal{X}} p_X(x)\rho_x, \tag{3.1}$$

que corresponde ao ensemble $\{p_X(x), \rho_x\}_{x\in\mathcal{X}}$. Futuramente, Bob quer acessar os dados armazenados (sabendo que correspondem à variável aleatória X). Para isso, ele deve realizar uma medição sobre A dada por um POVM $\{\Lambda_y\}_{y\in\mathcal{X}}$. Idealmente, Bob gostaria de diferenciar entre os elementos de $\mathcal{X}$, ou equivalentemente, entre as distribuições δ_x, que correspondem aos elementos extremais de $\mathcal{P}(\mathcal{X})$. Na teoria da informação clássica isso é possível, já que nesse caso $\mathrm{Ext}(\mathcal{P}(\mathcal{X}))$ é finito. Bob pode estar incerto de qual símbolo foi armazenado, mas ele sempre sabe diferenciar os possíveis resultados. Mas o espaço de estados puros de um sistema quântico é infinito (não-enumerável), o que impede que medições com finitos resultados possíveis diferenciem perfeitamente entre dois estados quânticos puros.

Definição 3.1.1 (Estados Distinguíveis)**.** Um conjunto de estados $\{\rho_x\}_{x\in\mathcal{X}} \in \mathcal{D}(\mathcal{H}_A)$ é dito **distinguível** se existe uma medição $\{\Lambda_y\}_{y\in\mathcal{X}}$ sobre A tal que:

$$\forall x, y \in \mathcal{X}, \quad \mathrm{Tr}\left\{\rho_x \Lambda_y\right\} = \delta_x(y), \tag{3.2}$$

ou seja, que se o estado de A é ρ_x, então o resultado da medição é x com probabilidade um. Dizemos que os estados ρ_x são **distinguíveis entre si**.

Em contraste com a teoria clássica onde os estados do sistemas são sempre distinguíveis, não é sempre possível distinguir entre diferentes estados de um sistema quânticos com apenas uma medição:

Proposição 3.1.2 ([11])**.** *Um conjunto de estados $\{\rho_x\}_{x\in\mathcal{X}} \in \mathcal{D}(\mathcal{H}_A)$ é distinguível se e somente se $\{\rho_x\}_{x\in\mathcal{X}}$ forem ortogonais entre si sob o produto interno de Hilbert-Schmidt.*

Assim, vemos que Alice pode armazenar informação clássica de uma fonte $\mathcal{X}$ perfeitamente em $\mathcal{H}_A$ caso $|\mathcal{X}| \leq d_A$, codificando os símbolos $x \in \mathcal{X}$ em estados ortogonais.

Exemplo 3.1.3. Um caso simples é quando $\mathcal{X} = \{0, 1\}$, $\mathcal{H}_A = (\mathbb{C}^{\mathcal{X}})^{\otimes n}$, e para cada $x \in \mathcal{X}$, $\rho_x = |x\rangle\langle x|$, onde $\{|x\rangle\}_{x\in\mathcal{X}}$ é a base computacional de $\mathcal{H}_A$. Aqui é possível armazenar n bits de informação clássica em n qubits. Para recuperar os dados clássicos, basta realizar a medição com respeito à base computacional.

Definição 3.1.4 (Estado Clássico)**.** Um estado $\rho^A \in \mathcal{D}(\mathcal{H}_A)$ é dito um **estado clássico** de A se, para $\{|x\rangle\}_{x\in\mathcal{X}}$ a base computacional de $\mathcal{H}_A$ e $p \in \mathcal{P}(\mathcal{X})$:

$$\rho^A = \sum_{x\in\mathcal{X}} p(x) \left|x\right\rangle\!\left\langle x\right| .$$

Um estado clássico determina (e é determinado por) uma distribuição $p \in \mathcal{P}(\mathcal{X})$. É fácil ver que se ρ é um estado clássico e $U \in \mathcal{B}(\mathcal{H}_A)$ é unitária, então $U\rho U^\dagger$ é estado clássico.

A existência de conjuntos de estados não-distinguíveis está intimamente associada com o seguinte resultado:

Proposição 3.1.5. *Não existe operador unitário $U \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_A)$ que, para quaisquer vetores normalizados $|\psi\rangle, |\varphi\rangle \in \mathcal{H}_A$, satisfaça:*

$$U(|\psi\rangle \otimes |\varphi\rangle) = |\psi\rangle \otimes |\psi\rangle .$$

Implicando que não existe maneira universal de clonar estados puros. Para uma demonstração, ver [10]. Um resultado análogo também vale para estados mistos [48][49].

3.1.2 Estados Clássico-Quânticos

No cenário de armazenamento e acesso de uma fonte clássica X em um sistema quântico A, assumimos que Bob conhece a distribuição p_X, assim seu objetivo é apenas determinar qual emissão de X foi armazenada. Esse é um detalhe importante pois caso o ensemble de Alice tenha estados não-distinguíveis, Bob não poderia reconstruir p_X. A incerteza clássica advinda de X se mistura com a incerteza quântica gerada pela não-distinguibilidade.

Para levar isso em conta, basta considerar o sistema de Alice como sendo o sistema bipartite $\mathbb{C}^{\mathcal{X}} \otimes \mathcal{H}_A$ e que Alice prepara o **ensemble clássico-quântico**:

$$\{p_X(x), |x\rangle\langle x|_X \otimes \rho_x^A\}_{x\in\mathcal{X}}, \tag{3.3}$$

onde $\{|x\rangle_X\}_{x\in\mathcal{X}}$ é a base computacional de $\mathbb{C}^{\mathcal{X}}$. Assim, não há dúvida quanto à distribuição p_X, que pode ser determinada realizando medições locais no sistema $\mathbb{C}^{\mathcal{X}}$. Os estados do ensemble são ortogonais entre si, de forma a separar a incerteza clássica da quântica.

Definição 3.1.6 (Estado Clássico-Quântico)**.** O estado misto correspondente ao ensemble clássico-quântico acima é dito um **estado clássico-quântico** e tem matriz densidade dada por:

$$\rho^{XA} = \sum_{x\in\mathcal{X}} p_X(x)\, |x\rangle\langle x|_X \otimes \rho_x^A. \tag{3.4}$$

Equivalentemente, um estado clássico-quântico em um sistema $\mathbb{C}^{\mathcal{X}} \otimes \mathcal{H}_A$ é um estado separável dado por (3.4) tal que os estados $\{|x\rangle\langle x|_X\}_{x\in\mathcal{X}}$ são distinguíveis.

Note que $\mathrm{Tr}_X\left\{\rho^{XA}\right\} = \rho^A$, com ρ^A dado em (3.1). Para qualquer ensemble $\{p_X(x), \rho_x\}_{x\in\mathcal{X}}$ com $\mathcal{X}$ finito, existe um estado clássico-quântico em $\mathcal{D}(\mathbb{C}^{\mathcal{X}} \otimes \mathcal{H}_A)$ associado, dado por (3.4).

3.1.3 Canais Clássico-Clássicos (C-C)

Vejamos agora que canais clássicos surgem como um caso particular de canal quântico.

Definição 3.1.7 (Canal Clássico-Clássico)**.** Chamamos de **Canal Clássico-Clássico** ou **Canal C-C** um canal quântico $(A, \mathcal{N}^{CC}, B)$ que leva estados clássicos (Definição 3.1.4) de A em estados clássicos de B.

É de se esperar que tais canais correspondam exatamente com os canais clássicos vistos no Capítulo 1. É o que mostraremos agora.

Teorema 3.1.8. *(i) Um canal quântico clássico-clássico $\mathcal{N}^{CC} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ determina um canal clássico $(\mathcal{X}, N, \mathcal{Y})$, dado por:*

$$N(y|x) = \mathrm{Tr}\left\{\mathcal{N}^{CC}(|x\rangle\langle x|)\, |y\rangle\langle y|\right\}.$$

(ii) Reciprocamente, dado um canal clássico $(\mathcal{X}, N, \mathcal{Y})$, existe um canal C-C $\mathcal{N}^{CC} : \mathcal{B}(\mathbb{C}^{\mathcal{X}}) \to \mathcal{B}(\mathbb{C}^{\mathcal{Y}})$ tal que:

$$\mathcal{N}^{CC}(|x\rangle\langle x|) = \sum_{y\in\mathcal{Y}} N(y|x)\, |y\rangle\langle y|\,.$$

Demonstração. Para ver (i), note que como $\mathcal{N}^{CC}$ é canal C-C, para cada $x \in \mathcal{X}$:

$$\mathcal{N}^{CC}(|x\rangle\langle x|) = \sum_{y\in\mathcal{Y}} p_{Y|X}(y|x)\, |y\rangle\langle y|\,,$$

com $p_{Y|X}(\cdot|x) \in \mathcal{P}(\mathcal{Y})$. Definindo $N(x) = p_{Y|X}(\cdot|x)$, temos:

$$N(y|x) = p_{Y|X}(y|x) = \mathrm{Tr}\left\{\mathcal{N}^{CC}(|x\rangle\langle x|)\, |y\rangle\langle y|\right\}.$$

Para ver (ii), considere o canal $\mathcal{N}^{CC}$ com a seguinte representação de Kraus:

$$\left\{\sqrt{N(y|x)}\,|y\rangle\langle x|\right\}_{x\in\mathcal{X},y\in\mathcal{Y}}.$$

Que isso é de fato uma representação de Kraus segue de N ser uma distribuição de probabilidade. Considere a ação do canal no estado clássico $\rho^X = \sum_{x\in\mathcal{X}} p_X(x)\,|x\rangle\langle x|$:

$$\begin{aligned}\mathcal{N}^{CC}(\rho^X) &= \sum_{x,y}\left(\sqrt{N(y|x)}\,|y\rangle\langle x|\right)\rho^X\left(\sqrt{N(y|x)}\,|x\rangle\langle y|\right)\\ &= \sum_{x,y} N(y|x)\,\langle x|\rho^X|x\rangle\,|y\rangle\langle y|\\ &= \sum_{x,y} N(y|x)p_X(x)\,|y\rangle\langle y|\\ &= \sum_{y}\left(\sum_{x} N(y|x)p_X(x)\right)|y\rangle\langle y|\,.\end{aligned}$$

Como $\sum_x N(\cdot|x)p_X(x) \in \mathcal{P}(\mathcal{Y})$, $\mathcal{N}^{CC}(\rho^X)$ é um estado clássico. Em particular, tomando $p_X = \delta_x$, temos:

$$\mathcal{N}^{CC}(|x\rangle\langle x|) = \sum_{y\in\mathcal{Y}} N(y|x)\,|y\rangle\langle y|\,.$$

□

Exemplo 3.1.9 (Canal Clássico Sem Ruído [12])**.** O canal clássico sem ruído $(\mathcal{X}, N(x) \equiv \delta_x, \mathcal{X})$ é representado como canal quântico C-C dado por:

$$\Delta(\rho) = \sum_{x\in\mathcal{X}} \langle x|\rho|x\rangle\,|x\rangle\,\langle x|\,,$$

que no contexto da teoria quântica é chamado de *completely dephasing channel*.

3.1.4 Canais Clássico-Quânticos (C-Q)

Canais quânticos que tomam como entrada estados clássicos tem um papel fundamental no que segue. Tais canais são por vezes definidos na literatura como mapas $\mathcal{N} : \mathcal{X} \to \mathcal{B}(\mathcal{H}_B)$, onde $\mathcal{X}$ é um conjunto finito de índices, ou seja, são uma codificação do alfabeto clássico $\mathcal{X}$ no sistema quântico B. Como aqui canais quânticos são sempre mapas entre operadores, vamos fazer um esforço a fim de obter uma definição adequada ao formalismo desenvolvido até aqui.

Definição 3.1.10 (Canais Clássico-Quânticos)**.** Um canal $\mathcal{N}^{CQ} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ é dito um **canal clássico-quântico** ou **canal C-Q** se existe uma base $\{|x\rangle_A\}_{x\in\mathcal{X}}$ de $\mathcal{H}_A$ e estados $\{\sigma_x^B\}_{x\in\mathcal{X}}$ de forma que a ação de $\mathcal{N}^{CQ}$ em um estado ρ^A é dada por:

$$\mathcal{N}^{CQ}(\rho^A) = \sum_{x\in\mathcal{X}} \langle x|\rho^A|x\rangle_A\,\sigma_x^B \tag{3.5}$$

Denotamos este canal por $(\mathcal{X}, \mathcal{N}^{CQ} \equiv \sigma_x^B, B)$.

Um canal C-Q tem o efeito de uma medição na base $\{|x\rangle_A\}_{x\in\mathcal{X}}$ sobre A e preparo do estado σ_x^B condicionado ao resultado da medição, por isso canais Q-C são também são

chamados de **canais de preparo**. Em particular, dado $p \in \mathcal{P}(\mathcal{X})$, a ação de $\mathcal{N}^{CQ}$ no estado clássico $\rho^A = \sum_{x\in\mathcal{X}} p(x)\,|x\rangle\langle x|$ é:

$$\mathcal{N}^{CQ}(\rho^A) = \sum_{x\in\mathcal{X}} p(x)\sigma_x^B,$$

levando distribuições de probabilidade sobre $\mathcal{X}$ em ensembles $\{p(x), \sigma_x^B\}$ de B. O mapa $x \mapsto \sigma_x^B$ determina inteiramente o canal C-Q. Se escrevermos:

$$\sigma_x^B = \sum_{y=1}^{d_B} |a_{xy}\rangle\langle a_{xy}|\,,$$

e $K_{xy} = |a_{xy}\rangle\langle x|_A$, então $\{K_{xy}\}_{x,y=1}^{d_A,d_B}$ é uma representação de Kraus para $\mathcal{N}^{CQ}$.

3.1.5 Canais Quântico-Clássicos (Q-C)

Definição 3.1.11 (Canal Quântico-Clássico)**.** Seja $\mathcal{Y}$ um conjunto finito com $\{|y\rangle\}_{y\in\mathcal{Y}}$ uma base ortonormal para $\mathbb{C}^{\mathcal{Y}}$. Um canal $\mathcal{N}^{QC} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathbb{C}^{\mathcal{Y}})$ é dito um **canal quântico-clássico** ou **canal Q-C** se existe uma medição $\{\Lambda_y\}_{y\in\mathcal{Y}}$ sobre A tal que, para todo $\rho^A \in \mathcal{D}(\mathcal{H}_{\mathcal{A}})$:

$$\mathcal{N}^{QC}(\rho^A) = \sum_{y\in\mathcal{Y}} \mathrm{Tr}\left\{\rho^A\Lambda_y\right\} |y\rangle\langle y|\,. \tag{3.6}$$

Denotamos este canal por $(A, \mathcal{N}^{QC}, \mathcal{Y})$.

Tomando $\{|j\rangle\}_{j=1}^{d_A}$ a base computacional de $\mathcal{H}_A$ e

$$K_{yj} = |y\rangle\langle j|\,\sqrt{\Lambda_y}, \tag{3.7}$$

então $\{K_{yj}\}_{j=1,y\in\mathcal{Y}}^{d_A}$ é uma representação de Choi-Kraus de $\mathcal{N}^{QC}$. Canais Q-C tem o efeito de realizar uma medição sobre A e registrar o resultado no sistema clássico $\mathcal{Y}$, por isso são chamados também de **canais de medição**.

Reciprocamente, para qualquer medição $\{\Lambda_y\}_{y\in\mathcal{Y}}$ sobre A, podemos definir um canal Q-C associado, dado pela expressão (3.6)

3.1.6 Caracterização de Canais *Entanglement-Breaking*

Lema 3.1.12. *Canais C-C, C-Q e Q-C são entanglement-breaking.*

Demonstração. Parte do Teorema 2.3.24, já que canais C-C, C-Q e Q-C têm representação de Choi-Kraus com operadores de posto um. □

Teorema 3.1.13. *Se $(A, \mathcal{N}^{EB}, B)$ é um canal EB, então existem canais $(A, \mathcal{N}^{QC}, \mathcal{X})$ Q-C e $(\mathcal{X}, \mathcal{N}^{CQ}, B)$ C-Q tais que $\mathcal{N}^{EB} = \mathcal{N}^{CQ} \circ \mathcal{N}^{QC}$.*

Demonstração. Como na demonstração do Teorema 2.3.24, $\mathcal{N}^{EB}$ tem representação de Choi-Kraus dada por $\{K_x \equiv |\varphi_x\rangle_B \langle\psi_x|_A\}_{x=1}^R$, onde vale (2.23), de forma que $\{|\psi_x\rangle\langle\psi_x|_A\}_{x=1}^R$

é um POVM. Tomando $\mathcal{X} = \{1, 2, \ldots, R\}$, seja $\mathcal{N}^{QC}$ o canal Q-C dado por (3.6) com $\Lambda_x = |\psi_x\rangle\langle\psi_x|_A$, e seja $\mathcal{N}^{CQ}$ o canal dado C-Q por (3.5) com $\sigma_x^B = |\varphi_x\rangle\langle\varphi_x|_B$. Então:

$$\begin{aligned}(\mathcal{N}^{CQ} \circ \mathcal{N}^{QC})(\rho^A) &= \sum_{x' \in \mathcal{X}} \langle x'| \left(\sum_{x \in \mathcal{X}} \langle\psi_x|\rho^A|\psi_x\rangle \, |x\rangle\langle x| \right) |x'\rangle \, |\varphi_x\rangle\langle\varphi_x|_B \\ &= \sum_{x \in \mathcal{X}} |\varphi_x\rangle \, \langle\psi_x|\rho^A|\psi_x\rangle \, \langle\varphi_x| \\ &= \sum_{x \in \mathcal{X}} K_x \rho^A K_x^\dagger \\ &= \mathcal{N}^{EB}(\rho^A).\end{aligned}$$

□

Ou seja, canais EB são canais Q-C-Q, também chamados de **canais de medição-preparo**.

3.2 Entropia Quântica

Vamos ver a versão quântica da entropia de Shannon, definida para estados quânticos, independentemente da escolha de matriz densidade para representá-la.

3.2.1 Definição

Definição 3.2.1 (Entropia Quântica)**.** Dado um estado $\rho \in \mathcal{D}(\mathcal{H}_A)$, a **entropia de von Neumann** ou **entropia quântica** de ρ é dada por:

$$S(\rho) = -\mathrm{Tr}\left\{\rho \log \rho\right\}. \tag{3.8}$$

Aqui, $\log(\rho)$ está bem definida para qualquer estado com autovalores não-nulos e é independente da escolha de matriz densidade [50].

Lema 3.2.2 ([50])**.** *Para qualquer $X \in \mathcal{B}(\mathcal{H}_A)$ diagonalizável para o qual $\log(X)$ está bem definido, se $X = UDU^\dagger$ com $D = \mathrm{diag}(x_1, x_2, \ldots, x_{d_A})$ e U unitário, então $\log(X) = U \log(D) U^\dagger$ onde*

$$\log(D) = \mathrm{diag}(\log(x_1), \log(x_2), \ldots, \log(x_{d_A})).$$

É fácil ver que a entropia quântica de um estado ρ é invariante por evolução unitária $\rho \mapsto U\rho U^\dagger$:

$$\begin{aligned}S(U\rho U^\dagger) &= -\mathrm{Tr}\left\{U\rho U^\dagger \log(U\rho U^\dagger)\right\} \\ &= -\mathrm{Tr}\left\{U\rho U^\dagger U \log(\rho) U^\dagger\right\} \\ &= -\mathrm{Tr}\left\{\rho \log \rho\right\} \\ &= S(\rho).\end{aligned}$$

Portanto, para calcular a entropia quântica de um estado, basta considerar a matriz densidade diagonal associada à sua decomposição espectral, lembrando da convenção $0 \cdot \log 0 = 0$.

Teorema 3.2.3. *Se ρ tem decomposição espectral dada por:*

$$\rho = \sum_{x=1}^{d_A} p(x) \left|x\right\rangle\!\left\langle x\right|_A , \tag{3.9}$$

então $S(\rho) = H(p)$.

Demonstração. Basta notar que na base espectral a matriz densidade é a matriz diagonal com entradas $p(x)$. Neste caso $\mathrm{Tr}\left\{\rho \log \rho\right\} = \sum_x p(x) \log(p(x)) = H(p)$. □

Corolário 3.2.4. *Se $\rho \in \mathbb{C}^{\mathcal{X}}$ é um estado clássico correspondente a $p \in \mathcal{P}(\mathcal{X})$, então $S(\rho) = H(p)$.*

Exemplo 3.2.5. Seja $\left|+\right\rangle = \frac{1}{\sqrt{2}}(\left|0\right\rangle + \left|1\right\rangle)$. Considere o estado misto:

$$\rho = \frac{1}{2} \left|0\right\rangle\!\left\langle 0\right| + \frac{1}{2} \left|+\right\rangle\!\left\langle +\right| .$$

Os autovalores de ρ são $\cos^2(\pi/8)$ e $\sin^2(\pi/8)$. A entropia da Shannon da mistura estatística $(\frac{1}{2}, \frac{1}{2})$ entre $\left|0\right\rangle\!\left\langle 0\right|$ e $\left|+\right\rangle\!\left\langle +\right|$ é $H_{bin}(\frac{1}{2}) = 1$, enquanto $S(\rho) = H_{bin}(\cos^2(\pi/8)) = 0.6$. Como veremos na Proposição 3.2.8, a divergência entre os valores se dá pela não-ortogonalidade entre $\left|0\right\rangle$ e $\left|+\right\rangle$.

Teorema 3.2.6. *Para todo $\rho \in \mathcal{D}(\mathcal{H}_A)$, $0 \leq S(\rho) \leq \log d_A$. E ainda:*

(1) $S(\rho) = 0$ se, e somente se, ρ é estado puro.

(2) $S(\rho) = \log d_A$ se, e somente se, ρ é o estado maximamente misto $\frac{1}{d_A}\mathbb{1}_A$.

Demonstração. Pelo Teorema 3.2.3, se ρ tem decomposição espectral (3.9), $S(\rho) = H(p)$. *(1)* segue pois $H(p) = 0$ se e somente se $p = \delta_x$ com $1 \leq x \leq d_A$, que ocorre exatamente quando ρ é um estado puro. Também por (3.9), $S(\rho) = H(p) = \log d_A$ se e somente se $p(x) = 1/d_A$. □

A seguinte proposição segue facilmente da Decomposição de Schmidt (Proposição 2.2.6).

Proposição 3.2.7. *Seja ρ^{AB} um estado puro do sistema bipartite AB, então:*

$$S(\rho^A) = S(\rho^B).$$

Proposição 3.2.8 (Entropia Quântica do Ensemble [11])**.** *Dado um ensemble $\{p(x), \rho_x\}_{x \in \mathcal{X}}$ em A, temos:*

$$S\left(\sum_{x \in \mathcal{X}} p(x)\rho_x\right) \leq \sum_{x \in \mathcal{X}} p(x) S(\rho_x) + H(p), \tag{3.10}$$

com igualdade se e somente se os estados são ortogonais. Em particular:

$$S\left(\sum_{x \in \mathcal{X}} p(x) \left|x\right\rangle\!\left\langle x\right| \otimes \rho_x^A\right) = H(p) + \sum_{x \in \mathcal{X}} p(x) S(\rho_x^A). \tag{3.11}$$

3.2.2 Entropia Relativa Quântica

Podemos também definir a versão quântica da entropia relativa (Definição 1.1.25).

Definição 3.2.9 (Entropia Relativa Quântica)**.** Dados estados $\rho, \sigma \in \mathcal{D}(\mathcal{H}_A)$, a **entropia relativa quântica** de ρ para σ é dada por:

$$D(\rho||\sigma) = \mathrm{Tr}\left\{\rho(\log\rho - \log\sigma)\right\}$$

se $\mathrm{supp}(\rho) \subset \mathrm{supp}(\sigma)$ e $+\infty$ caso contrário.

É fácil verificar o seguinte resultado, análogo ao Teorema 1.1.27:

Proposição 3.2.10. $S(\rho^A) = \log d_A - D(\rho^A||\frac{1}{d_A}\mathbb{1}_A)$.

Se ρ e σ forem estados clássicos sob a mesma base, então $D(\rho||\sigma)$ se reduz à entropia relativa clássica (Definição 1.1.25). Naquele caso a não-negatividade era garantida pela Desigualdade de Gibbs (1.1.26). Temos um resultado análogo no caso quântico:

Proposição 3.2.11 (Desigualdade de Klein [5, 13])**.** *Para quaisquer $\rho, \sigma \in \mathcal{D}(\mathcal{H}_A)$:*

$$D(\rho||\sigma) \geq 0,$$

com igualdade se e somente se $\rho = \sigma$.

Lema 3.2.12. *Para todo σ^{AB},* $\mathrm{Tr}\left\{\sigma^{AB}(X \otimes \mathbb{1}_B)\right\} = \mathrm{Tr}\left\{\mathrm{Tr}_B\left\{\sigma^{AB}\right\} X\right\}$.

Lema 3.2.13. *Para $\rho^A \in \mathcal{D}(\mathcal{H}_A)$ e $\rho^B \in \mathcal{D}(\mathcal{H}_B)$,*

$$\log(\rho^A \otimes \rho^B) = \log(\rho^A) \otimes \mathbb{1}_B + \mathbb{1}_A \otimes \log(\rho^B).$$

Teorema 3.2.14 (Sub-aditividade da Entropia Quântica)**.** *Para $\rho^{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$, $S(\rho^{AB}) \leq S(\rho^A) + S(\rho^B)$, com igualdade se, e somente se, $\rho^{AB} = \rho^A \otimes \rho^B$.*

Demonstração. Pela Desigualdade de Klein:

$$S(\rho^{AB}) \leq -\mathrm{Tr}\left\{\rho^{AB}\log(\rho^A \otimes \rho^B)\right\}$$

(e pelos Lemas 3.2.12 e 3.2.13, lembrando que $\rho^A = \mathrm{Tr}_B\left\{\rho^{AB}\right\}$)

$$\begin{aligned}
&= -\mathrm{Tr}\left\{\rho^{AB}(\log(\rho^A) \otimes \mathbb{1}_B + \mathbb{1}_A \otimes \log(\rho^B))\right\} \\
&= -\mathrm{Tr}\left\{\rho^{AB}(\log(\rho^A) \otimes \mathbb{1}_B)\right\} - \mathrm{Tr}\left\{\rho^{AB}(\mathbb{1}_A \otimes \log(\rho^B))\right\} \\
&= -\mathrm{Tr}\left\{\rho^{A}(\log(\rho^A) \otimes \mathbb{1}_B)\right\} - \mathrm{Tr}\left\{\rho^{B}(\mathbb{1}_A \otimes \log(\rho^B))\right\} \\
&= -\mathrm{Tr}\left\{\rho^{A}\log(\rho^A)\right\} - \mathrm{Tr}\left\{\rho^{B}\log(\rho^B)\right\} \\
&= S(\rho^A) + S(\rho^B),
\end{aligned}$$

onde usamos que:

$$\mathrm{Tr}\left\{\rho^A(\log(\rho^A) \otimes \mathbb{1}_B)\right\} = \mathrm{Tr}_A\left\{\mathrm{Tr}_B\left\{\rho^A(\log(\rho^A) \otimes \mathbb{1}_B)\right\}\right\} = \mathrm{Tr}_A\left\{\rho^A\log(\rho^A)\right\} = \mathrm{Tr}\left\{\rho^A\log(\rho^A)\right\}.$$

□

Proposição 3.2.15 (Desigualdade do Processamento de Dados Quântica [13])**.** *Dados estados $\rho, \sigma \in \mathcal{D}(\mathcal{H}_A)$, e um canal quântico $(A, \mathcal{N}, B)$:*

$$D(\mathcal{N}(\rho)||\mathcal{N}(\sigma)) \leq D(\rho||\sigma).$$

Em particular, considerando o traço parcial como canal quântico (Exemplo 2.3.14):

Corolário 3.2.16 (Monotonicidade da Entropia Relativa Quântica)**.** *Para estados $\rho^{AB}, \sigma^{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$:*

$$D(\rho^A||\sigma^A) \leq D(\rho^{AB}||\sigma^{AB}).$$

3.2.3 Informação Mútua Quântica

Apesar de não existir uma noção na teoria quântica análoga à probabilidade condicional, podemos reproduzir a fórmula do Teorema 1.1.12 usando a entropia quântica:

Definição 3.2.17 (Entropia Condicional Quântica)**.** Para um estado ρ^{AB} de um sistema bipartite AB, a **entropia condicional quântica** de A dado B no estado ρ^{AB} é:

$$S(A|B)_\rho = S(\rho^{AB}) - S(\rho^B).$$

Ao contrário da entropia condicional clássica, a entropia condicional quântica pode ser negativa. De fato, pela Proposição 3.2.6, para um estado puro ρ^{AB} do sistema bipartite AB, $S(A|B)_\rho = -S(\rho^B)$.

Definição 3.2.18 (Informação Mútua Quântica)**.** Para um estado ρ^{AB} de um sistema bipartite AB a **informação mútua quântica** entre A e B no estado ρ^{AB} é:

$$\mathcal{I}(A:B)_\rho = S(\rho^A) + S(\rho^B) - S(\rho^{AB}).$$

Analogamente à informação mútua clássica, temos:

$$\mathcal{I}(A:B)_\rho = S(\rho^A) - S(A|B)_\rho = S(\rho^B) - S(B|A)_\rho. \tag{3.12}$$

Teorema 3.2.19. *Para um estado ρ^{AB} de um sistema bipartite AB:*

$$\mathcal{I}(A:B)_\rho = D(\rho^{AB}||\rho^A \otimes \rho^B).$$

Demonstração. Pelos Lemas 3.2.12 e 3.2.13:

$$\begin{aligned} D(\rho^{AB}||\rho^A \otimes \rho^B) &= \mathrm{Tr}_{AB}\left\{\rho^{AB}(\log \rho^{AB} - \log \rho^A \otimes \rho^B)\right\} \\ &= \mathrm{Tr}_{AB}\left\{\rho^{AB}(\log \rho^{AB} - \log \rho^A \otimes \mathbb{1}_B - \mathbb{1}_A \otimes \log \rho^B)\right\} \\ &= S(\rho^A) + S(\rho^B) - S(\rho^{AB}) \\ &= \mathcal{I}(A:B)_\rho. \end{aligned}$$

□

Corolário 3.2.20. *Para qualquer estado ρ^{AB} de AB, $\mathcal{I}(A:B) \geq 0$.*

Teorema 3.2.21 (Subaditividade Forte da Entropia Quântica)**.** *Sempre temos que:*

$$S(\rho^{ABC}) + S(\rho^B) \leq S(\rho^{AB}) + S(\rho^{BC}).$$

Equivalentemente, $\mathcal{I}(A:B)_\rho \leq \mathcal{I}(A:BC)_\rho$.

Demonstração. Seja $\sigma^{ABC} = \rho^A \otimes \rho^{BC}$. Pelo Corolário 3.2.16:

$$D(\rho^{AB}||\sigma^{AB}) \leq D(\rho^{ABC}||\sigma^{ABC}) \tag{3.13}$$

e, pelo Teorema 3.2.19:

$$D(\rho^{ABC}||\sigma^{ABC}) = \mathcal{I}(A:BC)_\rho = S(\rho^A) + S(\rho^{BC}) - S(\rho^{ABC}).$$

Similarmente:

$$D(\rho^{AB}||\sigma^{AB}) = \mathcal{I}(A:B)_\rho = S(\rho^A) + S(\rho^B) - S(\rho^{AB}).$$

Substituindo em (3.13), obtemos o resultado. □

3.2.4 A Quantidade de Holevo

Vamos agora considerar as quantidades entrópicas quânticas de estados clássico-quânticos.

Teorema 3.2.22. *Se ρ^{XA} é o estado clássico-quântico associado ao ensemble $\{p_X(x), \rho_x^A\}_{x\in\mathcal{X}}$, então:*

$$S(A|X)_\rho = \sum_{x\in\mathcal{X}} p_X(x) S(\rho_x^A).$$

Demonstração. Pela Proposição 3.2.8:

$$S(A|X)_\rho = S(\rho^{XA}) - S(\rho^X) = H(p_X) + \sum_{x\in\mathcal{X}} p_X(x) S(\rho_x^A) - H(p_X)$$

□

Por (3.12) e pelo Teorema 3.2.22, a informação mútua quântica do estado clássico-quântico ρ^{XA} é:

$$\mathcal{I}(X:A)_\rho = S(\rho^A) - S(A|X)_\rho = S\left(\sum_{x\in\mathcal{X}} p_X(x)\rho_x^A\right) - \sum_{x\in\mathcal{X}} p_X(x) S(\rho_x^A). \tag{3.14}$$

Definição 3.2.23 (Quantidade de Holevo)**.** Dado um ensemble $\mathcal{E} = \{p_X(x), \rho_x^A\}_{x\in\mathcal{X}}$ de A, definimos:

$$\chi(\mathcal{E}) = \mathcal{I}(X:A)_\rho = S\left(\sum_{x\in\mathcal{X}} p_X(x)\rho_x^A\right) - \sum_{x\in\mathcal{X}} p_X(x) S(\rho_x^A), \tag{3.15}$$

com $\rho^{XA} = \sum_{x\in\mathcal{X}} p(x)\,|x\rangle\langle x| \otimes \rho_x^A$.

Por (3.15) e pelo Corolário 3.2.20, a quantidade de Holevo é sempre não-negativa. Como consequência obtemos a concavidade da entropia quântica.

Corolário 3.2.24 (Concavidade da Entropia Quântica)**.** *Dado um ensemble de A $\mathcal{E} = \{p_X(x), \rho_x^A\}_{x\in\mathcal{X}}$:*

$$S\left(\sum_{x\in\mathcal{X}} p_X(x)\rho_x^A\right) \geq \sum_{x\in\mathcal{X}} p_X(x) S(\rho_x^A).$$

3.2.5 O Teorema de Holevo

Para um ensemble $\{p_X(x), \rho_x\}_{x\in\mathcal{X}}$ e medição $\{\Lambda_y\}_{y\in\mathcal{Y}}$ fixados, temos exatamente o cenário de comunicação por canal que estudamos no Capítulo 1, onde o papel do ruído é feito pela incerteza devido à indistinguibilidade dos estados do ensemble. Definimos o canal clássico $N_\Lambda : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ por:

$$N_\Lambda(y|x) = \langle \rho_x, \Lambda_y \rangle_{HS} = \mathrm{Tr}\left\{\rho_x \Lambda_y\right\}.$$

Tomamos Y como sendo a resultado da medição, ou seja, é uma variável aleatória com valores em $\mathcal{Y}$ e com distribuição tal que $p_{X,Y} = p_X(x) N_\Lambda(y|x)$.

Definição 3.2.25 (Informação Acessível)**.** A **informação acessível** de um ensemble $\{p_X(x), \rho_x\}_{x\in\mathcal{X}}$ em $\mathcal{H}_A$, é dada por:

$$I_{acc}(\{p_X, \rho_x\}) = \sup_{\Lambda} I(X:Y),$$

onde o supremo é tomado sob todas as medições $\Lambda : \mathcal{Y} \to \mathcal{B}(\mathcal{H}_A)^+$ para qualquer $\mathcal{Y}$.

Ou seja, é a quantidade máxima de informação clássica que pode ser armazenada fielmente no ensemble $\{p_X(x), \rho_x\}_{x\in\mathcal{X}}$. Dado qualquer conjunto de estados quânticos $\{\rho_x\}_{x\in\mathcal{X}}$, temos:

$$\begin{aligned}\sup_{p_X\in\mathcal{P}(\mathcal{X})} I_{acc}(\{p_X, \rho_x\}) &= \sup_{p_X\in\mathcal{P}(\mathcal{X})} \sup_{\Lambda} I(X:Y) \\ &= \sup_{\Lambda} \sup_{p_X\in\mathcal{P}(\mathcal{X})} I(X:Y) \\ &= \sup_{\Lambda} C(N_\Lambda), \qquad (3.16)\end{aligned}$$

onde (3.16) segue do Teorema da Codificação de Canais. Se $\{p_X, \rho_x\}_{x\in\mathcal{X}}$ é um ensemble de estados distinguíveis e Λ é uma medição que os distingue, então:

$$I_{acc}(\{p_X, \rho_x\}) = H(X).$$

Proposição 3.2.26 ([51][52])**.** *Dado um ensemble $\{p_X, \rho_x\}$ em $\mathcal{H}_A$, existe uma medição $\Lambda : \mathcal{Y} \to \mathcal{B}(\mathcal{H}_A)^+$ tal que $I_{acc}(\{p_X, \rho_x\}) = I(X:Y)$ com $|\mathcal{Y}| \leq d_A^2$.*

Isso nos dá a cota superior $I_{acc}(\{p_X, \rho_x\}) < 2\log d_A$, por (1.3). Note que essa cota não depende do ensemble, que implica que, apesar de possuir infinitos estados quânticos, o sistema quântico A pode armazenar apenas uma quantidade finita de informação clássica.

Calcular a informação acessível de um ensemble não é um problema trivial e é de grande interesse (ver [53, 54]). A dificuldade é devido ao supremo ser sobre qualquer POVM. Vejamos porém que a quantidade de Holevo do ensemble sempre fornece uma cota superior para a informação acessível.

Teorema 3.2.27 (Teorema de Holevo)**.** *Para qualquer ensemble $\mathcal{E} = \{p_X(x), \rho_x^A\}_{x\in\mathcal{X}}$ de estados de $\mathcal{H}_A$, temos que:*

$$I_{acc}(\mathcal{E}) \leq \chi(\mathcal{E}), \qquad (3.17)$$

com igualdade se e somente se os operadores $p_X(x)\rho_x^A$ comutam entre si.

Demonstração. Seja ρ^{XA} o estado clássico-quântico (3.4) associado a $\mathcal{E}$. Por (3.15) e pelo Teorema 3.2.19, $\chi(\mathcal{E}) = \mathcal{I}(X:A)_\rho = D(\rho^{XA}||\rho^X \otimes \rho^A)$. Dada uma medição $\{\Lambda_y\}_{y\in\mathcal{Y}}$, seja $(A, \mathcal{N}, \mathcal{Y})$ o canal Q-C associado (Definição 3.1.11). Então:

$$\begin{aligned}(\mathbb{1}_X \otimes \mathcal{N})(\rho^{XA}) &= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \mathrm{Tr}\left\{\rho_x^A \Lambda_y\right\} |x\rangle\langle x| \otimes |y\rangle\langle y| \\ &= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} N_\Lambda(y|x)\, |x\rangle\langle x| \otimes |y\rangle\langle y| \\ &= \mathrm{Diag}(p_{X,Y}),\end{aligned}$$

onde $p_{X,Y}$ é o vetor de probabilidade em $\mathcal{P}(\mathcal{X}\times\mathcal{Y})$ definido por $p_{X,Y}(x,y) = p_X(x)N_\Lambda(y|x)$, que interpretamos como estado clássico em $\mathcal{D}(\mathbb{C}^{\mathcal{X}} \otimes \mathbb{C}^{\mathcal{Y}})$. Neste caso a informação mútua clássica e a informação mútua quântica coincidem e temos:

$$\begin{aligned} I(X:Y) &= D(p_{X,Y}||p_X \times p_Y) \\ &= D((\mathbb{1}_X \otimes \mathcal{N})(\rho^{XA})||\rho^X \otimes \mathcal{N}(\rho^A)) \\ &\leq D(\rho^{XA}||\rho^X \otimes \rho^A) \\ &= \chi(\mathcal{E}), \end{aligned}$$

com a desigualdade seguindo da Proposição 3.2.15. O teorema segue tomando o supremo entre as medições. Uma demonstração para o critério de igualdade pode ser encontrada em [13]. □

Corolário 3.2.28. *Para qualquer ensemble $\mathcal{E} = \{p_X(x), \rho_x^A\}_{x\in\mathcal{X}}$ de A, temos que:*

$$\sup_\Lambda C(N_\Lambda) \leq \sup_{p\in\mathcal{P}(\mathcal{X})} \chi(\mathcal{E}) \leq \log(d_A). \tag{3.18}$$

O corolário acima mostra que um qubit pode armazenar fielmente no máximo um bit de informação clássica.

Capítulo 4

Comunicação Clássica por Canais Quânticos

Nesse capítulo, segundo principalmente [34, 13, 14], vamos explorar o uso de canais quânticos na transmissão de informação clássica, ou seja, de como obter um canal clássico a partir de um canal quântico. Esse é um processo que pode ser feito de diversas maneiras, dependendo de quais recursos quânticos utilizamos, levando a diferentes noções de capacidade. Na Seção 4.1, introduzimos diversas quantidades relacionadas à transmissão de informação clássica por canais quânticos. Definimos formalmente a noção de código quântico e de capacidade clássica de canais quânticos, que são dadas interpretações operacionais em termos de esquemas de codificação. Enunciamos então o Teorema de HSW, que é o análogo quântico do Teorema da Codificação de Canais, e citamos algumas de suas consequências. A Seção 4.2 se dedica ao estudo da conjectura da aditividade, que permaneceu um dos principais problemas em aberto da teoria da informação quântica, que se mostrou falsa.

4.1 Capacidade Clássica de Canais Quânticos

Começamos ilustrando uma maneira de obter um canal clássico a partir de um canal quântico.

Dado o canal quântico $(A, \mathcal{N}, B)$, considere que Alice quer enviar a Bob uma letra do alfabeto $\mathcal{X}$. Para isso, ela codifica cada $x \in \mathcal{X}$ em um estado $\mathcal{Q}(x) = \rho_x^A \in \mathcal{D}(\mathcal{H}_A)$. Para enviar x, ela prepara o estado ρ_x^A e o envia por $\mathcal{N}$, de forma que Bob recebe o estado $\sigma_x^B = \mathcal{N}(\rho_x^A) \in \mathcal{D}(\mathcal{H}_B)$ que ele acessa realizando uma medição dada por um POVM $\{\Lambda_y\}_{y\in\mathcal{Y}}$ sobre o sistema B, obtendo um resultado $y \in \mathcal{Y}$. Se Alice quer transmitir a variável aleatória X com valores em $\mathcal{X}$, ela prepara o ensemble $\mathcal{E} = \{p_X(x), \rho_x^A\}$.

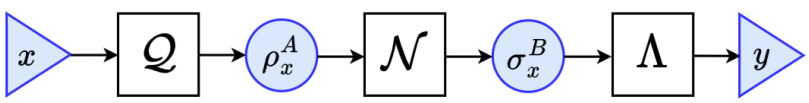

Figura 4.1: Esquema de comunicação clássica por um canal quântico. Triângulos representam informação clássica e círculos representam informação quântica.

Fixada uma codificação $\mathcal{Q}(x) = \rho_x^A$ e uma medição $\{\Lambda_y\}_{y\in\mathcal{Y}}$, obtemos um canal clássico $(\mathcal{X}, N_{\mathcal{Q},\Lambda}, \mathcal{Y})$ dado por:

$$N_{\mathcal{Q},\Lambda}(y|x) = \mathrm{Tr}\left\{\mathcal{N}(\rho_x^A)\Lambda_y\right\}. \tag{4.1}$$

A capacidade deste canal é dada pelo Teorema da Codificação de Canais:

$$C(N_{\mathcal{Q},\Lambda}) = \sup_{p_X\in\mathcal{P}(\mathcal{X})} I(X : N_{\mathcal{Q},\Lambda}(X)) = \sup_{p_X\in\mathcal{P}(\mathcal{X})} \left\{ H(N_{\mathcal{Q},\Lambda}(p_X)) - \sum_{x\in\mathcal{X}} p_X(x) N_{\mathcal{Q},\Lambda}(x) \right\}.$$

Otimizando em relação à medição, obtemos:

$$\begin{aligned}
\sup_{\Lambda} C(N_{\mathcal{Q},\Lambda}) &= \sup_{\Lambda} \sup_{p_X\in\mathcal{P}(\mathcal{X})} I(X : N_{\mathcal{Q},\Lambda}(X)) && (4.2)\\
&= \sup_{p_X\in\mathcal{P}(\mathcal{X})} \sup_{\Lambda} I(X : N_{\mathcal{Q},\Lambda}(X)) && (4.3)\\
&= \sup_{p_X\in\mathcal{P}(\mathcal{X})} I_{acc}(\{p_X(x), \mathcal{N}(\rho_x^A)\}). && (4.4)
\end{aligned}$$

Otimizando também em relação ao ensemble, ou seja, em relação a $\mathcal{Q}$ e p_X, temos um possível candidato à capacidade clássica de um canal quântico, que vamos definir formalmente adiante.

Definição 4.1.1 (Informação Acessível do Canal)**.** A **informação acessível** do canal quântico $(A, \mathcal{N}, B)$ é definida por:

$$I_{acc}(\mathcal{N}) = \sup_{\mathcal{Q},\Lambda} C(N_{\mathcal{Q},\Lambda}) = \sup_{\mathcal{E}} I_{acc}(\{p_X(x), \mathcal{N}(\rho_x^A)\}) \tag{4.5}$$

Com o supremo na expressão à direita em (4.5) acima sendo sobre os possíveis ensembles $\mathcal{E} = \{p_X(x), \rho_x^A\}$. Para facilitar a exposição, vamos introduzir a seguinte notação: Fixados um ensemble $\mathcal{E} = \{p_X(x), \mathcal{Q}(x)\}$ e um POVM $\{\Lambda_y\}_{y\in\mathcal{Y}}$, denotamos por:

$$I_{\mathcal{N}}(\mathcal{E} : \Lambda) = I(X : N_{\mathcal{Q},\Lambda}(X)) = H(N_{\mathcal{Q},\Lambda}(p_X)) - \sum_{x\in\mathcal{X}} p_X(x) N_{\mathcal{Q},\Lambda}(x) \tag{4.6}$$

a informação mútua clássica entre a entrada X com distribuição p_X e saída $N_{\mathcal{Q},\Lambda}(X)$, resultantes do esquema de codificação $\mathcal{Q}$ e decodificação Λ.

Proposição 4.1.2 ([55])**.** *O supremo sobre ensembles $\mathcal{E}$ de A em* (4.5) *pode ser substituído pelo máximo entre ensembles de estados puros $\{p(x), |\psi_x\rangle\langle\psi_x|\}_{x\in\mathcal{X}}$ com $|\mathcal{X}| \leq d_A^2 + 1$.*

Por (4.4), Proposição 3.2.26 e Proposição 4.1.2:

$$I_{acc}(\mathcal{N}) = \max_{\mathcal{E},\Lambda} I_{\mathcal{N}}(\mathcal{E} : \Lambda).$$

Assim como no caso clássico, a noção de capacidade deve levar em consideração múltiplos usos independentes do canal. Vamos ver que a informação acessível falha em capturar o comportamento desses casos, pois não é uma quantidade aditiva. Por exemplo, taxa de transmissão (que vamos definir para canais quânticos a seguir) maior é atingível codificando o canal produto $\mathcal{N}\otimes\mathcal{N}\otimes\mathcal{N}$. Nesse caso é possível atingir a taxa de transmissão:

$$\frac{1}{3} I_{acc}(\mathcal{N}\otimes\mathcal{N}\otimes\mathcal{N}),$$

que é uma quantidade pelo menos tão grande quanto $I_{acc}(\mathcal{N})$:

Teorema 4.1.3. *Para quaisquer canais quânticos* $(A,\mathcal{N}_1,B)$, $(C,\mathcal{N}_2,D)$*:*

$$I_{acc}(\mathcal{N}_1)+I_{acc}(\mathcal{N}_2)\leq I_{acc}(\mathcal{N}_1\otimes\mathcal{N}_2). \tag{4.7}$$

Demonstração. Sejam $\mathcal{Q}_1,\mathcal{Q}_2$ e Λ_1,Λ_2 tais que $I_{acc}(\mathcal{N}_1)=C(N_{\mathcal{Q}_1,\Lambda_1})$ e $I_{acc}(\mathcal{N}_2)=C(N_{\mathcal{Q}_2,\Lambda_2})$. Tomando $\mathcal{Q}=\mathcal{Q}_1\otimes\mathcal{Q}_2$ e $\Lambda=\Lambda_1\otimes\Lambda_2$ para $\mathcal{N}_1\otimes\mathcal{N}_2$, temos que $N_{\mathcal{Q},\Lambda}=N_{\mathcal{Q}_1,\Lambda_1}\times N_{\mathcal{Q}_2,\Lambda_2}$. Pela superaditividade da capacidade de canais clássicos (Teorema 1.4.2):

$$C(N_{\mathcal{Q}_1,\Lambda_1})+C(N_{\mathcal{Q}_2,\Lambda_2})\leq C(N_{\mathcal{Q}_1\otimes\mathcal{Q}_2,\Lambda_1\otimes\Lambda_2})\leq I_{acc}(\mathcal{N}_1\otimes\mathcal{N}_2).$$

□

Nosso interesse no comportamento assintótico de múltiplos usos independentes do canal nos leva a considerar a quantidade:

$$I_{reg}(\mathcal{N})=\sup_n\frac{1}{n}I_{acc}(\mathcal{N}^{\otimes n})=\lim_{n\to\infty}\frac{1}{n}I_{acc}(\mathcal{N}^{\otimes n}), \tag{4.8}$$

dita **informação acessível regularizada** de $\mathcal{N}$. A segunda igualdade em (4.8) segue da superaditividade da informação acessível (Teorema 4.1.3) (ver também appendix A em [56]). Esta expressão é de pouco uso computacionalmente. Ao contrário do caso para canais clássicos, essa expressão não se reduz ao caso de um único uso do canal, o motivo sendo a superaditividade estrita da informação acessível.

Proposição 4.1.4. *A informação acessível não é aditiva.*

Isto foi primeiro demonstrado por Holevo em [57]. Veja [58] para um exemplo de canal $\mathcal{N}$ com $I_{acc}(\mathcal{N})<\frac{1}{3}I_{acc}(\mathcal{N}\otimes\mathcal{N}\otimes\mathcal{N})$.

Assim como na discussão sobre informação acessível de um ensemble, aqui podemos obter uma cota superior mais simples de calcular. A expressão à direita em (4.5) e o Teorema de Holevo (3.2.27) sugerem a seguinte definição:

Definição 4.1.5 (Capacidade de Holevo)**.** Dado um canal quântico $(A,\mathcal{N},B)$, a **capacidade de Holevo** $\mathcal{N}$ é definida por:

$$\chi(\mathcal{N})=\sup_{\{p(x),\rho_x^A\}_{x\in\mathcal{X}}}\chi(\{p(x),\mathcal{N}(\rho_x^A)\}_{x\in\mathcal{X}}), \tag{4.9}$$

com supremo sobre todos os ensembles $\{p(x),\rho_x^A\}_{x\in\mathcal{X}}$ sobre qualquer $\mathcal{X}$ finito.

Proposição 4.1.6 ([55])**.** *O supremo sobre ensembles de A em* (4.9) *pode ser substituído pelo máximo entre ensembles de estados puros* $\{p(x),|\psi_x\rangle\langle\psi_x|\}_{x\in\mathcal{X}}$ *com* $|\mathcal{X}|\leq d_A^2$.

Pelo Teorema de Holevo (3.2.27) e (4.5), sempre temos:

$$I_{acc}(\mathcal{N})\leq\chi(\mathcal{N}). \tag{4.10}$$

Teorema 4.1.7. *Para quaisquer canais quânticos* $(A,\mathcal{N}_1,B)$, $(C,\mathcal{N}_2,D)$*:*

$$\chi(\mathcal{N}_1)+\chi(\mathcal{N}_2)\leq\chi(\mathcal{N}_1\otimes\mathcal{N}_2).$$

Demonstração. Segue pois para quaisquer ensembles $\{p(x), \rho_x^A\}_{x\in\mathcal{X}}$ e $\{q(y), \rho_y^C\}_{y\in\mathcal{Y}}$ em A e C respectivamente:

$$\begin{aligned}\chi(\{p(x), \mathcal{N}_1(\rho_x^A)\}) + \chi(\{q(y), \mathcal{N}_2(\rho_y^C)\}) &= \chi(\{p(x)q(y), \mathcal{N}_1(\rho_x^A) \otimes \mathcal{N}_2(\rho_y^C)\}) \\ &\leq \chi(\mathcal{N}_1 \otimes \mathcal{N}_2).\end{aligned}$$

□

Se e quando a capacidade de Holevo é ou não uma quantidade aditiva é uma questão delicada, à qual iremos dedicar a Seção 4.2. Devemos então considerar a **quantidade de Holevo regularizada**:

$$\chi_{reg}(\mathcal{N}) = \sup_n \frac{1}{n}\chi(\mathcal{N}^{\otimes n}) = \lim_{n\to\infty} \frac{1}{n}\chi(\mathcal{N}^{\otimes n}).$$

A definição de códigos para comunicação clássica por um canal quântico é análoga ao caso clássico (Definição 1.2.6):

Definição 4.1.8 (Código de Canal Quântico)**.** Um (M, n)-código para o canal quântico $(A, \mathcal{N}, B)$ consiste em:

- Um conjunto $\mathcal{M} = \{1, 2, \ldots, M\}$ de mensagens e conjuntos finitos $\mathcal{X}$ e $\mathcal{Y}$, ditos alfabetos de entrada e saída, respectivamente.
- Uma livro-código clássico $\mathcal{C}^n : \mathcal{M} \to \mathcal{X}^n$.
- Um canal C-Q $(\mathcal{X}^n, \mathcal{Q}^n, A^n)$ que age por $x^n \mapsto \rho_{x^n}^{A^n}$.
- Um canal Q-C $(B^n, \Lambda^n, \mathcal{Y}^n)$ associado à medição sobre B dada por $\{\Lambda_{y^n}^n\}_{y^n\in\mathcal{Y}^n}$.
- Uma decodificação clássica $\mathcal{D}^n : \mathcal{Y}^n \to \mathcal{M}$.

Para enviar a mensagem m codificada pela palavra-código $x^n(m)$, Alice prepara o estado $\mathcal{Q}^n(x^n(m)) = \rho_{x^n(m)}^{A^n}$ e o envia pelo canal $\mathcal{N}^{\otimes n}$. Bob recebe o estado $\mathcal{N}^{\otimes n}(\rho_{x^n(m)}^{A^n}) = \sigma_{x^n(m)}^{B^n} \in \mathcal{D}(\mathcal{H}_B^{\otimes n})$ e realiza a medição $\{\Lambda_{y^n}^n\}_{y^n\in\mathcal{Y}^n}$ obtendo uma palavra $y^n \in \mathcal{Y}^n$, que ele decodifica como $\mathcal{D}^n(y^n) = \hat{m}$. A probabilidade de que Bob decodifica corretamente a mensagem m é:

$$\sum_{y^n:\mathcal{D}_n(y^n)=m} \mathrm{Tr}\left\{\sigma_{x^n(m)}^{B^n}\Lambda_{y^n}\right\}.$$

Assim como no caso clássico, assumimos que as mensagens de $\mathcal{M}$ são emitidas uniformemente, induzindo uma distribuição $p_{X^n} \sim \mathcal{C}^n(W)$ em $\mathcal{X}^n$. Se X^n e Y^n denotam as entradas de Q^n e saídas de Λ^n respectivamente, a informação mútua clássica $I(X^n : Y^n)$ sob esse código é $I_{\mathcal{N}^{\otimes n}}(\mathcal{E}^n : \Lambda^n)$, dada por (4.6), com ensemble $\mathcal{E}^n = \{p_{X^n}(x^n), \rho_{x^n(m)}^{A^n}\}_{x^n\in\mathcal{X}^n}$, por isso denotamos esse código por $(\mathcal{E}^n, \Lambda^n)$. Omitimos os índices n quando o contexto for claro.

Definição 4.1.9. Dado um (M, n)-código $(\mathcal{E}, \Lambda)$ para o canal quântico $(A, \mathcal{N}, B)$, definimos:

- Probabilidade condicional de erro para a mensagem m:

$$\begin{aligned} P_e(\mathcal{E}, \Lambda)(m) &= 1 - \sum_{y^n : \mathcal{D}_n(y^n) = m} \operatorname{Tr}\left\{\sigma^{B^n}_{x^n(m)} \Lambda_{y^n}\right\} \\ &= \sum_{y^n : \mathcal{D}_n(y^n) = m} \operatorname{Tr}\left\{\sigma^{B^n}_{x^n(m)} (\mathbb{1}_B - \Lambda_{y^n})\right\}. \end{aligned}$$

- Probabilidade média de erro:

$$P_e(\mathcal{E}, \Lambda) = \frac{1}{M} \sum_{m \in \mathcal{M}} P_e(\mathcal{E}, \Lambda)(m).$$

- A taxa de transmissão é dada por:

$$R = \frac{\log M}{n}.$$

No que segue, denotamos por:

$$p_e(M, n) = \min_{(\mathcal{E}, \Lambda)} P_e(\mathcal{E}, \Lambda) \tag{4.11}$$

a probabilidade média de erro minimizada entre todos os (M, n)-códigos $(\mathcal{E}, \Lambda)$ para $\mathcal{N}$.

Definição 4.1.10 (Capacidade Clássica)**.** Uma taxa $R \geq 0$ é dita atingível para o canal quântico $(A, \mathcal{N}, B)$ se $R = 0$ ou se existe uma sequência de (M_n, n)-códigos para $\mathcal{N}$ tais que

$$R = \lim_{n \to \infty} \frac{\log M_n}{n}, \qquad \lim_{n \to \infty} p_e(M_n, n) = 0.$$

A **capacidade clássica** de $\mathcal{N}$ é definida por:

$$C(\mathcal{N}) = \sup\{R \geq 0 \,|\, R \text{ é atingível para } \mathcal{N}\}.$$

Essa definição é a versão quântica do que chamamos de capacidade operacional no Capítulo 1 e, assim como naquele caso, essa definição não indica uma maneira de calcular a capacidade clássica de um canal específico. No caso clássico isso foi remediado pelo Teorema da Codificação de Canais, que fornece a capacidade como a capacidade informacional, relativa a um único uso do canal, apesar da definição operacional de capacidade clássica depender do comportamento assintótico de múltiplos usos independentes do canal. Isso se deu pela aditividade da capacidade informacional.

Teorema 4.1.11 ([13])**.** *A capacidade clássica de um canal quântico $(A, \mathcal{N}, B)$ é dada por:*

$$C(\mathcal{N}) = I_{reg}(\mathcal{N}) = \lim_{n \to \infty} \frac{1}{n} \max_{\mathcal{E}, \Lambda} I_{\mathcal{N}^{\otimes n}}(\mathcal{E} : \Lambda).$$

Demonstração. Para ver que $C \leq I_{reg}(\mathcal{N})$, considere uma sequência de $(2^{nR}, n)$-códigos $(\mathcal{E}^n, \Lambda^n)$ de taxa $R > 0$. De forma idêntica à demonstração da parte recíproca do Teorema da Codificação de Canais (1.3.2), obtemos:

$$P_e(\mathcal{E}^n, \Lambda^n) \geq 1 - \frac{I(X^n : Y^n)}{nR} - \frac{1}{nR}$$

$$
\begin{aligned}
&= 1 - \frac{I_{\mathcal{N}^{\otimes n}}(\mathcal{E}^n, \Lambda^n)}{nR} - \frac{1}{nR} \\
&\geq 1 - \frac{1}{nR} \max_{\mathcal{E},\Lambda} I_{\mathcal{N}^{\otimes n}}(\mathcal{E}, \Lambda) - \frac{1}{nR},
\end{aligned}
$$

onde $I(X^n : Y^n) = I_{\mathcal{N}^{\otimes n}}(\mathcal{E}^n, \Lambda^n)$ por (4.6). Como $(\mathcal{E}^n, \Lambda^n)$ é arbitrário, podemos substituir $P_e(\mathcal{E}^n, \Lambda^n)$ por $p_e(2^{nR}, n)$ nas desigualdades acima. Tomando $n \to \infty$:

$$
\lim_{n\to\infty} p_e(2^{nR}, n) \geq 1 - \frac{I_{reg}(\mathcal{N})}{R}.
$$

Logo, para $R > I_{reg}(\mathcal{N})$, $\lim_{n\to\infty} p_e(2^{nR}, n) > 0$ e portanto $C(\mathcal{N}) \leq I_{reg}(\mathcal{N})$.

Para ver que $C \geq I_{reg}(\mathcal{N})$, basta mostrar que qualquer taxa $R < I_{reg}(\mathcal{N})$ é atingível. Seja então tal R, e seja n_0 tal que $n_0 R < I_{acc}(\mathcal{N}^{\otimes n_0})$. Tome $\mathcal{Q}$ e Λ^{n_0} tais que $I_{acc}(\mathcal{N}^{\otimes n_0}) = C(N_{\mathcal{Q},\Lambda^{n_0}})$ como na equação (4.5) (a Proposição 4.1.2 garante suas existências). Pelo Teorema da Codificação de Canais temos que $n_0 R$ é uma taxa atingível para $N_{\mathcal{Q},\Lambda^{n_0}}$, logo existe uma sequência de $(2^{n(n_0 R)}, n)$-códigos clássicos para $N_{\mathcal{Q},\Lambda^{n_0}}$ tais que $P_e^n \to 0$. Combinando esses códigos clássicos com a codificação $\mathcal{Q}$ e medição Λ^{n_0}, obtemos $(2^{n(n_0 R)}, n)$-códigos quânticos e concluímos que $n_0 R$ é uma taxa atingível para $\mathcal{N}^{\otimes n_0}$. Esses mesmos códigos quânticos podem ser vistos como códigos de blocos de tamanho nn_0 para $\mathcal{N}$, e portanto $p_e(nn_0, 2^{n(n_0 R)}) \to 0$ para $\mathcal{N}$.

Agora, para qualquer $n' > 0$, podemos obter n suficientemente grande tal que:

$$
nn_0 \leq n' \leq (n+1)n_0.
$$

Então temos, escolhendo R' de forma que $(1 + \frac{1}{n})R \leq R' < I_{acc}(\mathcal{N}^{\otimes n})$:

$$
p_e(2^{n'R}, n') \leq p_e(2^{(n+1)n_0 R}, nn_0) \leq p_e(2^{n(n_0 R')}, nn_0) \to 0,
$$

onde a primeira e segunda desigualdades expressam que aumenta a probabilidade de erro codificando com taxas maiores e blocos menores, e taxas maiores com blocos de mesmo tamanho, respectivamente. □

4.1.1 Esquemas de Codificação

Em um (M, n)-código quântico geral, tanto Alice quanto Bob podem utilizar emaranhamento como recurso. Na Seção 1.4, vimos que correlações clássicas entre entradas de um canal DSM não podem utilizadas para atingir taxas de transmissão maiores. No contexto quântico, temos à disposição não só entradas com correlações clássicas, mas também correlações quânticas, na forma de estados emaranhados. Supondo que um canal quântico preserve emaranhamento entre suas entradas, é possível que Bob detecte tais correlações. Pelo Teorema da Não-Comunicação (Teorema 2.2.18), isso só é possível se a medição de Bob for uma medição conjunta sobre o sistema composto de todas as componentes das entradas.

É então instrutivo estudar os diferentes esquemas de codificação de um canal quântico obtidos pela utilização ou não de emaranhamento, tanto por Alice quanto por Bob. A vantagem desta perspectiva é que quando Alice codifica palavras $x^n \in \mathcal{X}^n$ exclusivamente

por estados produto, tomando, em termos da Definição 4.1.8, $\mathcal{Q}^n = \mathcal{Q}^{\otimes n}$ para algum canal C-Q $(\mathcal{X}, \mathcal{Q}, A)$, então o canal resultante $\mathcal{N}^{\otimes n} \circ \mathcal{Q}^{\otimes n}$ é um canal C-Q. Similarmente, se Bob realiza somente medições locais em cada componente da saída, dadas por $\Lambda^n = \Lambda^{\otimes n}$ para algum canal Q-C $(B, \Lambda, \mathcal{Y})$, então o canal resultante $\Lambda^{\otimes n} \circ \mathcal{N}^{\otimes n}$ é um canal Q-C. Vejamos em mais detalhe os quatro esquemas que resultam da utilização ou não de emaranhamento por Alice ou Bob:

PP: (Estados Produto e Medições Produto) Neste caso, fixamos um canal C-Q $(\mathcal{X}, \mathcal{Q}, A)$ com $\mathcal{Q}(x) = \rho_x^A$ e um canal Q-C $(B, \Lambda, \mathcal{Y})$, e definimos para cada $n > 1$ um (M, n)-código de forma que $\mathcal{Q}^n = (\mathcal{Q})^{\otimes n}$ e $\Lambda^n = (\Lambda)^{\otimes n}$, ou seja, tais que:

$$\mathcal{Q}^n(x^n) = \mathcal{Q}(x_1) \otimes \mathcal{Q}(x_2) \otimes \cdots \otimes \mathcal{Q}(x_n) = \rho_{x_1}^A \otimes \rho_{x_2}^A \otimes \cdots \otimes \rho_{x_n}^A,$$

e

$$\Lambda_{y^n} = \Lambda_{y_1} \otimes \Lambda_{y_2} \otimes \cdots \otimes \Lambda_{y_n}.$$

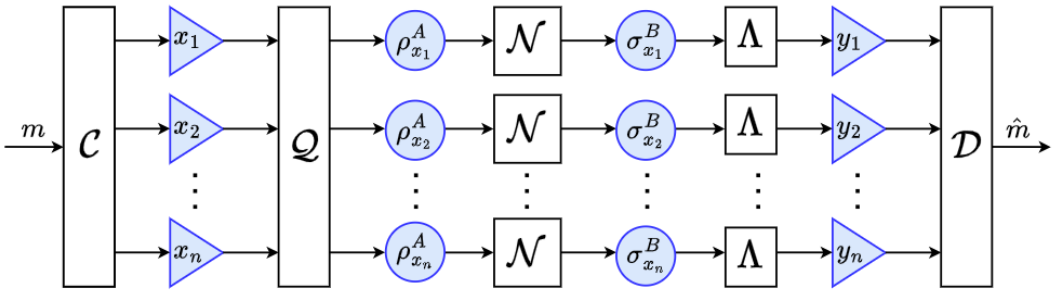

Figura 4.2: Esquema PP: estados produto e medições produto.

EP: (Estados Emaranhados e Medições Produto) Neste caso, fixamos um canal Q-C $(B, \Lambda, \mathcal{Y})$, e definimos para cada $n > 1$ um (M, n)-código onde $\mathcal{Q}^n$ é qualquer e $\Lambda^n = (\Lambda)^{\otimes n}$.

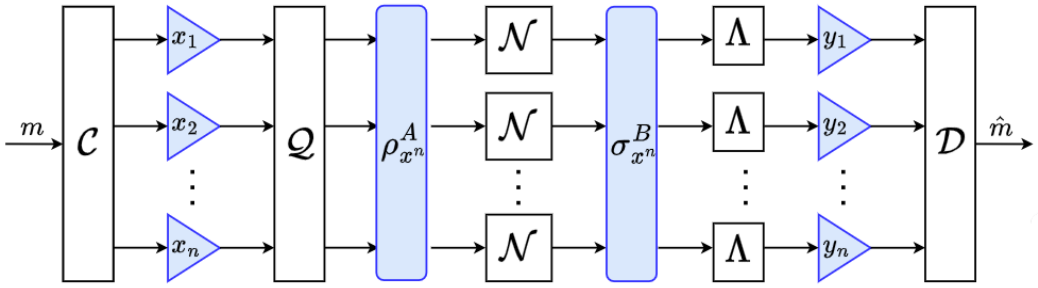

Figura 4.3: Esquema EP: estados emaranhados e medições produto.

PE: (Estados Produto e Medições Conjuntas) Neste caso, fixamos um canal C-Q $(\mathcal{X}, \mathcal{Q}, A)$ com $\mathcal{Q}(x) = \rho_x^A$ e definimos para cada $n > 1$ um (M, n)-código com $\mathcal{Q}^n = (\mathcal{Q})^{\otimes n}$ e Λ qualquer.

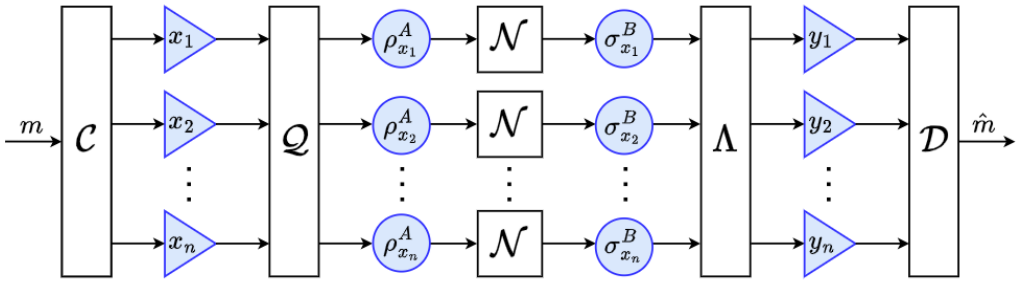

Figura 4.4: Esquema PE: estados produto e medições conjuntas.

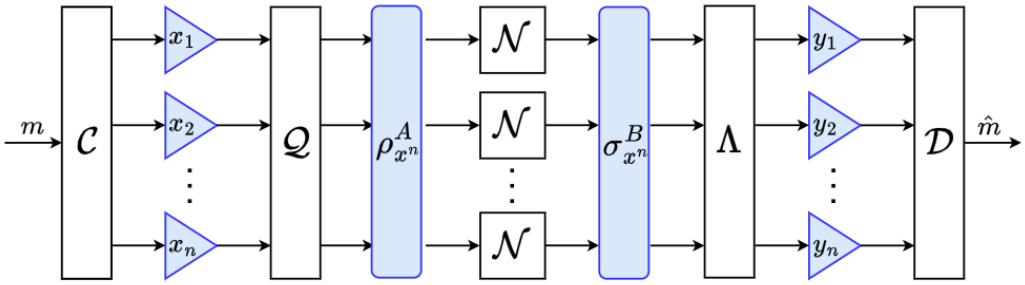

Figura 4.5: Esquema EE: estados emaranhados e medições conjuntas.

EE: (Estados Emaranhados e Medições Conjuntas) Neste caso consideramos (M, n)-códigos sem restrições.

Para cada esquema, temos uma noção adequada de capacidade clássica.

Definição 4.1.12 (Capacidades [14])**.** Para um canal quântico $(A, \mathcal{N}, B)$, definimos:

$$C_{PP}(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n} \max_{\mathcal{E}^{\otimes n}, \Lambda^{\otimes n}} I_{\mathcal{N}^{\otimes n}}(\mathcal{E}^{\otimes n}, \Lambda^{\otimes n}); \tag{4.12}$$

$$C_{EP}(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n} \max_{\mathcal{E}, \Lambda^{\otimes n}} I_{\mathcal{N}^{\otimes n}}(\mathcal{E}, \Lambda^{\otimes n}); \tag{4.13}$$

$$C_{PE}(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n} \max_{\mathcal{E}^{\otimes n}, \Lambda} I_{\mathcal{N}^{\otimes n}}(\mathcal{E}^{\otimes n}, \Lambda); \tag{4.14}$$

$$C_{EE}(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n} \max_{\mathcal{E}, \Lambda} I_{\mathcal{N}^{\otimes n}}(\mathcal{E}, \Lambda). \tag{4.15}$$

Pelo Teorema 4.1.11, sempre temos que:

$$C_{EE}(\mathcal{N}) = C(\mathcal{N}).$$

Teorema 4.1.13. *Para qualquer canal quântico $(A, \mathcal{N}, B)$:*

$$C_{PP}(\mathcal{N}) = I_{acc}(\mathcal{N}).$$

Demonstração. Seja $N_{\mathcal{Q}^{\otimes n}, \Lambda^{\otimes n}}$ o canal como em (4.1), com $\mathcal{Q}^{\otimes n}$ um esquema de codificação em estados produto de A^n e $\Lambda^{\otimes n}$ uma medição produto sobre B^n.

A probabilidade de transição de $N_{\mathcal{Q}^{\otimes n}, \Lambda^{\otimes n}}$ é dada por:

$$p_{Y^n|X^n}(y_1, y_2, \ldots, y_n | x_1(m), x_2(m), \ldots, x_n(m))$$

$$
\begin{aligned}
&= \mathrm{Tr}\left\{(\Lambda_{y_1}\otimes\Lambda_{y_2}\cdots\Lambda_{y_n})(\mathcal{N}(\rho^A_{x_1(m)})\otimes\mathcal{N}(\rho^A_{x_2(m)})\otimes\cdots\mathcal{N}(\rho^A_{x_n(m)}))\right\}\\
&= \prod_{i=1}^{n}\mathrm{Tr}\left\{\Lambda(y_i)\mathcal{N}(\rho^A_{x_i(m)})\right\}\\
&= \prod_{i=1}^{n} p_{Y|X}(y_i|x_i(m)).
\end{aligned}
$$

Ou seja, $N_{\mathcal{Q}^{\otimes n},\Lambda^{\otimes n}} = N^{\times n}_{\mathcal{Q},\Lambda}$. Pela aditividade da capacidade de canais clássicos (Seção 1.4) temos que $C(N_{\mathcal{Q}^{\otimes n},\Lambda^{\otimes n}}) = nC(N_{\mathcal{Q},\Lambda})$. Isso implica que:

$$\sup_{\mathcal{E}^{\otimes n},\Lambda^{\otimes n}} I_{\mathcal{N}^{\otimes n}}(\mathcal{E}^{\otimes n},\Lambda^{\otimes n}) = n\sup_{\mathcal{E},\Lambda} I_{\mathcal{N}}(\mathcal{E},\Lambda) = nI_{acc}(\mathcal{N}).$$

Substituindo em (4.12), obtemos o resultado.

□

Teorema 4.1.14. *Para qualquer canal quântico $(A,\mathcal{N},B)$:*

$$C_{EP}(\mathcal{N}) = C_{PP}(\mathcal{N}) = I_{acc}(\mathcal{N}). \tag{4.16}$$

Demonstração. A demonstração que segue é adaptada de [13] (ver [14] para uma demonstração alternativa). Note que sempre vale a desigualdade:

$$C_{PP}(\mathcal{N}) \leq C_{EP}(\mathcal{N}),$$

já que os ensembles produto $\mathcal{E}^{\otimes n}$ em (4.12) são casos particulares de ensembles $\mathcal{E}$ em (4.13). Vamos mostrar a desigualdade reversa para estabelecer a igualdade. Para isso, basta mostrar que, para dois canais $(A_1,\mathcal{N}_1,B_1)$ e $(A_2,\mathcal{N}_2,B_2)$ quaisquer:

$$\max_{\mathcal{E},\Lambda^1\otimes\Lambda^2} I_{\mathcal{N}_1\otimes\mathcal{N}_2}(\mathcal{E},\Lambda^1\otimes\Lambda^2) \leq \max_{\mathcal{E}^1,\Lambda^1} I_{\mathcal{N}_1}(\mathcal{E}^1,\Lambda^1) + \max_{\mathcal{E}^2,\Lambda^2} I_{\mathcal{N}_2}(\mathcal{E}^2,\Lambda^2), \tag{4.17}$$

já que por superaditividade da informação acessível (4.1.3), a desigualdade acima implica em particular que $C_{EP}(\mathcal{N}) \leq C_{PP}(\mathcal{N})$.

Para obter (4.17), considere um ensemble $\mathcal{E} = \{p_X,\rho^{A_1A_2}_x\}_{x\in\mathcal{X}}$ como entrada de $\mathcal{N}_1\otimes\mathcal{N}_2$ e uma medição produto $\Lambda^1\otimes\Lambda^2$ sobre o sistema B_1B_2. O canal clássico resultante tem probabilidade de transição (como em (4.1)) dada por:

$$p_{Y_1,Y_2|X}(y_1,y_2|x) = \mathrm{Tr}\left\{\rho^{A_1A_2}_x(\mathcal{N}_1\otimes\mathcal{N}_2)^*(\Lambda^1_{y_1}\otimes\Lambda^2_{y_2})\right\} = \mathrm{Tr}\left\{\rho^{A_1A_2}_x\mathcal{N}_1^*(\Lambda^1_{y_1})\otimes\mathcal{N}_2^*(\Lambda^2_{y_2})\right\},$$

que podemos expressar como:

$$p_{Y_1,Y_2|X}(y_1,y_2|x) = p_{Y_2|X}(y_2|x)p_{Y_1|,Y_2,X}(y_1|y_2,x),$$

onde

$$p_{Y_2|X}(y_2|x) = \mathrm{Tr}\left\{\rho^{A_2}_x\mathcal{N}_2^*(\Lambda^2_{y_2})\right\} \qquad p_{Y_1|,Y_2,X}(y_1|y_2,x) = \mathrm{Tr}\left\{\rho^{A_1}_{y_2,x}\mathcal{N}_1^*(\Lambda^1_{y_1})\right\},$$

com

$$\rho^{A_2}_x = \mathrm{Tr}_{A_1}\left\{\rho^{A_1A_2}_x\right\} \qquad \rho^{A_1}_{y_2,x} = \frac{\mathrm{Tr}_{A_2}\left\{\rho^{A_1A_2}_x(\mathbb{1}_{A_1}\otimes\mathcal{N}_2^*(\Lambda^2_{y_2}))\right\}}{\mathrm{Tr}\left\{\rho^{A_1A_2}_x(\mathbb{1}_{A_1}\otimes\mathcal{N}_2^*(\Lambda^2_{y_2}))\right\}}.$$

Tomando $\mathcal{E}^1 = \{p_{X_1}(y_2, x), \rho^{A_1}_{y_2,x}\}$ e $\mathcal{E}^2 = \{p_{X_2}(x), \rho^{A_2}_x\}$, temos, pela definição de informação mútua clássica:

$$I(Y_1, Y_2 : X) = H(Y_1, Y_2) - H(Y_1, Y_2|X)$$

(e pela subaditividade da entropia e regra da cadeia para entropia condicional)

$$\begin{aligned} &\leq H(Y_1) - H(Y_1|Y_2, X) + H(Y_2) - H(Y_2|X) \\ &= I(Y_1 : Y_2, X) + I(Y_2 : X) \\ &= I_{\mathcal{N}_1}(\mathcal{E}^1, \Lambda^1) + I_{\mathcal{N}_2}(\mathcal{E}^2, \Lambda^2). \end{aligned}$$

Tomando o máximo entre os ensembles, obtemos (4.17). □

Note que esse resultado era esperado, já que são necessárias medições conjuntas para detectar emaranhamento.

4.1.2 Capacidade Clássica de Canais C-Q

Os esquemas de codificação com estados produto (PP e PE) resultam em um canal C-Q $(\mathcal{X}^n, \mathcal{N}^{\otimes n} \circ \mathcal{Q}^{\otimes n} \equiv \sigma^{B^n}_{x^n}, B^n)$, logo para esses casos podemos nos restringir à análise de canais C-Q. Já vimos que a capacidade resultante do esquema PP é a informação acessível do canal. Já o esquema PE requer um estudo mais cuidadoso.

Teorema 4.1.15 (Capacidade de Holevo de Canais C-Q)**.** *A Capacidade de Holevo de um canal C-Q $(\mathcal{X}, \mathcal{N} \equiv \sigma^B_x, B)$ é dado por:*

$$\max_{p_X} \mathcal{I}(X : B)_{\rho^{XB}} = \max_{p_X} \left\{ S\left(\sum_{x\in\mathcal{X}} p_X(x)\sigma^B_x \right) - \sum_{x\in\mathcal{X}} p_X(x) S(\sigma^B_x) \right\}, \tag{4.18}$$

onde

$$\rho^{XB} = \sum_{x\in\mathcal{X}} p_X(x) \left|x\right\rangle\!\left\langle x\right| \otimes \sigma^B_x.$$

Ou seja, podemos nos restringir a otimizar sob os ensembles do tipo $\mathcal{E} = \{p_X(x), |x\rangle\langle x|\}_{x\in\mathcal{X}}$ em (4.9).

Demonstração. Seja $\{q(i), \rho_i\}_{i\in\Sigma}$ um ensemble em $\mathbb{C}^{\mathcal{X}}$ tal que $\chi(\mathcal{N}) = \chi(\{q(i), \mathcal{N}(\rho_i)\})$ (que existe pela Proposição 4.1.6). Considere as decomposições espectrais de ρ_i:

$$\rho_i = \sum_{j=1}^{|\mathcal{X}|} \alpha_{ij} \left|\phi_{ij}\right\rangle\!\left\langle\phi_{ij}\right|.$$

Então, como $\mathcal{N}$ é canal C-Q:

$$\begin{aligned} \mathcal{N}(\rho_i) &= \sum_{x\in\mathcal{X}} \langle x|\rho_i|x\rangle\, \sigma^B_x \\ &= \sum_{x\in\mathcal{X}} \sum_{j=1}^{|\mathcal{X}|} \alpha_{ij} \langle x|\phi_{ij}\rangle \langle\phi_{ij}|x\rangle\, \sigma^B_x \\ &= \sum_{j=1}^{|\mathcal{X}|} \sum_{x\in\mathcal{X}} \alpha_{ij} \langle x|\phi_{ij}\rangle^2\, \sigma^B_x. \end{aligned}$$

Note que:

$$p(x) = \sum_{i\in\Sigma}\sum_{j=1}^{|\mathcal{X}|} q(i)\alpha_{ij} \langle x|\phi_{ij}\rangle^2$$

é uma distribuição de probabilidade em $\mathcal{X}$. Temos:

$$\begin{aligned} S\left(\sum_{i\in\Sigma} q(i)\mathcal{N}(\rho_i)\right) &= S\left(\sum_{i\in\Sigma} q(i)\left(\sum_{j=1}^{|\mathcal{X}|}\sum_{x\in\mathcal{X}} \alpha_{ij}\langle x|\phi_{ij}\rangle^2 \sigma_x^B\right)\right) \\ &= S\left(\sum_{x\in\mathcal{X}}\left(\sum_{i\in\Sigma}\sum_{j=1}^{|\mathcal{X}|} q(i)\alpha_{ij}\langle x|\phi_{ij}\rangle^2\right)\sigma_x^B\right) \\ &= S\left(\sum_{x\in\mathcal{X}} p(x)\sigma_x^B\right). \end{aligned} \tag{4.19}$$

Por concavidade da entropia quântica:

$$\begin{aligned} -\sum_{i\in\Sigma} q(i)S(\mathcal{N}(\rho_i)) &= -\sum_{i\in\Sigma} q(i) S\left(\sum_{j=1}^{|\mathcal{X}|}\sum_{x\in\mathcal{X}} \alpha_{ij}\langle x|\phi_{ij}\rangle^2 \sigma_x^B\right) \\ &\leq -\sum_{i\in\Sigma} q(i)\sum_{j=1}^{|\mathcal{X}|}\sum_{x\in\mathcal{X}} \alpha_{ij}\langle x|\phi_{ij}\rangle^2 S\left(\sigma_x^B\right) \\ &= -\sum_{x\in\mathcal{X}} p(x)S(\sigma_x^B). \end{aligned} \tag{4.20}$$

Somando (4.19) e (4.20):

$$\chi(\{q(i),\rho_i\}) \leq S\left(\sum_{x\in\mathcal{X}} p(x)\sigma_x^B\right) - \sum_{x\in\mathcal{X}} p(x)S(\sigma_x^B) = \chi(\{p(x),\sigma_x^B\}).$$

Mas escolhemos $\{q(i),\rho_i\}$ como um ensemble que maximiza a quantidade de Holevo, logo $\chi(\{q(i),\rho_i\}) = \chi(\{p(x),\sigma_x^B\} = \chi(\mathcal{N})$. A expressão (4.18) segue então de (3.15). □

Finalmente, podemos enunciar o análogo quântico do Teorema da Codificação de Canais:

Teorema 4.1.16 (Teorema de Holevo-Schumacher-Westmoreland (HSW))**.** *Para todo canal C-Q* $(\mathcal{X},\mathcal{N}\equiv\sigma_x^B,B)$*:*

$$C(\mathcal{N}) = \chi(\mathcal{N})$$

Note que, pelo Teorema 4.1.15, a capacidade de Holevo é, assim como a capacidade informacional de um canal clássico, o supremo de uma informação mútua. A demonstração do Teorema de HSW, à qual vamos dedicar o próximo capítulo, é similar ao Teorema da Codificação de Canais. Vejamos algumas de suas aplicações importantes.

Teorema 4.1.17. *Para qualquer canal quântico* $(A,\mathcal{N},B)$*:*

$$C_{PE}(\mathcal{N}) = \chi(\mathcal{N}) \tag{4.21}$$

Demonstração. Seja $(\mathcal{X}, \mathcal{Q}, A)$ o canal C-Q correspondente ao esquema PE e sejam $\tilde{\mathcal{N}}^{\otimes n} = \mathcal{N}^{\otimes n} \circ \mathcal{Q}^{\otimes n}$, também canais C-Q. Note que, para cada $n \geq 1$:

$$\max_{\mathcal{E}^{\otimes n}, \Lambda} I_{\mathcal{N}^{\otimes n}}(\mathcal{E}^{\otimes n}, \Lambda) = \max_{\mathcal{E}, \Lambda} I_{\tilde{\mathcal{N}}^{\otimes n}}(\mathcal{E} : \Lambda),$$

já que, para qualquer estado $\rho_{x^n}^{A^n}$ de um ensemble $\mathcal{E}$ de A^n, $\tilde{\mathcal{N}}^{\otimes n}(\rho_{x^n}^{A^n})$ é um estado produto.

Por (4.14) e pelo Teorema 4.1.11:

$$C_{PE}(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n} \max_{\mathcal{E}^{\otimes n}, \Lambda} I_{\mathcal{N}^{\otimes n}}(\mathcal{E}^{\otimes n}, \Lambda) = \lim_{n\to\infty} \frac{1}{n} \max_{\mathcal{E}, \Lambda} I_{\tilde{\mathcal{N}}^{\otimes n}}(\mathcal{E} : \Lambda) = C(\tilde{\mathcal{N}}) = \chi(\tilde{\mathcal{N}}).$$

Basta notar agora que $\chi(\tilde{\mathcal{N}}) = \chi(\mathcal{N})$. Para isso, seja $\tilde{\mathcal{E}} = \{p_X(x), |x\rangle\langle x|\}_{x\in\mathcal{X}}$ tal que $\chi(\tilde{\mathcal{N}}) = \chi(\{p_X(x), \tilde{\mathcal{N}}(|x\rangle\langle x|)\})$. Então:

$$\chi(\tilde{\mathcal{N}}) = S\left(\sum_{x\in\mathcal{X}} p_X(x)\mathcal{N} \circ \mathcal{Q}(|x\rangle\langle x|)\right) - \sum_{x\in\mathcal{X}} p_X(x) S(\mathcal{N} \circ \mathcal{Q}(|x\rangle\langle x|)) \leq \chi(\mathcal{N}).$$

Por outro lado, se $\mathcal{E} = \{p_X(x), \rho_x^A\}_{x\in\mathcal{X}}$ é tal que $\chi(\mathcal{N}) = \chi(\{p_X(x), \mathcal{N}(\rho_x^A)\})$, então definindo $\mathcal{Q}$ por $\mathcal{Q}(x) = \rho_x^A$ temos:

$$\chi(\mathcal{N}) = \chi(\{p_X(x), \mathcal{N} \circ \mathcal{Q}(|x\rangle\langle x|)\}) \leq \chi(\tilde{\mathcal{N}}).$$

□

Note que, para um canal C-Q $(\mathcal{X}, \mathcal{N} \equiv \sigma_x^B, B)$ e $p_X \in \mathcal{P}(\mathcal{X})$, denotando por σ o estado misto correspondente ao ensemble $\{p_X(x), \sigma_x^B\}_{x\in\mathcal{X}}$:

$$\sigma = \sum_{x\in\mathcal{X}} p_X(x)\sigma_x^B,$$

então $S(\sigma)$ corresponde ao primeiro termo em $\chi(\{p_X(x), \sigma_x^B\})$, logo:

$$C(\mathcal{N}) \leq \chi(\{p_X(x), \sigma_x^B\}) \leq S(\sigma) \leq \log d_B.$$

Proposição 4.1.18 ([13]). *Para um canal C-Q , $C(\mathcal{N}) = I_{acc}(\mathcal{N})$ se, e somente se os operadores $p_X(x)\sigma_x^B$ comutam, com $p \in \mathcal{P}(\mathcal{X})$ tal que $\chi(\mathcal{N}) = \chi(\{p(x), \sigma_x^B\})$.*

4.1.3 Capacidade Clássica de Canais Quânticos

Como corolário do Teorema de HSW, obtemos o seguinte resultado, também referido na literatura como Teorema de Holevo-Schumacher-Westmoreland, que nos dá a capacidade clássica de um canal quântico qualquer em termos da capacidade de Holevo de suas extensões sem memória.

Teorema 4.1.19. *Para todo canal quântico $(A, \mathcal{N}, B)$:*

$$C(\mathcal{N}) = \chi_{reg}(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n}\chi(\mathcal{N}^{\otimes n}). \tag{4.22}$$

Em particular, se para todo $n > 0$, $\chi(\mathcal{N}^{\otimes n}) = n\chi(\mathcal{N})$, então $C(\mathcal{N}) = \chi(\mathcal{N})$.

Demonstração. Que $C(\mathcal{N}) \leq \chi_{reg}(\mathcal{N})$ segue do Teorema 4.1.11 e do Teorema de Holevo 3.2.27. Para ver que $\chi_{reg}(\mathcal{N}) \leq C(\mathcal{N})$, seja $R < \chi_{reg}(\mathcal{N})$, $n_0 > 0$ e um ensemble $\mathcal{E}^{n_0} = \{p_{n_0}(x), \rho_x^{n_0}\}_{x\in\mathcal{X}_0}$ para algum $\mathcal{X}_0$ tais que:

$$n_0 R < \chi(\{p_{n_0}(x), \mathcal{N}^{\otimes n_0}(\rho_x^{n_0})\}_{x\in\mathcal{X}_0}).$$

Considere então o canal C-Q $(\mathcal{X}_0, \tilde{\mathcal{N}} \equiv \mathcal{N}^{\otimes n_0}(\rho_x^{n_0}), B^{n_0})$. Pelo Teorema de HSW:

$$n_0 R < \chi(\{p_{n_0}(x), \mathcal{N}^{\otimes n_0}(\rho_x^{(n_0)})\}_{x\in\mathcal{X}_0}) \leq C(\mathcal{N}^{\otimes n_0}),$$

ou seja, $n_0 R$ é uma taxa atingível para $\tilde{\mathcal{N}}$. Denotando por $\tilde{p}_e(M, n)$ a probabilidade mínima de erro de $\tilde{\mathcal{N}}$, temos:

$$p_e(2^{(nn_0)R}, nn_0) \leq \tilde{p}_e(2^{n(n_0 R)}, n) \to 0,$$

e similarmente à demonstração do Teorema 4.1.11, podemos concluir que R é uma taxa atingível para $\mathcal{N}$. □

Note que, apesar de sempre termos $I_{acc}(\mathcal{N}) \leq \chi(\mathcal{N})$ (4.10), temos também que $I_{reg}(\mathcal{N}) = \chi_{reg}(\mathcal{N}) = C(\mathcal{N})$. Eis um exemplo de cálculo da capacidade clássica utilizando o Teorema 4.1.19:

Teorema 4.1.20 (Canais Quânticos de Capacidade Zero)**.** *Dado um canal quântico $(A, \mathcal{N}, B)$, $C(\mathcal{N}) = 0$ se e somente se $\mathcal{N}$ é constante, ou seja, existe $\sigma^B \in \mathcal{D}(\mathcal{H}_B)$ tal que $\mathcal{N} = \sigma^B \mathrm{Tr}\{\cdot\}$.*

Demonstração. Se $\mathcal{N}$ é constante, então para qualquer ensemble $\mathcal{E} = \{p_X(x), \rho_x^A\}_{x\in\mathcal{X}}$ de A:

$$\chi(\mathcal{E}) = S\left(\sum_{x\in\mathcal{X}} p_X(x)\sigma^B \mathrm{Tr}\left\{\rho_x^A\right\}\right) - \sum_{x\in\mathcal{X}} p_X(x) S(\sigma^B \mathrm{Tr}\left\{\rho_x^A\right\}) = 0,$$

logo $\chi(\mathcal{N}) = 0$. É fácil ver que $\chi(\mathcal{N}^{\otimes n})$ é constante para todo $n > 0$. Segue do Teorema 4.1.19 que $C(\mathcal{N}) = 0$. Suponha agora que $\mathcal{N}$ não é constante, ou seja que existem $\rho_1^A, \rho_2^A \in \mathcal{D}(\mathcal{H}_A)$ tais que $\mathcal{N}(\rho_1^A) \neq \mathcal{N}(\rho_2^A)$. Considere um estado clássico-quântico:

$$\rho^{XA} = \frac{1}{2}|1\rangle\langle 1| \otimes \rho_1^A + \frac{1}{2}|2\rangle\langle 2| \otimes \rho_2^A,$$

onde $\mathcal{X} = \{1, 2\}$. Note que:

$$\chi(\{\frac{1}{2}, \rho_x^A\}_{x\in\mathcal{X}}) = \mathcal{I}(X:A)_\rho = D((\mathbb{1}_X \otimes \mathcal{N})\rho^{XA} || \rho^X \otimes \rho^A) = 0$$

se e somente se $(\mathbb{1}_X \otimes \mathcal{N})\rho^{XA} = \rho^X \otimes \rho^A$, ou seja:

$$\frac{1}{2}|1\rangle\langle 1| \otimes \mathcal{N}(\rho_1^A) + \frac{1}{2}|2\rangle\langle 2| \otimes \mathcal{N}(\rho_2^A) = \frac{\mathbb{1}_X}{2} \otimes \left(\frac{1}{2}\mathcal{N}(\rho_1^A) + \frac{1}{2}\mathcal{N}(\rho_2^A)\right)$$

se e somente se $\mathcal{N}(\rho_1^A) = \mathcal{N}(\rho_2^A)$. Logo temos que $\chi(\{\frac{1}{2}, \rho_x^A\}_{x\in\mathcal{X}}) > 0$ e portanto $C(\mathcal{N}) \geq \chi(\mathcal{N}) > 0$. □

4.2 A Conjectura da Aditividade

Vimos no Capítulo 1 que o Teorema da Codificação de Canais fornece uma expressão para a capacidade de qualquer canal clássico dependendo somente de um uso do canal (a capacidade informacional), e que isso advém do fato de que a capacidade de canais clássicos é sempre aditiva. Analogamente, o Teorema 4.1.19 mostra que a capacidade clássica de um canal quântico é dada pela capacidade de Holevo, que também depende apenas de um uso do canal, caso sua capacidade de Holevo satisfaça uma hipótese de aditividade.

Se a capacidade de Holevo é ou não aditiva não é uma questão fácil, já que, como vimos na Seção 1.4, uma demonstração da aditividade da capacidade de canais clássicos utiliza propriedades que não tem análogo na teoria quântica. Essa questão foi considerada primeiro implicitamente em [59], e formulada como uma conjectura em [60]. Uma versão um pouco mais geral da conjectura que se tornou popular é a seguinte:

Conjectura da Aditividade: Para quaisquer canais quânticos $\mathcal{N}_1$ e $\mathcal{N}_2$:

$$\chi(\mathcal{N}_1 \otimes \mathcal{N}_2) = \chi(\mathcal{N}_1) + \chi(\mathcal{N}_2).$$

Em particular, se a conjectura é verdadeira, então para qualquer canal quântico $\mathcal{N}$, temos que $C(\mathcal{N}) = \chi_{reg}(\mathcal{N}) = \chi(\mathcal{N})$ e portanto, que $C_{EE}(\mathcal{N}) = C_{PE}(\mathcal{N})$.

Na direção de elucidar a conjectura, se demostrou que a conjectura é satisfeita se restrita a diversas classes de canais quânticos. Por exemplo, canais ideais [20], canais depolarizadores [61] e canais unitários de qubits [62] satisfazem a conjectura. Veremos a seguir uma demonstração, por [15], de que canais EB também satisfazem a conjectura.

A conjectura da aditividade permaneceu aberta até 2009, quando Hastings [8] demonstrou, a partir de uma versão quântica do método de codificação aleatória, a existência de um contraexemplo, que vamos discutir brevemente.

4.2.1 Capacidade Clássica de Canais EB

Vejamos um importante exemplo de classe de canais quânticos para os quais a capacidade de Holevo é aditiva, e portanto, pelo Teorema 4.1.19, cujas capacidades clássicas coincidem com suas capacidades de Holevo. Note que a demonstração do Teorema 4.2.2 a seguir não utiliza o Teorema de HSW (Teorema 4.1.16).

Lema 4.2.1. *Para $\rho^{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ e um canal quântico $(B, \mathcal{N}, C)$:*

$$\mathrm{Tr}_A\left\{(\mathrm{id}_A \otimes \mathcal{N})(\rho^{AB})\right\} = \mathcal{N}(\mathrm{Tr}_A\left\{\rho^{AB}\right\}) = \mathcal{N}(\rho^B).$$

Demonstração. Basta comparar as expressões (2.9) e (2.16). □

Teorema 4.2.2 (Aditividade dos Canais EB [15])**.** *Para $(A, \mathcal{M}, C)$ um canal quântico qualquer e $(B, \mathcal{N}^{EB}, D)$ um canal EB:*

$$\chi(\mathcal{M} \otimes \mathcal{N}^{EB}) = \chi(\mathcal{M}) + \chi(\mathcal{N}^{EB}).$$

Demonstração. Pelo Teorema 4.1.7, basta mostrar que $\chi(\mathcal{M}\otimes\mathcal{N}^{EB}) \leq \chi(\mathcal{M})+\chi(\mathcal{N}^{EB})$. Para isso, considere $\{p(x),\rho_x^{AB}\}_{x\in\mathcal{X}}$ tal que $\chi(\mathcal{M}\otimes\mathcal{N}^{EB}) = \chi(\{p(x),(\mathcal{M}\otimes\mathcal{N}^{EB})(\rho_x^{AB})\})$, que existe pela Proposição 4.1.6. Como $\mathcal{N}^{EB}$ é canal EB, para cada $x \in \mathcal{X}$, $(\mathrm{id}_A \otimes \mathcal{N}^{EB})(\rho_x^{AB})$ é separável, logo pela Proposição 2.2.9, existe $\mathcal{Y}$ finito e $q_x \in \mathcal{P}(\mathcal{Y})$ tais que:

$$(\mathrm{id}_A \otimes \mathcal{N}^{EB})(\rho_x^{AB}) = \sum_{y\in\mathcal{Y}} q_x(y)\,|a_{xy}\rangle\langle a_{xy}| \otimes |d_{xy}\rangle\langle d_{xy}|\,,$$

com $|a_{xy}\rangle \in \mathcal{H}_A$ e $|d_{xy}\rangle \in \mathcal{H}_D$. Seja $\sigma_x^{YCD} \in \mathcal{D}(\mathbb{C}^{\mathcal{Y}} \otimes \mathcal{H}_C \otimes \mathcal{H}_D)$ dado por:

$$\sigma_x^{YCD} = \sum_{y\in\mathcal{Y}} q_x(y)\,|y\rangle\langle y| \otimes \mathcal{M}(|a_{xy}\rangle\langle a_{xy}|) \otimes |d_{xy}\rangle\langle d_{xy}|\,.$$

Pela subaditividade forte da entropia quântica (Teorema 3.2.21):

$$S(\sigma_x^{CD}) \geq S(\sigma_x^{YCD}) - S(\sigma_x^{YD}) + S(\sigma_x^{D}), \tag{4.23}$$

onde

$$\sigma_x^{CD} = \sum_{y\in\mathcal{Y}} q_x(y)\mathcal{M}(|a_{xy}\rangle\langle a_{xy}|) \otimes |d_{xy}\rangle\langle d_{xy}| = (\mathcal{M}\otimes\mathcal{N}^{EB})(\rho_x^{AB}). \tag{4.24}$$

Como σ_x^{YCD} é um estado clássico-quântico, pelo Teorema 3.2.8 e pela aditividade da entropia quântica para estados produto (Teorema 3.2.14):

$$\begin{aligned} S(\sigma_x^{YCD}) &= H(q_x) + \sum_{y\in\mathcal{Y}} q_x(y) S(\mathcal{M}(|a_{xy}\rangle\langle a_{xy}|) \otimes |d_{xy}\rangle\langle d_{xy}|) \\ &= H(q_x) + \sum_{y\in\mathcal{Y}} q_x(y) S(\mathcal{M}(|a_{xy}\rangle\langle a_{xy}|)) + \sum_{y\in\mathcal{Y}} q_x(y) S(|d_{xy}\rangle\langle d_{xy}|) \\ &= H(q_x) + \sum_{y\in\mathcal{Y}} q_x(y) S(\mathcal{M}(|a_{xy}\rangle\langle a_{xy}|)), \end{aligned} \tag{4.25}$$

onde usamos que $S(|d_{xy}\rangle\langle d_{xy}|) = 0$, pelo Teorema 3.2.6. Similarmente:

$$S(\sigma_x^{YD}) = H(q_x) + \sum_{y\in\mathcal{Y}} q_x(y) S(|d_{xy}\rangle\langle d_{xy}|) = H(q_x). \tag{4.26}$$

Notando que:

$$\sigma_x^{C} = \sum_{y\in\mathcal{Y}} q_x(y)\mathcal{M}(|a_{xy}\rangle\langle a_{xy}|) \tag{4.27}$$

e, pelo Lema 4.2.1:

$$\sigma_x^{D} = \sum_{y\in\mathcal{Y}} q_x(y)\,|d_{xy}\rangle\langle d_{xy}| = \mathrm{Tr}_A\left\{(\mathrm{id}_A \otimes \mathcal{N}^{EB})(\rho_x^{AB})\right\} = \mathcal{N}^{EB}(\rho_x^{B}). \tag{4.28}$$

Substituindo (4.25), (4.26) e (4.28) em (4.23):

$$S((\mathcal{M}\otimes\mathcal{N}^{EB})(\rho_x^{AB})) \geq \sum_{y\in\mathcal{Y}} q_x(y) S(\mathcal{M}(|a_{xy}\rangle\langle a_{xy}|) + S(\mathcal{N}^{EB}(\rho_x^{B}))$$

e somando as desigualdades para cada $x \in \mathcal{X}$, obtemos:

$$\sum_{x\in\mathcal{X}} p(x)S((\mathcal{M}\otimes\mathcal{N}^{EB})(\rho_x^{AB})) \geq \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x)q_x(y)S(\mathcal{M}(|a_{xy}\rangle\langle a_{xy}|)+\sum_{x\in\mathcal{X}} p(x)S(\mathcal{N}^{EB}(\rho_x^B)).$$

Finalmente, pela subaditividade da entropia quântica e pela desigualdade acima:

$$\begin{aligned}\chi(\{p(x),(\mathcal{M}\otimes\mathcal{N}^{EB})(\rho_x^{AB})\}) &= S\left(\sum_{x\in\mathcal{X}} p(x)(\mathcal{M}\otimes\mathcal{N}^{EB})(\rho_x^{AB})\right) - \sum_{x\in\mathcal{X}} p(x)S((\mathcal{M}\otimes\mathcal{N}^{EB})(\rho_x^{AB}))\\ &\leq S(\rho^C)+S(\rho^D)-\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x)q_x(y)S(\mathcal{M}(|a_{xy}\rangle\langle a_{xy}|))\\ &\quad -\sum_{x\in\mathcal{X}} p(x)S(\mathcal{N}^{EB}(\rho_x^B)),\end{aligned}$$

onde, por (4.24):

$$\rho^{CD}=\sum_{x\in\mathcal{X}} p(x)(\mathcal{M}\otimes\mathcal{N}^{EB})(\rho_x^{AB})=\sum_{x\in\mathcal{X}} p(x)\sigma_x^{CD}.$$

Logo, por (4.27) e (4.28):

$$\rho^C=\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x)q_x(y)\mathcal{M}(|a_{xy}\rangle\langle a_{xy}|),\quad \rho^D=\sum_{x\in\mathcal{X}} p(x)\sigma_x^D=\sum_{x\in\mathcal{X}} p(x)\mathcal{N}^{EB}(\rho_x^B)$$

concluímos que:

$$\chi(\mathcal{M}\otimes\mathcal{N}^{EB})=\chi(\{p(x),(\mathcal{M}\otimes\mathcal{N}^{EB})(\rho_x^{AB})\})\leq\chi(\mathcal{M})+\chi(\mathcal{N}^{EB}).$$

□

4.2.2 Conjecturas Equivalentes

No esforço de melhor entender a aditividade (ou não) da capacidade de Holevo, diversas outras quantidades relacionadas foram consideradas, com suas próprias conjecturas de aditividade, que por fim se mostraram equivalentes à conjectura da aditividade e ofereceram novos métodos para demonstrar a aditividade para certas classes de canais, assim como potencialmente obter contraexemplos. Vejamos algumas dessas conjecturas equivalentes.

Definição 4.2.3 (Entropia de Saída Mínima)**.** Dado um canal quântico $(A,\mathcal{N},B)$, sua **entropia de saída mínima** é a quantidade $S_{min}(\mathcal{N})$ dada por:

$$S_{min}(\mathcal{N})=\min_{\rho\in\mathcal{D}(\mathcal{H}_A)} S(\mathcal{N}(\rho)). \tag{4.29}$$

Podemos pensar em $S_{min}(\mathcal{N})$ como uma medida do quão ruidoso é o canal $\mathcal{N}$.

Pela concavidade da entropia quântica, em (4.29) basta minimizar entre os estados puros de A. Em particular, para canais C-C, a entropia quântica em (4.29) é a entropia de Shannon, e neste caso a entropia de saída mínima é aditiva, pois pela Proposição 1.1.4, um estado puro de um sistema clássico composto é um estado produto. Já para sistemas quânticos compostos AB, a existência de estados puros emaranhados implica em:

$$\mathrm{Ext}(\mathcal{D}(\mathcal{H}_A\otimes\mathcal{H}_B))\supsetneq\mathrm{Ext}(\mathcal{D}(\mathcal{H}_A))\times\mathrm{Ext}(\mathcal{D}(\mathcal{H}_B)),$$

e portanto a aditividade da entropia de saída mínima não segue diretamente.

O exemplo a seguir sugere que a aditividade da entropia de saída mínima está relacionada à aditividade da capacidade de Holevo.

Exemplo 4.2.4 ([61])**.** Lembre que no Exemplo 2.3.5 definimos o canal depolarizador $D_p : \mathcal{H}_A \to \mathcal{H}_A$, para $p \in [0,1]$, dado por

$$D_p(\rho) = (1-p)\frac{\mathbb{1}_A}{d_A}\mathrm{Tr}\{\rho\} + p\rho.$$

A capacidade de Holevo de D_p é dada por:

$$\chi(D_p) = \log d_A - S_{min}(D_p).$$

Note que, dado outro canal depolarizador $D_q : \mathcal{H}_B \to \mathcal{H}_B$, se a entropia de saída mínima é aditiva, então:

$$\begin{aligned}\chi(D_p \otimes D_q) &= \log(d_A \cdot d_B) - S_{min}(D_p \otimes D_q) \\ &= \log d_A + \log d_B - S_{min}(D_p) - S_{min}(D_q) \\ &= \chi(D_p) + \chi(D_q).\end{aligned}$$

De fato, para canais depolarizadores, a entropia de saída mínima, e portanto a capacidade de Holevo, são aditivas.

No geral, um contraexemplo da aditividade da entropia de saída mínima é um contraexemplo da aditividade da capacidade de Holevo:

Proposição 4.2.5 ([34])**.** *Para quaisquer dois canais quânticos $\mathcal{N}_1$ e $\mathcal{N}_2$, se*

$$S_{min}(\mathcal{N}_1 \otimes \mathcal{N}_2) < S_{min}(\mathcal{N}_1) + S_{min}(\mathcal{N}_2),$$

então:

$$\chi(\mathcal{N}_1 \otimes \mathcal{N}_2) > \chi(\mathcal{N}_1) + \chi(\mathcal{N}_2).$$

Intuitivamente, se o uso emaranhado dos canais $\mathcal{N}_1$ e $\mathcal{N}_2$ é menos ruidoso que o uso independente deles, então é possível transmitir fielmente mais informação pelos canais emaranhados do que por usos independentes deles.

Outra quantidade relacionada à capacidade de Holevo é uma medida de emaranhamento. Lembrando que, pela Proposição 2.2.6, se σ^{AB} é um estado puro de um sistema quântico bipartite AB:

$$S(\sigma^A) = S(\sigma^B).$$

Essa quantidade serve como definição para quantidade de emaranhamento presente em um estado puro [63]. Uma extensão adequada para estados mistos é dada pela definição a seguir [64].

Definição 4.2.6 (Emaranhamento de Formação)**.** Dado um estado $\sigma^{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$, o **emaranhamento de formação** de σ^{AB} é a quantidade:

$$E_F(\sigma^{AB}) = \min_{\mathcal{E}(\sigma)} \left\{ \sum_{x \in \mathcal{X}} p(x) S(|\psi_x\rangle\langle\psi_x|) \right\},$$

com $\mathcal{E}(\sigma)$ ensembles do tipo $\{p(x), |\psi_x\rangle\langle\psi_x|\}_{x\in\mathcal{X}}$ tais que $\sigma^{AB} = \sum_{x\in\mathcal{X}} p(x)\,|\psi_x\rangle\langle\psi_x|$.

Proposição 4.2.7 ([16]). *As seguintes afirmações são equivalentes:*

1. *(Aditividade da Capacidade de Holevo) Para quaisquer canais quânticos $\mathcal{N}_1$ e $\mathcal{N}_2$:*

$$\chi(\mathcal{N}_1 \otimes \mathcal{N}_2) = \chi(\mathcal{N}_1) + \chi(\mathcal{N}_2). \tag{4.30}$$

2. *(Aditividade da Entropia de Saída Mínima) Para quaisquer dois canais quânticos $\mathcal{N}_1$ e $\mathcal{N}_2$:*

$$S_{min}(\mathcal{N}_1 \otimes \mathcal{N}_2) = S_{min}(\mathcal{N}_1) + S_{min}(\mathcal{N}_2). \tag{4.31}$$

3. *(Aditividade da Entropia de Formação) Para estados $\sigma_1 \in \mathcal{D}(\mathcal{H}_{A_1} \otimes \mathcal{H}_{B_1})$ e $\sigma_2 \in \mathcal{D}(\mathcal{H}_{A_2} \otimes \mathcal{H}_{B_2})$:*

$$E_F(\sigma_1 \otimes \sigma_2) = E_F(\sigma_1) + E_F(\sigma_2).$$

4. *(Superaditividade Forte da Entropia de Formação) Para um estado $\sigma \in \mathcal{D}(\mathcal{H}_{A_1} \otimes \mathcal{H}_{A_2} \otimes \mathcal{H}_{B_1} \otimes \mathcal{H}_{B_2})$:*

$$E_F(\sigma) \geq E_F(\mathrm{Tr}_2\{\sigma\}) + E_F(\mathrm{Tr}_1\{\sigma\}),$$

onde $E_F(\sigma)$ é calculado sobre a bipartição $A_1A_2|B_1B_2$, $\mathrm{Tr}_1\{\cdot\}$ é o traço parcial do sistema A_1B_1 e $\mathrm{Tr}_2\{\cdot\}$ é o traço parcial do sistema A_2B_2.

4.2.3 O Contraexemplo de Hastings

Vamos descrever o primeiro contraexemplo da conjectura da aditividade, obtido por Hastings [8], que utiliza o método probabilístico para demonstrar a existência de um canal cuja entropia de saída mínima não é aditiva. Os detalhes da demonstração envolvem a teoria de matrizes aleatórias e está muito além do escopo deste trabalho. Para mais detalhes, também ver [65, 66].

Para $n > 0$, sejam $\mathcal{U}_n$ o grupo de matrizes unitárias em $M_n(\mathbb{C})$ e μ_n a medida de Haar normalizada em $\mathcal{U}_n$, que é a (única) medida de probabilidade em $\mathcal{U}_n$ invariante por produto à esquerda ou direita por elementos de $\mathcal{U}_n$ (para mais sobre a medida de Haar, ver [67, 68]). Considere um canal quântico dado por:

$$\mathcal{N}(\rho) = \sum_{i=1}^{d} p_i U_i \rho U_i^\dagger,$$

onde $d > 0$, $\{p_i\}_{i=1}^d$ uma distribuição de probabilidade e U_i selecionados aleatoriamente de $\mathcal{U}_n$ pela medida de Haar. Seja então o canal:

$$\overline{\mathcal{N}}(\rho) = \sum_{i=1}^{d} p_i \overline{U}_i \rho \overline{U}_i^\dagger,$$

é possível então mostrar que existem n e d com $1 \ll d \ll n$ e $\{p_i\}_{i=1}^d$ tais que, com probabilidade não-nula:

$$S_{min}(\mathcal{N} \otimes \overline{\mathcal{N}}) < S_{min}(\mathcal{N}) + S_{min}(\overline{\mathcal{N}}) = 2S_{min}(\mathcal{N}).$$

Logo, pelo método probabilístico, deve existir pelo menos um canal $\mathcal{N}$ que satisfaz a relação acima, e pela Proposição 4.2.5, tal $\mathcal{N}$ é um contraexemplo da conjectura da aditividade.

Note que a demonstração não é construtiva. De fato, não há ainda um contraexemplo concreto para a conjectura da aditividade. Para mais sobre a conjectura da aditividade, ver [19, 17, 18].

Capítulo 5

O Teorema de Holevo-Schumacher-Westmoreland

Este capítulo se dedica à demonstração do Teorema de Holevo-Schumacher-Westmoreland, enunciado no Capítulo 4 (Teorema 4.1.16), que é o análogo quântico do Teorema da Codificação de Canais 1.3.1. Fixamos, no que segue, $(\mathcal{X}, \mathcal{N} \equiv \sigma_x^B, B)$ um canal C-Q. Lembre que $p_e(M,n) = \min_{(\mathcal{E},\Lambda)} P_e(\mathcal{E},\Lambda)$ é a mínima probabilidade média de erro para (M,n)-códigos $(\mathcal{E},\Lambda)$. Assim como o teorema clássico, vamos dividir a demonstração em duas partes:

Parte Direta: Para $R < \chi(\mathcal{N})$, usamos o método de codificação aleatória para mostrar a existência de uma sequência de códigos cujas taxas tendem à R mas com probabilidade máxima de erros tendendo a zero.

Parte Recíproca: Usamos o Teorema de Holevo e a aditividade da capacidade de Holevo para canais C-Q para mostrar que, se $R > \chi(\mathcal{N})$, então:

$$\lim_{n\to\infty} p_e(2^{nR}, n) > 0.$$

Vamos começar pela parte recíproca, que podemos demonstrar com os resultados do Capítulo 4. Para demonstrar a parte direta, vamos desenvolver a teoria de tipicalidade quântica.

5.1 Parte Recíproca

Lema 5.1.1. *Para todo $n \geq 1$, $\chi(\mathcal{N}^{\otimes n}) = n\chi(\mathcal{N})$. Ou seja, a capacidade de Holevo é aditiva para canais C-Q.*

Demonstração. Segue do Lema 3.1.12 e do Teorema 4.2.2, pois um canal C-Q é um caso particular de Canal EB, cuja capacidade de Holevo é aditiva. □

Demonstração da Parte Recíproca. Temos diretamente do Teorema de Holevo que para todo $n > 0$, $I_{acc}(\mathcal{N}^{\otimes n}) \leq \chi(\mathcal{N}^{\otimes n})$. Pela demonstração do Teorema 4.1.11, obtemos que:

$$\lim_{n\to\infty} p_e(2^{nR}, n) \geq 1 - \frac{1}{nR} \max_{\mathcal{E},\Lambda} I_{\mathcal{N}^{\otimes n}}(\mathcal{E},\Lambda) - \frac{1}{nR} \geq 1 - \lim_{n\to\infty} \frac{\chi(\mathcal{N}^{\otimes n})}{nR} = 1 - \frac{\chi(\mathcal{N})}{R},$$

com a última igualdade partindo do Lema 5.1.1. Assim, se $R > \chi(\mathcal{N})$, então temos $\lim_{n\to\infty} p_e(2^{nR}, n) > 0$.

□

Em [69, 70] foi demonstrada uma versão forte da parte recíproca, dando explicitamente um cota inferior para a probabilidade de erro de um código, mostrando que $\lim_{n\to\infty} p_e(2^{nR}, n) = 1$ caso a taxa assintótica dos códigos for maior que a capacidade de Holevo do canal. Para uma discussão, ver [71].

5.2 Tipicalidade Quântica

Nesta seção, vamos desenvolver a versão quântica da propriedade da equipartição assintótica vista na Seção 1.1.5.

5.2.1 Subespaços Típicos

Analogamente às sequências de variáveis aleatórias X^n i.i.d., vamos considerar estados do sistema quântico composto B^n do tipo:

$$\sigma^{\otimes n} = (\sigma^B)^{\otimes n} = \overbrace{\sigma^B \otimes \sigma^B \cdots \otimes \sigma^B}^{n \text{ vezes}},$$

onde σ^B tem decomposição espectral

$$\sigma^B = \sum_{y\in\mathcal{Y}} p_Y(y)\, |y\rangle\langle y|_B \tag{5.1}$$

com $|\mathcal{Y}| = d_B$, e portanto $(\sigma^B)^{\otimes n}$ tem decomposição espectral

$$\sigma^{\otimes n} = \sum_{y^n\in\mathcal{Y}^n} p_{Y^n}(y^n)\, |y^n\rangle\langle y^n|_{B^n}\,,$$

onde $|y^n\rangle = |y_1\rangle \otimes |y_2\rangle \otimes \cdots \otimes |y_{d_B}\rangle$ é o autovetor correspondente ao autovalor $y^n = y_1 y_2 \ldots y_{d_B}$ e $p_{Y^n}(y^n) = p_Y(y_1)p_Y(y_2)\cdots p_Y(y_n)$.

Definição 5.2.1 (Subespaço Típico)**.** O **subespaço δ-típico** $T_\delta^{B^n}$ associado ao estado σ^B como em (5.1) é o subespaço vetorial de $\mathcal{H}_B^{\otimes n}$ dado por:

$$T_\delta^{B^n} = \text{span}\{|y^n\rangle_{B^n} \mid y^n \in T_\delta^{Y^n}\},$$

onde $T_\delta^{Y^n}$ é conjunto δ-típico em relação a $Y \equiv p_Y$.

Chamamos de **projetor típico** o projetor $\Pi_\delta^{B^n}$ do subespaço $T_\delta^{B^n}$, ou seja:

$$\Pi_\delta^{B^n} = \sum_{y^n\in T_\delta^{Y^n}} |y^n\rangle\langle y^n|_{B^n}\,.$$

Para um $\delta > 0$ fixo, dizemos que um estado ρ^{B^n} é típico em relação a σ^B se os vetores associados à sua decomposição espectral pertencem a $T_\delta^{B^n}$, ou equivalentemente, se $\Pi_\delta^{B^n}\rho^{B^n} = \rho^{B^n}\Pi_\delta^{B^n}$.

Vejamos que subespaços típicos satisfazem resultados análogos aos Teoremas 1.1.33, 1.1.34 e 1.1.35, e suas demonstrações seguem diretamente desses teoremas. Tomamos $T_\delta^{B^n}$ associado ao estado σ^B dado por (5.1).

Teorema 5.2.2 (Equipartição). *Vale a desigualdade de operadores:*

$$2^{-n(S(\sigma^B)+\delta)}\Pi_\delta^{B^n} \leq \Pi_\delta^{B^n}\sigma^{\otimes n}\Pi_\delta^{B^n} \leq 2^{-n(S(\sigma^B)-\delta)}\Pi_\delta^{B^n}.$$

Equivalentemente, aplicando qualquer vetor $|y^n\rangle \in T_\delta^{B^n}$:

$$2^{-n(S(\sigma^B)+\delta)} \leq p_{Y^n}(y^n) \leq 2^{-n(S(\sigma^B)-\delta)}.$$

Teorema 5.2.3 (Acúmulo de Probabilidade). *Para todo $\epsilon \in (0,1)$ e $\delta > 0$, existe $n \in \mathbb{N}$ tal que*

$$\mathrm{Tr}\left\{\sigma^{\otimes n}\Pi_\delta^{B^n}\right\} \geq 1-\epsilon.$$

Teorema 5.2.4 (Dimensão). *Para todo $\epsilon \in (0,1)$, existe $n \in \mathbb{N}$ tal que:*

$$2^{n(S(\sigma^B)-\delta)} \leq \mathrm{Tr}\left\{\Pi_\delta^{B^n}\right\} \leq 2^{n(S(\sigma^B)+\delta)}.$$

Note que $\mathrm{Tr}\left\{\Pi_\delta^{B^n}\right\} = \dim(T_\delta^{B^n})$.

5.2.2 Tipicalidade Condicional Quântica

Considere o ensemble $\{p_X(x), \sigma_x^B\}_{x\in\mathcal{X}}$ em B, e σ^{XB} o estado clássico-quântico correspondente:

$$\sigma^{XB} = \sum_{x\in\mathcal{X}} p_X(x)\,|x\rangle\langle x| \otimes \sigma_x^B.$$

Pelo Teorema 3.2.22:

$$S(B|X)_\sigma = \sum_{x\in\mathcal{X}} p_X(x) S(\sigma_x^B). \tag{5.2}$$

Para cada $x \in \mathcal{X}$, escrevemos a decomposição espectral de σ_x^B como:

$$\sigma_x^B = \sum_{y\in\mathcal{Y}} p_{Y|X}(y|x)\,|y_x\rangle\langle y_x|_B\,,$$

onde novamente $|\mathcal{Y}| = d_B$. Pelo Teorema 3.2.3 e pela expressão (1.1) temos:

$$S(\sigma_x^B) = H(Y|X=x) = -\sum_{y\in\mathcal{Y}} p_{Y|X}(y|x)\log(p_{Y|X}(y|x)).$$

Substituindo em (5.2):

$$S(B|X)_\sigma = H(Y|X),$$

com $X \equiv p_X$ e $Y \equiv p_X \cdot p_{Y|X}$. Para n cópias do ensemble $\{p_X(x), \sigma_x^B\}_{x\in\mathcal{X}}$, corresponde o estado clássico quântico:

$$\sigma^{X^n B^n} = \sum_{x^n \in \mathcal{X}^n} p_{X^n}(x^n) \, |x^n\rangle\langle x^n| \otimes \sigma_{x^n}^{B^n},$$

onde

$$\sigma_{x^n}^B = \sum_{y^n \in \mathcal{Y}^n} p_{Y^n|X^n}(y^n|x^n) \, |y_{x^n}^n\rangle\langle y_{x^n}^n|_{B^n} . \tag{5.3}$$

Interpretando amostras independentes do ensemble $\{p_X(x), \sigma_x^B\}_{x\in\mathcal{X}}$ como emissões independentes de um canal clássico-quântico, uma entrada $x^n \in \mathcal{X}^n$ corresponde à saída $y^n \in \mathcal{Y}^n$ com probabilidade $p_{Y^n|X^n}(y^n|x^n)$ dada em (5.3).

Definição 5.2.5 (Subespaço Condicionalmente Típico)**.** Dado o ensemble $\{p_X(x), \sigma_x^B\}_{x\in\mathcal{X}}$ em B, o **subespaço condicionalmente δ-típico** com $x^n \in \mathcal{X}^n$ é dado por:

$$T_\delta^{B^n|x^n} = \text{span}\{|y_{x^n}^n\rangle_{B^n} \mid \left|\overline{H}(y^n|x^n) - S(B|X)_\sigma\right| < \delta\},$$

onde $\overline{H}(y^n|x^n)$ é dado na Definição 1.2.18.

Chamamos de **projetor condicionalmente típico** o projetor $\Pi_\delta^{B^n|x^n}$ do subespaço $T_\delta^{B^n|x^n}$, ou seja:

$$\Pi_\delta^{B^n|x^n} = \sum_{y^n \in T_\delta^{Y^n|x^n}} |y_{x^n}^n\rangle\langle y_{x^n}^n|_{B^n} .$$

Para um $\delta > 0$ fixo, dizemos que um estado ρ^{B^n} é típico em relação a σ^B se $\Pi_\delta^{B^n|x^n} \rho^{B^n} = \rho^{B^n} \Pi_\delta^{B^n|x^n}$.

Os seguintes teoremas seguem diretamente dos Teoremas 1.2.20, 1.2.21, 1.2.22:

Teorema 5.2.6 (Equipartição)**.** *Vale a desigualdade de operadores:*

$$2^{-n(S(B|X)_\sigma+\delta)} \Pi_\delta^{B^n|x^n} \leq \Pi_\delta^{B^n|x^n} \sigma_{x^n}^{\otimes n} \Pi_\delta^{B^n|x^n} \leq 2^{-n(S(B|X)_\sigma-\delta)} \Pi_\delta^{B^n|x^n}.$$

Equivalentemente, aplicando qualquer vetor $|y_{x^n}^n\rangle \in T_\delta^{B^n|x^n}$:

$$2^{-n(H(Y|X)+\delta)} \leq p_{Y^n|X^n}(y^n|x^n) \leq 2^{-n(H(Y|X)-\delta)}.$$

Teorema 5.2.7 (Acúmulo (em Média) de Probabilidade)**.** *Para todo $\epsilon \in (0,1)$ e $\delta > 0$, existe $n \in \mathbb{N}$ tal que:*

$$\mathbb{E}_{X^n} \left\{ \text{Tr} \left\{ \sigma_{X^n}^{B^n} \Pi_\delta^{B^n|X^n} \right\} \right\} \geq 1 - \epsilon. \tag{5.4}$$

Ou seja, em média com respeito a p_{X^n}, a probabilidade de encontrar o estado aleatório $\sigma_{X^n}^{B^n}$ no subespaço típico $T_\delta^{B^n|X^n}$ se aproxima de 1 para n suficientemente grande.

Teorema 5.2.8 (Dimensão). *Para todo $x^n \in \mathcal{X}^n$:*

$$\dim(T_\delta^{B^n|x^n}) = \mathrm{Tr}\left\{\Pi_\delta^{B^n|x^n}\right\} \leq 2^{n(S(B|X)_\sigma+\delta)}. \tag{5.5}$$

Além disso, temos que, para todo $\epsilon \in (0,1)$, existe $\delta > 0$ e n suficientemente grande tais que:

$$\mathbb{E}_{X^n}\left\{\mathrm{Tr}\left\{\Pi_\delta^{B^n|X^n}\right\}\right\} \geq (1-\epsilon)2^{n(S(B|X)_\sigma-\delta)}. \tag{5.6}$$

5.3 Parte Direta

Estamos prontos para demonstrar (segundo principalmente [12]) a parte direta do Teorema de HSW:

> Se $0 < R < \chi(\mathcal{N})$, então existe uma sequência de $(2^{nR}, n)$-códigos quânticos $(\mathcal{E}^n, \Lambda^n)$ tais que $P_e(\mathcal{E}^n, \Lambda^n) \to 0$ e portanto $p_e(2^{nR}, n) \to 0$.

Podemos assumir, sem perda de generalidade, que, para cada $n > 0$, o canal Q-C Λ^n associado a cada código tem contradomínio $\mathcal{M}$ (em termos da Definição 4.1.8, estamos tomando $\Lambda^n := \mathcal{D} \circ \Lambda$).

Fixamos $p_X \in \mathcal{P}(\mathcal{X})$ tal que $\chi(\mathcal{N}) = \chi(\{p_X(x), \sigma_x^B\})$ e $n > 0$. Identicamente à demonstração do Teorema 1.3.1, vamos considerar códigos $\mathcal{C}^n : \mathcal{M} \to \mathcal{X}^n$ aleatórios com distribuição dada por (1.15):

$$p_{X^n}(x^n(m)) = \prod_{i=1}^n p_X(x_i(m)) = \prod_{i=1}^n p_X(x_i).$$

No que segue, para simplificar a notação, vamos denotar $\mathcal{C}^n$ por $\mathcal{C}$ quando n for fixo.

Seja $\delta > 0$ tal que $R < \chi(\mathcal{N}) - 3\delta$. Para cada $x^n \in \mathcal{X}^n$ vamos denotar por

$$\Pi_{x^n} := \Pi_\delta^{B^n|x^n},$$

o projetor condicionalmente típico de x^n associado ao ensemble $\mathcal{E} = \{p_X(x), \sigma_x^B\}_{x\in\mathcal{X}}$. Escrevemos:

$$\sigma = \sum_{x\in\mathcal{X}} p_X(x)\sigma_x^B,$$

e denotamos por:

$$\Pi_\sigma := \Pi_\delta^{B^n},$$

o projetor típico associado ao estado $\sigma^{\otimes n}$.

Note que escolhemos p_X tal que, se $\sigma^{XB} = \sum_{x\in\mathcal{X}} p_X(x)\,|x\rangle\langle x| \otimes \sigma_X^B$, então:

$$\chi(\mathcal{N}) = \mathcal{I}(X:B) = S(\sigma) - S(B|X)_\sigma. \tag{5.7}$$

Analogamente ao caso clássico, onde utilizamos funções indicadoras dos conjuntos típicos para identificar erros de transmissão, podemos utilizar projetores típicos para identificar a presença de estados em subespaços típicos.

Seja:

$$P = \sum_{m\in\mathcal{M}} \Pi_\sigma \Pi_{x^n(m)} \Pi_\sigma.$$

Note que $P > 0$, já que é a soma da conjugação de matrizes positivas por matrizes positivas (ver [72]).

Note que os estados que pertencem à imagem de P são exatamente os estados típicos em relação a σ que são condicionalmente típicos com algum $x^n(m) \in \mathcal{X}^n$ em relação a $\mathcal{E}$.

Para cada $m \in \mathcal{M}$, considere os operadores positivos:

$$P_m = P^{-\frac{1}{2}} \Pi_\sigma \Pi_{x^n(m)} \Pi_\sigma P^{-\frac{1}{2}},$$

onde aqui denotamos por A^{-1} a **pseudo-inversa de Moore–Penrose** de $A > 0$, que é a matriz não-negativa tal que $AA^{-1} = A^{-1}A = \Pi_A$, com Π_A a projeção na imagem de A.

Denotando por Π_P a projeção sobre a imagem de P, note que:

$$\sum_{m\in\mathcal{M}} P_m = \Pi_P.$$

Bob pode então realizar a seguinte medição sobre B^n:

$$\Lambda_m = P_m + \frac{1}{M}(\mathbb{1}_{B^n} - \Pi_P).$$

De fato, $\mathbb{1}_{B^n} - \Pi_P \geq 0$ pois $\Pi_P \leq \mathbb{1}_{B^n}$ e $\sum_m \Lambda_m = \mathbb{1}_{B^n}$.

A probabilidade média de erro com o código resultante $(\mathcal{E}, \Lambda)$ é:

$$P_e(\mathcal{E}, \Lambda) = \frac{1}{M} \sum_{m\in\mathcal{M}} \mathrm{Tr}\left\{\sigma^{B^n}_{x^n(m)}(\mathbb{1}_{B^n} - \Lambda_m)\right\} \leq \frac{1}{M} \sum_{m\in\mathcal{M}} \mathrm{Tr}\left\{\sigma^{B^n}_{x^n(m)}(\mathbb{1}_{B^n} - P_m)\right\} \qquad (5.8)$$

Nosso objetivo é obter uma estimativa para $\mathbb{E}_\mathcal{C}\{P_e(\mathcal{E}, \Lambda)\}$.

Na demonstração da parte direta do Teorema 1.3.1, obtemos uma cota superior da probabilidade de erro considerando três tipos de erro, a desigualdade (1.17):

$$\mathbb{E}_\mathcal{C}\left\{P_e\right\} \leq \mathbb{E}_\mathcal{C}\{p_{Y^n|\mathcal{C}}\{\mathcal{E}_0(m)\}\} + \mathbb{E}_\mathcal{C}\{p_{Y^n|\mathcal{C}}\{\mathcal{E}_1(m)\}\} + \mathbb{E}_\mathcal{C}\{p_{Y^n|\mathcal{C}}\{\mathcal{E}_2(m)\}\}.$$

Essa desigualdade segue simplesmente da subaditividade da probabilidade. No caso quântico, este tipo de relação se torna não-trivial (ver [73]) devido à não-comutatividade dos operadores envolvidos, em contraste às funções indicadores do caso clássico. A seguinte desigualdade cumpre esse papel na teoria quântica.

Proposição 5.3.1 (Desigualdade de Hayashi-Nagaoka [12, 74, 75])**.** *Para $S, T \in \mathcal{B}(\mathcal{H})^+$ com $S \leq \mathbb{1}$, temos que:*

$$\mathbb{1} - (S+T)^{-\frac{1}{2}} S (S+T)^{-\frac{1}{2}} \leq 2(\mathbb{1} - S) + 4T,$$

onde a inversa é a pseudo-inversa de Moore–Penrose.

Não é imediato que a desigualdade de Hayashi-Nagaoka fornece uma cota similar a (1.17), mas vamos ver que este é o caso.

Considerando que $P = \Pi_\sigma \Pi_{x^n(m)} \Pi_\sigma + (P - \Pi_\sigma \Pi_{x^n(m)} \Pi_\sigma)$,

$$\begin{aligned} \mathbb{1}_{B^n} - P_m &= \mathbb{1}_{B^n} - P^{-\frac{1}{2}} \Pi_\sigma \Pi_{x^n(m)} \Pi_\sigma P^{-\frac{1}{2}} \\ &\leq 2(\mathbb{1}_{B^n} - \Pi_\sigma \Pi_{x^n(m)} \Pi_\sigma) + 4(P - \Pi_\sigma \Pi_{x^n(m)} \Pi_\sigma). \end{aligned}$$

Substituindo em (5.8):

$$\begin{aligned} P_e(\mathcal{E}, \Lambda) &\leq \frac{1}{M} \sum_{m \in \mathcal{M}} \operatorname{Tr}\left\{ \sigma^{B^n}_{x^n(m)} (\mathbb{1}_{B^n} - P_m) \right\} \\ &\leq \frac{2}{M} \sum_{m \in \mathcal{M}} \operatorname{Tr}\left\{ \sigma^{B^n}_{x^n(m)} (\mathbb{1}_{B^n} - \Pi_\sigma \Pi_{x^n(m)} \Pi_\sigma) \right\} + \frac{4}{M} \sum_{m \in \mathcal{M}} \operatorname{Tr}\left\{ \sigma^{B^n}_{x^n(m)} (P - \Pi_\sigma \Pi_{x^n(m)} \Pi_\sigma) \right\} .QQ \end{aligned} \tag{5.9}$$

Note que a primeira parcela à direita da desigualdade acima é múltiplo da probabilidade da interseção dos eventos:

$\mathcal{E}_0(m) : \sigma^{B^n}_{x^n(m)}$ não é típico em relação à σ.

$\mathcal{E}_1(m) : \sigma^{B^n}_{x^n(m)}$ é típico em relação à σ mas $\sigma^{B^n}_{x^n(m)}$ não é condicionalmente típico com algum $x^n(m)$ relação à $\mathcal{E}$.

Já a segunda parcela é múltipla da probabilidade do evento:

$\mathcal{E}_2(m) : \sigma^{B^n}_{x^n(m)}$ é típico em relação à σ e $\sigma^{B^n}_{x^n(m)}$ é condicionalmente típico com $x^n(\hat{m})$ em relação à $\mathcal{E}$, onde $\hat{m} \neq m$.

Por definição $\sigma^{B^n}_{x^n(m)}$ é condicionalmente típico com $x^n(m)$ em relação à $\mathcal{E}$, logo:

$$\sigma^{B^n}_{x^n(m)} \Pi_{x^n(m)} = \Pi_{x^n(m)} \sigma^{B^n}_{x^n(m)} \tag{5.10}$$

Vamos começar analisando o primeiro termo do lado direito de (5.9). Para isso, vamos usar a seguinte identidade, cuja verificação é imediata:

$$\Pi_\sigma \Pi_{x^n(m)} \Pi_\sigma = \Pi_\sigma \Pi_{x^n(m)} + \Pi_{x^n(m)} \Pi_\sigma - \Pi_{x^n(m)} + (\mathbb{1}_{B^n} - \Pi_\sigma) \Pi_{x^n(m)} (\mathbb{1}_{B^n} - \Pi_\sigma)$$

Verificamos que:

$$\begin{aligned} \operatorname{Tr}\left\{ \sigma^{B^n}_{x^n(m)} \Pi_\sigma \Pi_{x^n(m)} \Pi_\sigma) \right\} &= \operatorname{Tr}\left\{ \sigma^{B^n}_{x^n(m)} \Pi_\sigma \Pi_{x^n(m)} \right\} + \operatorname{Tr}\left\{ \sigma^{B^n}_{x^n(m)} \Pi_{x^n(m)} \Pi_\sigma \right\} \\ &\quad - \operatorname{Tr}\left\{ \sigma^{B^n}_{x^n(m)} \Pi_{x^n(m)} \right\} + \operatorname{Tr}\left\{ \sigma^{B^n}_{x^n(m)} (\mathbb{1}_{B^n} - \Pi_\sigma) \Pi_{x^n(m)} (\mathbb{1}_{B^n} - \Pi_\sigma) \right\} \\ &\geq \operatorname{Tr}\left\{ \sigma^{B^n}_{x^n(m)} \Pi_\sigma \Pi_{x^n(m)} \right\} + \operatorname{Tr}\left\{ \sigma^{B^n}_{x^n(m)} \Pi_{x^n(m)} \Pi_\sigma \right\} - \operatorname{Tr}\left\{ \sigma^{B^n}_{x^n(m)} \Pi_{x^n(m)} \right\} \end{aligned}$$

que pela propriedade cíclica do traço e por (5.10):

$$\begin{aligned} &= \operatorname{Tr}\left\{ \sigma^{B^n}_{x^n(m)} \Pi_{x^n(m)} \Pi_\sigma \right\} + \operatorname{Tr}\left\{ \sigma^{B^n}_{x^n(m)} \Pi_{x^n(m)} \Pi_\sigma \right\} - \operatorname{Tr}\left\{ \sigma^{B^n}_{x^n(m)} \Pi_{x^n(m)} \right\} \\ &= \operatorname{Tr}\left\{ \sigma^{B^n}_{x^n(m)} \Pi_{x^n(m)} (2\Pi_\sigma - \mathbb{1}_{B^n}) \right\} \end{aligned}$$

É fácil verificar que:

$$\sigma^{B^n}_{x^n(m)} \Pi_{x^n(m)} (2\Pi_\sigma - \mathbb{1}_{B^n}) = \sigma^{B^n}_{x^n(m)} (2\Pi_\sigma - \mathbb{1}_{B^n}) + \sigma^{B^n}_{x^n(m)} (\mathbb{1}_{B^n} - \Pi_{x^n(m)}) (\mathbb{1}_{B^n} - 2\Pi_\sigma)$$

Como $\Pi_\sigma \leq \mathbb{1}_{B^n}$, $2\Pi_\sigma - \mathbb{1}_{B^n} \leq \mathbb{1}_{B^n}$ e portanto $\mathbb{1}_{B^n} - 2\Pi_\sigma \geq -\mathbb{1}_{B^n}$, temos:

$$\begin{aligned}
\mathrm{Tr}\left\{\sigma^{B^n}_{x^n(m)}\Pi_\sigma\Pi_{x^n(m)}\Pi_\sigma)\right\} &\geq \mathrm{Tr}\left\{\sigma^{B^n}_{x^n(m)}\Pi_{x^n(m)}(2\Pi_\sigma - \mathbb{1}_{B^n})\right\} \\
&= \mathrm{Tr}\left\{\sigma^{B^n}_{x^n(m)}(2\Pi_\sigma - \mathbb{1}_{B^n})\right\} + \mathrm{Tr}\left\{\sigma^{B^n}_{x^n(m)}(\mathbb{1}_{B^n} - \Pi_{x^n(m)})(\mathbb{1}_{B^n} - 2\Pi_\sigma)\right\} \\
&\geq \mathrm{Tr}\left\{\sigma^{B^n}_{x^n(m)}(2\Pi_\sigma - \mathbb{1}_{B^n})\right\} - \mathrm{Tr}\left\{\sigma^{B^n}_{x^n(m)}(\mathbb{1}_{B^n} - \Pi_{x^n(m)})\right\} \\
&= 2\mathrm{Tr}\left\{\sigma^{B^n}_{x^n(m)}\Pi_\sigma\right\} + \mathrm{Tr}\left\{\sigma^{B^n}_{x^n(m)}\Pi_{x^n(m)}\right\} - 2 \qquad (5.11)
\end{aligned}$$

Note que:

$$\mathbb{E}_C\left\{\mathrm{Tr}\left\{\sigma^{B^n}_{x^n(m)}\Pi_\sigma\right\}\right\} = \mathrm{Tr}\left\{\mathbb{E}_C\left\{\sigma^{B^n}_{x^n(m)}\Pi_\sigma\right\}\right\},$$

e que, como o código de uma mensagem é escolhida independentemente da mensagem:

$$\mathbb{E}_{\mathcal{C}}\left\{\sigma^{B^n}_{x^n(m)}\right\} = \sum_{x^n\in\mathcal{X}^n} p_{X^n}(x^n)\sigma^{B^n}_{x^n} = \sigma^{\otimes n}. \qquad (5.12)$$

Portanto:

$$\mathbb{E}_{\mathcal{C}}\left\{\mathrm{Tr}\left\{\sigma^{B^n}_{x^n(m)}\Pi_\sigma\right\}\right\} = \mathrm{Tr}\left\{\mathbb{E}_{\mathcal{C}}\left\{\sigma^{B^n}_{x^n(m)}\right\}\Pi_\sigma\right\} = \mathrm{Tr}\left\{\sigma^{\otimes n}\Pi_\sigma\right\}, \qquad (5.13)$$

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}}\left\{\frac{2}{M}\sum_{m\in\mathcal{M}}\mathrm{Tr}\left\{\sigma^{B^n}_{x^n(m)}(\mathbb{1}_{B^n} - \Pi_\sigma\Pi_{x^n(m)}\Pi_\sigma)\right\}\right\} \\
&= \frac{2}{M}\sum_{m\in\mathcal{M}}\mathbb{E}_{\mathcal{C}}\left\{\mathrm{Tr}\left\{\sigma^{B^n}_{x^n(m)}(\mathbb{1}_{B^n} - \Pi_\sigma\Pi_{x^n(m)}\Pi_\sigma)\right\}\right\} \\
&= \frac{2}{M}\sum_{m\in\mathcal{M}}\mathbb{E}_{\mathcal{C}}\left\{1 - \mathrm{Tr}\left\{(\sigma^{B^n}_{x^n(m)}\Pi_\sigma\Pi_{x^n(m)}\Pi_\sigma)\right\}\right\}
\end{aligned}$$

que por (5.11):

$$\begin{aligned}
&\leq \frac{2}{M}\sum_{m\in\mathcal{M}}\mathbb{E}_{\mathcal{C}}\left\{3 - 2\mathrm{Tr}\left\{\sigma^{B^n}_{x^n(m)}\Pi_\sigma\right\} - \mathrm{Tr}\left\{\sigma^{B^n}_{x^n(m)}\Pi_{x^n(m)}\right\}\right\} \\
&= \frac{2}{M}\sum_{m\in\mathcal{M}} 3 - 2\mathbb{E}_C\left\{\mathrm{Tr}\left\{\sigma^{B^n}_{x^n(m)}\Pi_\sigma\right\}\right\} - \mathbb{E}_C\left\{\mathrm{Tr}\left\{\sigma^{B^n}_{x^n(m)}\Pi_{x^n(m)}\right\}\right\}
\end{aligned}$$

e por (5.13):

$$= \frac{2}{M}\sum_{m\in\mathcal{M}} 3 - 2\mathrm{Tr}\left\{\sigma^{\otimes n}\Pi_\sigma\right\} - \mathbb{E}_C\left\{\mathrm{Tr}\left\{\sigma^{B^n}_{x^n(m)}\Pi_{x^n(m)}\right\}\right\}$$

Portanto, quando $n \to \infty$, pelos Teoremas 5.2.3 e 5.2.7:

$$\mathbb{E}_{\mathcal{C}^n}\left\{\frac{2}{M}\sum_{m\in\mathcal{M}}\mathrm{Tr}\left\{\sigma^{B^n}_{x^n(m)}(\mathbb{1}_{B^n} - \Pi_\sigma\Pi_{x^n(m)}\Pi_\sigma)\right\}\right\} \to 0, \qquad (5.14)$$

pois $\mathrm{Tr}\{\sigma^{\otimes n}\Pi_\sigma\} \to 1$ e $\mathbb{E}_{\mathcal{C}^n}\left\{\mathrm{Tr}\left\{\sigma^{B^n}_{x^n(m)}\Pi_{x^n(m)}\right\}\right\} \to 1$.

Vamos agora analisar o segundo termo do lado direito de (5.9). Note que:

$$\begin{aligned}\mathrm{Tr}\left\{\sigma_{x^n(m)}^{B^n}(P-\Pi_\sigma\Pi_{x^n(m)}\Pi_\sigma)\right\} &= \sum_{\hat{m}\neq m}\mathrm{Tr}\left\{\sigma_{x^n(m)}^{B^n}(\Pi_\sigma\Pi_{x^n(\hat{m})}\Pi_\sigma)\right\}\\ &= \sum_{\hat{m}\neq m}\mathrm{Tr}\left\{(\Pi_\sigma\sigma_{x^n(m)}^{B^n}\Pi_\sigma)\Pi_{x^n(\hat{m})}\right\}\end{aligned}$$

Como a $x^n(m)$ e $x^n(\hat{m})$ são independentes sob $\mathcal{C}$, e por (5.12):

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\left\{\mathrm{Tr}\left\{(\Pi_\sigma\sigma_{x^n(m)}^{B^n}\Pi_\sigma)\Pi_{x^n(\hat{m})}\right\}\right\} &= \mathrm{Tr}\left\{\mathbb{E}_{\mathcal{C}}\left\{(\Pi_\sigma\sigma_{x^n(m)}^{B^n}\Pi_\sigma)\Pi_{x^n(\hat{m})}\right\}\right\}\\ &= \mathrm{Tr}\left\{\mathbb{E}_{\mathcal{C}}\left\{(\Pi_\sigma\sigma_{x^n(m)}^{B^n}\Pi_\sigma)\right\}\mathbb{E}_{\mathcal{C}}\left\{\Pi_{x^n(\hat{m})}\right\}\right\}\\ &= \mathrm{Tr}\left\{(\Pi_\sigma\sigma^{\otimes n}\Pi_\sigma)\Pi_{x^n}\right\}\end{aligned}$$

Portanto, temos:

$$\begin{aligned}&\mathbb{E}_{\mathcal{C}}\left\{\frac{4}{M}\sum_{m\in\mathcal{M}}\mathrm{Tr}\left\{\sigma_{x^n(m)}^{B^n}(P-\Pi_\sigma\Pi_{x^n(m)}\Pi_\sigma)\right\}\right\}\\ &= \frac{4}{M}\sum_{m\in\mathcal{M}}\sum_{\hat{m}\neq m}\mathbb{E}_{\mathcal{C}}\left\{\mathrm{Tr}\left\{(\Pi_\sigma\sigma_{x^n(m)}^{B^n}\Pi_\sigma)\Pi_{x^n(\hat{m})}\right\}\right\}\\ &= \frac{4}{M}\sum_{m\in\mathcal{M}}\sum_{\hat{m}\neq m}\mathrm{Tr}\left\{(\Pi_\sigma\sigma^{\otimes n}\Pi_\sigma)\Pi_{x^n}\right\}\\ &= 4(M-1)\mathrm{Tr}\left\{(\Pi_\sigma\sigma^{\otimes n}\Pi_\sigma)\Pi_{x^n}\right\}\\ &\leq 4(M-1)2^{-n(S(\sigma)-\delta)}\mathrm{Tr}\left\{\Pi_\sigma\Pi_{x^n}\right\} &(5.15)\\ &\leq 4(M-1)2^{-n(S(\sigma)-\delta)}\mathrm{Tr}\left\{\Pi_{x^n}\right\} &(5.16)\\ &\leq 4(M-1)2^{-n(S(\sigma)-\delta)}2^{n(S(B|X)_\sigma+\delta)} &(5.17)\\ &\leq 4M2^{-n(S(\sigma)-\delta)}2^{n(S(B|X)_\sigma+\delta)}\\ &= 4(2^{n(R-(S(\sigma)-S(B|X)_\sigma)+2\delta)})\\ &= 4(2^{-n(\chi(\mathcal{N})-2\delta-R)}) &(5.18)\\ &\leq 4(2^{-n\delta}), &(5.19)\end{aligned}$$

onde (5.15) segue do Teorema 5.2.2, $\Pi_\sigma\sigma^{\otimes n}\Pi_\sigma \leq 2^{-n(S(\sigma^B)-\delta)}\Pi_\sigma$ e pela monotonicidade do traço ([76]). Já (5.16) segue do fato de que, para matrizes positivas X e Y, $AXA^* \leq AYA^*$ para qualquer matriz A, e portanto, como $\Pi_\sigma \leq \mathbb{1}_{B^n}$:

$$\begin{aligned}\mathrm{Tr}\left\{\Pi_\sigma\Pi_{x^n}\right\} &= \mathrm{Tr}\left\{\Pi_{x^n}\Pi_\sigma\Pi_{x^n}\right\}\\ &\leq \mathrm{Tr}\left\{\Pi_{x^n}\mathbb{1}_{B^n}\Pi_{x^n}\right\}\\ &= \mathrm{Tr}\left\{\Pi_{x^n}\right\}.\end{aligned}$$

(5.17) segue do Teorema 5.2.8. (5.18), que em algumas referências é chamada de **Lema do Empacotamento Quântico**, segue de $M = 2^{nR}$ e da escolha de σ^B como o estado que satisfaz (5.7). (5.19) segue da escolha de δ tal que $R < \chi(\mathcal{N}) - 3\delta$, ou seja, tal que $\delta < \chi(\mathcal{N}) - 2\delta - R$. Logo, tomando $n \to \infty$, obtemos:

$$\mathbb{E}_{\mathcal{C}^n}\left\{\frac{4}{M}\sum_{m\in\mathcal{M}}\mathrm{Tr}\left\{\sigma_{x^n(m)}^{B^n}(P-\Pi_\sigma\Pi_{x^n(m)}\Pi_\sigma)\right\}\right\} \to 0. \tag{5.20}$$

Finalmente, de (5.9), (5.14) e (5.20), temos que, com $n \to \infty$:

$$\mathbb{E}_{\mathcal{C}^n} \{P_e(\mathcal{E}^n, \Lambda^n)\} \to 0.$$

Concluímos que deve existir uma sequência de códigos $\mathcal{C}^n$ tais que, para os $(2^{nR}, n)$-códigos associados $(\mathcal{E}^n, \Lambda^n)$, temos $P_e(\mathcal{E}^n, \Lambda^n) \to 0$ e portanto que $p_e(M, n) \to 0$.

Referências Bibliográficas

[1] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[2] O. Freire Junior and I. M. Greca, "Informação e teoria quântica," *Scientiae Studia*, vol. 11, p. 11–33, Jan 2013.

[3] C. Bennett and P. Shor, "Quantum information theory," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2724–2742, 1998.

[4] B. Schumacher, "Quantum coding," *Phys. Rev. A*, vol. 51, pp. 2738–2747, Apr 1995.

[5] B. N. Lima, L. M. Cioletti, M. d. O. T. Cunha, and G. A. Braga, "Entropia: introdução à teoria matemática da (des) informação,"

[6] A. Holevo, "The capacity of the quantum channel with general signal states," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 269–273, 1998.

[7] B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channels," *Phys. Rev. A*, vol. 56, pp. 131–138, Jul 1997.

[8] M. B. Hastings, "Superadditivity of communication capacity using entangled inputs," *Nature Physics*, vol. 5, p. 255–257, Mar. 2009.

[9] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley, 2012.

[10] M. M. Wilde, *Quantum information theory*. Cambridge university press, 2013.

[11] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.

[12] J. Watrous, *The Theory of Quantum Information*. Cambridge University Press, 2018.

[13] A. S. Holevo, *Quantum Systems, Channels, Information*. Berlin, Boston: De Gruyter, 2013.

[14] C. King and M. B. Ruskai, "Capacity of quantum channels using product measurements," *Journal of Mathematical Physics*, vol. 42, pp. 87–98, 01 2001.

[15] P. W. Shor, "Additivity of the classical capacity of entanglement-breaking quantum channels," *Journal of Mathematical Physics*, vol. 43, p. 4334–4340, Sept. 2002.

[16] P. W. Shor, "Equivalence of additivity questions in quantum information theory," *Communications in Mathematical Physics*, vol. 246, p. 473–473, Apr. 2004.

[17] C. King, "Remarks on the additivity conjectures for quantum channels," 01 2010.

[18] K. Matsumoto, *On Additivity Questions*, pp. 133–164. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.

[19] A. Holevo, "The additivity problem in quantum information theory," *Proceedings oh the International Congress of Mathematicians, Vol. 3, 2006-01-01, ISBN 978-3-03719-022-7, pags. 999-1018*, vol. 3, 01 2015.

[20] G. G. Amosov, A. S. Holevo, and R. F. Werner, "On some additivity problems in quantum information theory," 2000.

[21] I. Savov, "Network information theory for classical-quantum channels," 2012.

[22] D. Applebaum, *Probability and Information: An Integrated Approach*. Cambridge University Press, 2008.

[23] K. Parthasarathy, *Coding theorems of classical and quantum information theory*. Texts and Readings in Mathematics, Hindustan Book Agency, 2013.

[24] Y. Polyanskiy and Y. Wu, *Information Theory: From Coding to Learning*. Cambridge University Press, 2025.

[25] D. MacKay, *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.

[26] F. Benatti, *Dynamics, information and complexity in quantum systems*, vol. 2009. 01 2009.

[27] E. Witten, "A mini-introduction to information theory," *La Rivista del Nuovo Cimento*, vol. 43, p. 187–227, Mar. 2020.

[28] A. O. Lopes and J. K. Mengue, "On information gain, kullback-leibler divergence, entropy production and the involution kernel," 2021.

[29] F. Nielsen, "An elementary introduction to information geometry," *Entropy*, vol. 22, p. 1100, Sept. 2020.

[30] P. Billingsley, *Probability and Measure*. Wiley Series in Probability and Statistics, Wiley, 2012.

[31] J. Wolfowitz, "Strong converse of the coding theorem for semicontinuous channels," *Illinois Journal of Mathematics*, vol. 3, no. 4, pp. 477–489, 1959.

[32] A. E. Gamal and Y.-H. Kim, "Lecture notes on network information theory," 2011.

[33] J. Watrous, "Understanding quantum information and computation," 2025.

[34] A. Mueller-Hermes, "Lecture notes - quantum information theory.".

[35] R. B. Griffiths, *Consistent Quantum Theory*. Cambridge University Press, 2001.

[36] D. A. Lidar, "Lecture notes on the theory of open quantum systems," 2020.

[37] K. Hoffman and R. Kunze, *Linear Algebra*. Prentice-Hall, 1971.

[38] N. Jeevanjee, *An Introduction to Tensors and Group Theory for Physicists*. An Introduction to Tensors and Group Theory for Physicists, Birkhäuser Boston, 2011.

[39] V. Paulsen, *Completely Bounded Maps and Operator Algebras*. Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2003.

[40] G. Adesso, M. Cianciaruso, and T. R. Bromley, "An introduction to quantum discord and non-classical correlations beyond entanglement," 2016.

[41] M. Jiang, S. Luo, and S. Fu, "Channel-state duality," *Phys. Rev. A*, vol. 87, p. 022310, Feb 2013.

[42] J. Maziero, "Computing partial traces and reduced density matrices," *International Journal of Modern Physics C*, vol. 28, no. 01, p. 1750005, 2017.

[43] A. Jamiołkowski, "Linear transformations which preserve trace and positive semidefiniteness of operators," *Reports on Mathematical Physics*, vol. 3, no. 4, pp. 275–278, 1972.

[44] M. Frembs and E. G. Cavalcanti, "Variations on the choi–jamiołkowski isomorphism," *Journal of Physics A: Mathematical and Theoretical*, vol. 57, p. 265301, June 2024.

[45] G. Homa, A. Ortega, and M. Koniorczyk, "Choi representation of completely positive maps in brief," *Zeitschrift für Naturforschung A*, vol. 79, no. 12, pp. 1123–1133, 2024.

[46] M.-D. Choi, "Completely positive linear maps on complex matrices," *Linear Algebra and its Applications*, vol. 10, no. 3, pp. 285–290, 1975.

[47] M. Horodecki, P. W. Shor, and M. B. Ruskai, "Entanglement breaking channels," *Reviews in Mathematical Physics*, vol. 15, no. 06, pp. 629–641, 2003.

[48] A. Kalev and I. Hen, "No-broadcasting theorem and its classical counterpart," *Physical Review Letters*, vol. 100, May 2008.

[49] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher, "Noncommuting mixed states cannot be broadcast," *Physical Review Letters*, vol. 76, p. 2818–2821, Apr. 1996.

[50] N. Higham, *Functions of Matrices: Theory and Computation*. Other Titles in Applied Mathematics, Society for Industrial and Applied Mathematics, 2008.

[51] E. Davies, "Information and quantum measurement," *IEEE Transactions on Information Theory*, vol. 24, no. 5, pp. 596–599, 1978.

[52] J. Preskill, "Lecture notes for physics 229: Quantum information and computation.".

[53] A. S. Holevo and A. V. Utkin, "Quantum accessible information and classical entropy inequalities," 2025.

[54] M. Sasaki, S. M. Barnett, R. Jozsa, M. Osaki, and O. Hirota, "Accessible information and optimal strategies for real symmetrical quantum sources," *Phys. Rev. A*, vol. 59, pp. 3325–3335, May 1999.

[55] A. Fujiwara and H. Nagaoka, "Operational capacity and pseudoclassicality of a quantum channel," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 1071–1086, 1998.

[56] H. Barnum, M. A. Nielsen, and B. Schumacher, "Information transmission through a noisy quantum channel," *Phys. Rev. A*, vol. 57, pp. 4153–4175, Jun 1998.

[57] A. S. Holevo, "Bounds for the quantity of information transmitted by a quantum communication channel," 1973.

[58] M. Sasaki, K. Kato, M. Izutsu, and O. Hirota, "A demonstration of superadditivity in the classical capacity of a quantum channel," *Physics Letters A*, vol. 236, no. 1, pp. 1–4, 1997.

[59] A. Holevo, "Problems in the mathematical theory of quantum communication channels," *Reports on Mathematical Physics*, vol. 12, no. 2, pp. 273–278, 1977.

[60] C. H. Bennett, C. A. Fuchs, and J. A. Smolin, *Entanglement-Enhanced Classical Communication on a Noisy Quantum Channel*, pp. 79–88. Boston, MA: Springer US, 1997.

[61] C. King, "The capacity of the quantum depolarizing channel," *IEEE Transactions on Information Theory*, vol. 49, no. 1, pp. 221–229, 2003.

[62] C. King, "Additivity for unital qubit channels," *Journal of Mathematical Physics*, vol. 43, p. 4641–4653, Oct. 2002.

[63] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, "Concentrating partial entanglement by local operations," *Physical Review A*, vol. 53, p. 2046–2052, Apr. 1996.

[64] G. Vidal, W. Dür, and J. I. Cirac, "Entanglement cost of bipartite mixed states," *Physical Review Letters*, vol. 89, June 2002.

[65] M. Fukuda, C. King, and D. K. Moser, "Comments on hastings' additivity counterexamples," *Communications in Mathematical Physics*, vol. 296, p. 111–143, Feb. 2010.

[66] F. G. S. L. Brandão and M. Horodecki, "On hastings' counterexamples to the minimum output entropy additivity conjecture," *Open Systems & Information Dynamics*, vol. 17, p. 31–52, Mar. 2010.

[67] A. A. Mele, "Introduction to haar measure tools in quantum information: A beginner's tutorial," *Quantum*, vol. 8, p. 1340, May 2024.

[68] B. Simon, *Representations of Finite and Compact Groups*. Graduate Studies in Mathematics, American Mathematical Society, 2025.

[69] T. Ogawa and H. Nagaoka, "Strong converse to the quantum channel coding theorem," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2486–2489, 1999.

[70] A. Winter, "Coding theorem and strong converse for quantum channels," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2481–2485, 1999.

[71] M. Mosonyi and T. Ogawa, "Strong converse exponent for classical-quantum channel coding," *Communications in Mathematical Physics*, vol. 355, p. 373–426, June 2017.

[72] R. Bhatia, *Positive Definite Matrices*. Princeton Series in Applied Mathematics, Princeton University Press, 2015.

[73] H.-C. Cheng, "Simple and tighter derivation of achievability for classical communication over quantum channels," *PRX Quantum*, vol. 4, Nov. 2023.

[74] M. Hayashi and H. Nagaoka, "General formulas for capacity of classical-quantum channels," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1753–1768, 2003.

[75] J. C. A. Barata and M. S. Hussein, "The moore–penrose pseudoinverse: A tutorial review of the theory," *Brazilian Journal of Physics*, vol. 42, p. 146–165, Dec. 2011.

[76] E. Carlen, "Trace inequalities and quantum entropy: An introductory course," vol. 529, 01 2010.