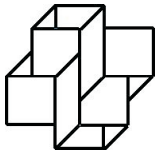

Laboratório Nacional de Computação Científica

Programa de Pós Graduação em Modelagem Computacional

Uma nova metodologia para o cálculo da informação acessível

Por

Michael Ferreira de Souza

PETRÓPOLIS, RJ - BRASIL

MARÓ DE 2007

UMA NOVA METODOLOGIA PARA O CÁLCULO DA INFORMAÇÃO ACESSÍVEL

Michael Ferreira de Souza

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DE FORMAÇÃO DE RECURSOS HUMANOS DO LABORATÓRIO NACIONAL DE COMPUTAÇÃO CIENTÍFICA COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM MODELAGEM COMPUTACIONAL

Aprovada por:

Carlile Campos Lavor, D.Sc. - UNICAMP

Renato Portugal, D.Sc. - LNCC

Nelson Maculan Filho, D.Sc. - UFRJ

Gilson Giraldi, D.Sc. - LNCC

PETRÓPOLIS, RJ - BRASIL

MARÓCO DE 2007

SOUZA, MICHAEL

Uma nova metodologia para o cálculo da informação acessível [Petropólis] 2007

XXIII, num.páginas p. 29,7 cm (LNCC/MCT, M.Sc., Modelagem Computacional, 2007)

Dissertação - Laboratório Nacional de Computação Científica, LNCC/MCT

1. Informação Acessível, Branch and Bound, Aritmética Intervalar

I. LNCC/MCT II. Título (série)

"E o que aumenta em ciência aumenta em trabalho." (Eclesiastes 1:18b)

Para meus pais, irmãos, esposa e amigos.

Agradecimentos

Agradeço ...

A Deus por ter me dado forças e a oportunidade de realizar este trabalho.

A meus pais pelo amor gratuito.

A meus irmãos por tornarem a minha vida maus dinâmica.

À minha esposa pela paciência e companheirismo durante esta jornada.

A todos os professores que ao longo da minha vida têm contribuído para minha formação, em especial, os professores Carlile C. Lavore Renato Portugal pela orientação e amizade.

E, por último apenas por uma limitação da língua, às amizades que o mestrado tornou possíveis.

Resumo da Dissertação apresentada ao LNCC/MCT como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

UMA NOVA METODOLOGIA PARA O CÁLCULO DA INFORMAÇÃO ACESSÍVEL

Michael Ferreira de Souza

Marćo , 2007

Orientador(es): Carlile Campos Lavor, D.Sc. - UNICAMP

Renato Portugal, D.Sc. - LNCC

Programa: Modelagem Computacional

O uso de sistemas quânticos como parte de sistemas de comunicação tem sido fonte de interessantes problemas muitos ainda sem solução. No presente trabalho, apresentamos os conceitos básicos em teoria da informação e mecânica quântica necessários ao entendimento do problema do cálculo da informação acessível, cuja solução maximiza a informação mútua de Shannon para um canal definido por um ensemble de estados quânticos dados **a priori**. Propomos o uso do método de otimização global **Branch and Bound** aliado à aritmética intervalar para a estimação de limites mais precisos que os teóricos disponíveis para a informação acessível. Experimentos numéricos e resultados relacionados são apresentados.

Abstract of Dissertation presented to LNCC/MCT as a partial fulfillment of the requirements for the degree of Master of Sciences (M.Sc.)

A NEW APPROACH TO CALCULATE THE ACCESSIBLE INFORMATION

Michael Ferreira de Souza

March , 2007

Advisor(s): Carlile Campos Lavor, D.Sc. - UNICAMP

Renato Portugal, D.Sc. - LNCC

Program: Computational Modeling

The use of quantum systems as part of the communication systems has been source of interesting problems many without solution. In the present work, we show the basic concepts of information theory and quantum mechanics necessary to understand the accessible information problem, whose solution maximizes the Shannon mutual information for a channel defined by an ensemble of quantum states given **a priori**. In order to estimate more precise bounds for accessible information, we propose the use of Branch and Bound method with interval arithmetic. Numerical experiments and related results are exhibited.

Sumário

Lista de Figuras

Figura

Lista de Tabelas

Tabela

Capítulo 1

Introdução

Para muitos, a informação é o conceito fundamental de nosso tempo: vivemos a era da informação. O desenvolvimento da computação nos moldes propostos por Alan Turing solidificaram o bit como o elemento fundamental da informação, sua materialização. Boa parte dos créditos pelo nível atual do conhecimento disponível sobre as possibilidades de manipulação da informação são devidos a C. E. Shannon. Seus trabalhos deram origem a teoria da informação e, talvez, o mais importante destes seja o já célebre artigo [Shanon (1948)], pois nele Shannon respondeu questões fundamentais como a determinação da quantidade mínima de recursos físicos necessários ao armazenamento da informação e a formulação de critérios objetivos, dados por entropias, que permitem identificar possíveis relações entre diferentes fontes de informação.

O avanço científico tem conduzido a exploração de novos paradigmas de computação e processamento da informação. A década de 1920 é marcada pelo surgimento de uma das teorias físicas indispensáveis aos nossos dias, a mecânica quântica. Muitos trabalhos sobre o uso de sistemas quânticos no processamento da informação têm sido publicados, pois o paradigma de computação baseado nas propriedades de sistemas quânticos traz novas possibilidades para o tratamento da informação através da inserção de um novo elemento de informação: o bit quântico ou qbit.

O qbit é um recurso físico tangível que, diferentemente de seu análogo clássico,

o bit, tem a possibilidade de estar em uma **superposição** de estados, na verdade, uma combinação linear de estados fundamentais. Uma outra intrigante particularidade do qbit é o **emaranhamento**. Graças a ele, um sistema quântico formado por dois ou mais qbits pode assumir um estado que não é um simples produto dos estados de suas partes. A superposição e o emaranhamento, ao mesmo tempo que expandem as possibilidades do processamento e comunicação da informação, trazem consigo desafios sem paralelos clássicos. Na verdade, vimos surgir nas últimas décadas um novo campo de pesquisa: **a teoria da informação quântica**. A teoria da informação quântica analisa as possibilidades dos sistemas regidos pelas leis da mecânica quântica serem usados como sistemas de informação. As possibilidades dos sistemas quânticos de processamento da informação não são completamente conhecidas. No entanto, trabalhos recentes demonstram que estes sistemas possuem recursos completamente distintos dos seus pares clássicos e, em alguns casos, esses recursos são até superiores [Nielsen & Chuang (2003)].

As diferenças entre os sistemas de processamento da informação clássico e quântico são devidas à disparidade entre as propriedades dos seus ítens de informação: o bit (ou bit clássico) e o qbit (ou bit quântico). Uma destas disparidades diz respeito a capacidade de distinguir diferentes estados de um sistema de informação. No caso clássico, não há qualquer característica intrínseca aos sistemas que impeça esta distinção. Por outro lado, os postulados da mecânica quântica e, mais ainda, os experimentos quânticos mostram que diferentes estados de um sistema quântico podem não ser perfeitamente distinguíveis.

Muitas pesquisas têm sido feitas com intuito de quantificar quão bem se pode distinguir estados quânticos arbitrários. O problema da distinção de estados quânticos admite muitas formulações e cada uma delas dá origem a diferentes critérios qualitativos sobre a capacidade de distinção [Fuchs (1995)]. Um dos critérios mais utilizados é o da **informação acessível**.

A informação acessível especifica a quantidade máxima de informação que uma única medição pode fornecer sobre o estado de um sistema. Por simplicidade,

o estado é aleatoriamente escolhido em um conjunto pré-fixado.

O problema do cálculo da informação acessível envolve a determinação do máximo de uma função objetivo denominada **informação mútua**. Grosso modo, a informação mútua é uma medida da quantidade de informação que duas variáveis aleatórias compartilham. No caso da informação acessível, a informação mútua é utilizada para avaliar a relação existente entre a variável que representa o estado do sistema e a que representa a inferência sobre este estado.

Duas das dificuldades encontradas na avaliação da capacidade de distinção dos estados quânticos, a partir da informação acessível, são: a não-linearidade da função objetivo (informação mútua) e a existência de muitos pontos críticos. Alguns resultados teóricos, tais como soluções para casos particulares [Davies (1978)] e a fixação de limites ([Holevo (1973)], [Fuchs (1995)]), foram alcançados. Contudo, o problema geral continua sem solução. Neste cenário, uma alternativa natural é o uso de métodos numéricos de otimização para o cálculo da informação acessível.

No presente trabalho, utilizaremos o bem conhecido método numérico de otimização **branch and bound** (BB) para explorar o problema do cálculo da informação acessível. O BB, por ser um método determinístico de otimização, i.e., por garantidamente encontrar todas as soluções de problemas de otimização em tempo finito, para uma tolerância previamente especificada, oferece recursos convenientes à transposição das dificuldades inerentes ao problema.

As diferentes versões do método BB destacam-se umas das outras pelos critérios utilizados na divisão do espaço de busca (**branch**) e pelas técnicas aplicadas à estimação de cotas para a função objetivo (**bound**). Empregaremos nos experimentos numéricos deste trabalho a **aritmética intervalar** para a geração de limites válidos para a função objetivo.

Inicialmente desenvolvida por Moore [Moore (1962)] para o tratamento de **erros de arredondamento**, a **aritmética intervalar** passou a ser utilizada em problemas de otimização [Robinson (1973)] graças à sua capacidade de estimação da imagem de funções. Mais precisamente, a aritmética intervalar se constitui como

um conjunto de técnicas que fornecem intervalos, preferencialmente os menores, que contêm a imagem exata das funções.

No capítulo 2, apresentaremos o conceito de entropia e seu significado através de uma abordagem pragmática. Além disso, faremos uma revisão de instrumentos da teoria da informação clássica que permitem avaliar possíveis correlações entre variáveis aleatórias distintas. A mecânica e a informação quânticas, que são a base para a definição do problema do cálculo da informação acessível, serão abordadas no capítulo 3. Ainda no capítulo 3, enunciaremos o problema central deste trabalho: o cálculo da informação acessível. No capítulo 4, definiremos o método numérico que será utilizado e sua fundamentação teórica. Os experimentos numéricos e seus resultados serão apresentados no capítulo 5.

Capítulo 2

Teoria da informação clássica

A teoria moderna da informação clássica é marcada pelo célebre artigo de Shannon [Shanon (1948)] publicado em 1948. Nele, Shannon faz três grandes contribuições para a teoria da informação: a formulação matemática do conceito de fonte de informação clássica e os teoremas de codificação da fonte e codificação do canal. O desenvolvimento da teoria da informação clássica tem permitido um aprofundamento do conhecimento sobre as variáveis aleatórias, mais especificamente, sobre possíveis relações entre suas distribuições de probabilidade. Neste capítulo, apresentaremos alguns resultados da teoria da informação clássica que nos permitirão na seção 3.2 abordar o problema da informação acessível com mais propriedade.

O problema central da teoria da informação é reproduzir exata ou aproximadamente em um ponto uma informação selecionada em outro ponto.

A figura 2.1 é um modelo esquemático de um sistema de informação genérico.

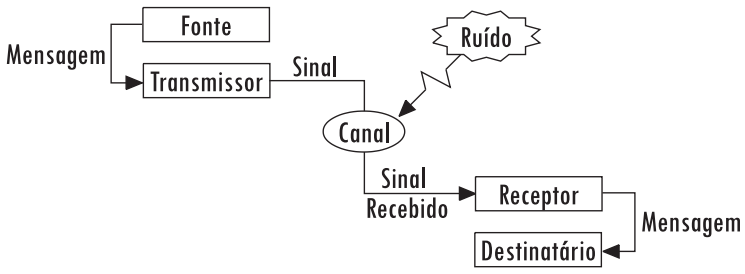

Figura 2.1: Diagrama esquemático de sistema de informação genérico.

Nesta figura, representamos uma fonte de informação que produz mensagens que serão enviadas a um destinatário através de um canal. Este simples modelo é geral o bastante para abarcar os casos particulares de comunicação via telefone ou de armazenamento de dados em um CD para consulta futura.

Definição 1 (Fonte de informação). *No caso mais simples, uma fonte de informação pode ser definida como uma seqüência* $X_1, X_2, \ldots, X_n$ *de variáveis aleatórias e uma mensagem como uma seqüência* $x_1, x_2, \ldots, x_n$ *formada por seus possíveis valores.*

Uma fonte de informação formada por variáveis aleatórias X_i **independentes** e **igualmente distribuídas** é denominada fonte de informação **i.i.d.**, formalmente temos:

Definição 2 (Fonte de informação i.i.d.). *A fonte* $X_1, X_2, \ldots, X_n$ *é* ***i.i.d.*** *se, e somente se,* $\mathrm{p}(x_1, x_2, ..., x_n) = \mathrm{p}(x_1)\mathrm{p}(x_2)...\mathrm{p}(x_n)$ *para toda seqüência* $x_1, x_2, ..., x_n$ *de valores da fonte e* $\mathrm{p}(X_i \leq x) = \mathrm{p}(X_j \leq x)$ *para todos* $i, j \in \{1, 2, \ldots, n\}$.

Fontes de informação i.i.d. são simplificações de fontes reais de informação. As regras e os processos de formação de palavras de linguagens reais fazem com que determinadas letras (símbolos) apareçam juntas com alta freqüência, i.e., não ocorre de fato a independência de usos da fonte. Por exemplo, na língua portuguesa a letra "q"aparece freqüentemente seguida da letra "u". No entanto, para fins práticos o modelo de fontes i.i.d. produz resultados satisfatórios que podem ser generalizados para modelos fontes de informação mais sofisticados.

A definição dada por Shannon para fonte de informação não é a única existente, mas é sem dúvida a mais frutífera [Nielsen & Chuang (2003)] e deixa claro o papel central das variáveis aleatórias na teoria da informação.

A determinação da quantidade mínima de recursos físicos necessários ao armazenamento da informação é uma questão de importância fundamental para qualquer implementação física de sistemas de informação. Neste contexto, definimos a entropia de Shannon para uma variável aleatória X.

Definição 3 (Entropia de Shannon). *Seja X uma variável aleatória que assume valor x com probabilidade* $\mathrm{p}(X = x) = \mathrm{p}(x)$. *A entropia de Shannon da variável X, que será denotada por $H(X)$, é definida como*

$$H(X) \equiv -\sum_x \mathrm{p}(x) \log_2 \mathrm{p}(x). \tag{2.1}$$

A entropia de Shannon representa a quantidade média de bits (no caso em que o logaritmo da Eq. (2.1) é tomado na base 2) necessários para armazenar os possíveis resultados de uma variável aleatória. Por exemplo, imagine uma variável aleatória X que assume os valores $0, 1, 2$ e 3 com as probabilidades $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}$ e $\frac{1}{8}$, respectivamente. Se utilizarmos um código de compressão que associe os resultados $0, 1, 2$ e 3 às seqüências $0, 10, 100$ e 101, então, em média, utilizaremos $\frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 + \frac{1}{8}\cdot 3 + \frac{1}{8}\cdot 3 = 7/4$ bits por uso da fonte, ou seja, exatamente o valor de $H(X) = 7/4$. O termo código de compressão é coerente, pois numa abordagem ingênua, usaríamos 2 bits para armazenar os 4 valores possíveis, ao invés dos $7/4$ bits obtidos pela codificação sugerida. Esta interpretação é formalmente validada pelo teorema de Shannon para a **codificação da fonte** [Shanon (1948)].

Teorema 1 (Teorema de codificação da fonte). *Seja $X_1, X_2, \ldots, X_n$ uma fonte informação **i.i.d.**. Então as n saídas de $X_1, X_2, \ldots, X_n$ podem ser comprimidas em*

$$nH(X) + o(n) \ \textit{bits} \tag{2.2}$$

e restauradas ao original com probabilidade que se aproxima de 1, à medida que n tende a ∞.

Geralmente, interpretamos a probabilidade de um evento de acordo com a **teoria da freqüência**, onde a probabilidade é uma informação quantitativa da ocorrência de um evento, quando o experimento é realizado um grande número de vezes. Alternativamente, a idéia por trás do resultado de Shannon fundamenta-se na **teoria da tendência** onde a probabilidade de um evento é interpretada como sendo uma **informação** sobre parte do resultado de um único experimento.

A demonstração do teorema de codificação da fonte baseia-se no argumento de que apenas um subconjunto próprio do conjunto de todas as possíveis seqüências $x_1, \ldots, x_n$ formadas pelas n saídas da fonte $X_1, X_2, \ldots, X_n$ precisa de fato ser codificado em bits, este subconjunto é chamado subconjunto típico. A idéia é que, para n suficientemente grande, a probabilidade de o uso da fonte gerar uma seqüência $x_1, \ldots, x_n$ que não pertença a este subconjunto típico é aproximadamente nula.

O teorema de codificação da fonte determina a quantidade de redundância que pode ser eliminada sem que se comprometa o significado da mensagem.

A entropia de Shannon é também definida para um par (X, Y) de variáveis aleatórias e, neste caso, recebe o nome de entropia conjunta.

Definição 4 (Entropia Conjunta). *Sejam X e Y variáveis aleatórias e* $\mathrm{p}(X = x, Y = y) = \mathrm{p}_{x,y}$. *A entropia conjunta do par (X, Y), denotada por $H(X, Y)$, é definida como*

$$H(X, Y) \equiv -\sum_{x,y} \mathrm{p}_{x,y} \log_2 \mathrm{p}_{x,y}. \tag{2.3}$$

Naturalmente, a interpretação da entropia conjunta é análoga à da entropia de Shannon, ou seja, a quantidade média de bits necessários ao armazenamento dos possíveis valores do par (X, Y) ou uma medida da ignorância com relação ao seu valor.

Por simplicidade, consideraremos a entrada do canal de um sistema de informação como sendo mensagens $x \in \{x_1, x_2, ..., x_n\}$ produzidas em cada uso da fonte X. No caso mais geral, o canal sofre alguma forma de ruído e uma entrada x pode ser alterada para uma saída $y \neq x$. Portanto, a saída do canal é geralmente uma variável aleatória $Y \neq X$ que assume valores y com probabilidades p_y dependentes tanto da entrada x, quanto do tipo de ruído existente no canal (ver figura 2.2). É essencial determinar a capacidade que o destinatário possui de reconstruir, a partir da saída y do canal, sua entrada x. Ou seja, determinar a quantidade de informação que as variáveis X e Y possuem em comum, compartilham. Com este propósito, definimos a informação mútua entre duas variáveis aleatórias.

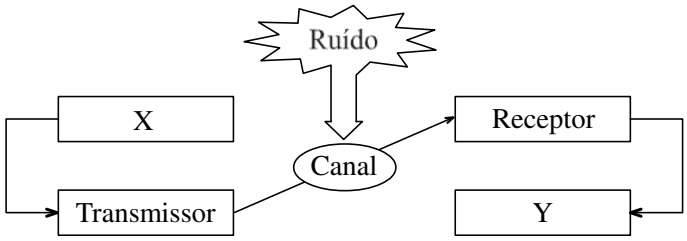

Figura 2.2: Na presença de ruído, a variável Y de saída do canal é diferente da variável de entrada X.

Definição 5 (Informação Mútua). *Sejam X e Y duas variáveis aleatórias. A informação mútua entre X e Y, denotada por $H(X:Y)$, é definida como*

$$H(X:Y) \equiv H(X) + H(Y) - H(X,Y). \tag{2.4}$$

Se denotarmos $P(X=x)$, $P(Y=y)$, $P(X=x,Y=y)$ e $P(Y=y|X=x)$, respectivamente, por p_x, p_y, $\mathrm{p}_{y,x}$ e $\mathrm{p}_{y|x}$, podemos obter uma útil expressão para a informação mútua a partir da Eq. (2.4). De fato,

$$\begin{aligned}
H(X:Y) &= H(X) + H(Y) - H(X,Y) && (2.5)\\
&= -\sum_x \mathrm{p}_x \log_2 \mathrm{p}_x - \sum_y \mathrm{p}_y \log_2 \mathrm{p}_y + \sum_{x,y} \mathrm{p}_{y,x} \log_2 \mathrm{p}_{y,x} && (2.6)\\
&= -\sum_{x,y} \mathrm{p}_{y,x} \log_2 \mathrm{p}_x - \sum_{y,x} \mathrm{p}_{y,x} \log_2 \mathrm{p}_y + \sum_{x,y} \mathrm{p}_{y,x} \log_2 \mathrm{p}_{y,x} && (2.7)\\
&= \sum_{x,y} \mathrm{p}_{y,x} \log_2 \frac{\mathrm{p}_{y,x}}{\mathrm{p}_x \mathrm{p}_y} && (2.8)\\
&= \sum_{x,y} \mathrm{p}_x \mathrm{p}_{y|x} \log_2 \frac{\mathrm{p}_x \mathrm{p}_{y|x}}{\mathrm{p}_x \mathrm{p}_y} && (2.9)
\end{aligned}$$

Segue que

$$H(X:Y) = \sum_{x,y} \mathrm{p}_x \mathrm{p}_{y|x} \log_2 \frac{\mathrm{p}_{y|x}}{\mathrm{p}_y}. \tag{2.10}$$

Um conceito importante que nos auxiliará a dar um significado concreto a informação mútua é o conceito de **cadeia de Markov**.

Definição 6. *Uma cadeia markoviana é uma seqüência de variáveis aleatórias*

$X_1 \to X_2 \to \ldots$, *tal que* X_{n+1} *depende apenas de* X_n. *Mais formalmente,*

$$\mathrm{p}(X_{n+1} = x_{n+1} | X_n = x_n, \ldots, X_1 = x_1) = \mathrm{p}(X_{n+1} = x_{n+1} | X_n = x_n). \tag{2.11}$$

Se definirmos as variáveis aleatórias X e Y como sendo respectivamente a entrada e a saída do canal, e a variável Z como sendo a tentativa de reconstruir X a partir de Y, então teremos uma cadeia markoviana $X \to Y \to Z$.

A **desigualdade de processamento de dados**, apresentada a seguir, permite atribuir um significado útil a informação mútua e, de forma objetiva, evidencia a conexão entre a definição da informação mútua e capacidade de reconstrução da mensagem.

Teorema 2 (Desigualdade do processamento de dados). *Suponha que* $X \to Y \to Z$ *seja uma cadeia markoviana. Resulta que*

$$H(X) \geq H(X:Y) \geq H(X:Z). \tag{2.12}$$

Além disso, a primeira desigualdade é saturada se, e somente se, dado Y*, for possível reconstruir* X*.*

Como exemplo, imagine uma fonte binária eqüiprovável X, i.e., uma fonte de informação que assume cada um dos valores 0 ou 1 com probabilidade $1/2$ e, além disso, um canal bit-flip $\mathcal{C}_\mathrm{p} : \{0,1\} \to \{0,1\}$ que apresenta probabilidade $\mathrm{p} \in [0,1]$ de produzir para uma dada entrada x uma saída $\mathcal{C}_\mathrm{p}(x) = \neg x$ (ver figura 2.3).

Assumindo que Y é a variável de saída do canal $\mathcal{C}_\mathrm{p}$, podemos representar $\mathcal{C}_\mathrm{p}$ pela matriz de probabilidades condicionais M_p dada por

$$M_\mathrm{p} = \begin{pmatrix} \mathrm{p}(Y=0|X=0) & \mathrm{p}(Y=0|X=1) \\ \mathrm{p}(Y=1|X=0) & \mathrm{p}(Y=1|X=1) \end{pmatrix} = \begin{pmatrix} 1-\mathrm{p} & \mathrm{p} \\ 1-\mathrm{p} & \mathrm{p} \end{pmatrix}. \tag{2.13}$$

Denote por Z nossa tentativa de determinar, a partir da saída Y, o estado da variável de entrada X. A estratégia de reconstrução será simplesmente tomar

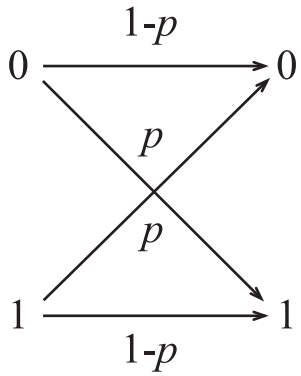

Figura 2.3: O canal bit-flip $\mathcal{C}_\mathrm{p}$ possui probabilidade p de produzir para uma entrada x uma saída $\mathcal{C}_\mathrm{p}(x) = \neg x$.

$Z = Y$, quando $\mathrm{p} \leq 1/2$, e $Z = \neg Y$, quando $\mathrm{p} > 1/2$. Neste caso, assumindo que $\mathrm{p}(X = x) = \mathrm{p}_x$, $\mathrm{p}(Z = z) = \mathrm{p}_z$ e $\mathrm{p}(Z = z|X = x) = \mathrm{p}_{z|x}$, a entropia da variável de entrada X é dada por

$$H(X) = -\sum_{x=0}^{1} \mathrm{p}_x \log_2 \mathrm{p}_x = -[0.5 \log_2(0.5) + 0.5 \log_2(0.5)] = 1 \tag{2.14}$$

e a informação mútua das variáveis X e Z por

$$\begin{aligned} H(X:Z) &= \sum_{x,z=0}^{1} \mathrm{p}_x \mathrm{p}_{z|x} \log_2 \frac{\mathrm{p}_{z|x}}{\mathrm{p}_z} && (2.15)\\ &= \sum_{z=0}^{1} 0.5(1-\mathrm{p}) \log_2 \frac{1-\mathrm{p}}{0.5} + 0.5\mathrm{p} \log_2 \frac{\mathrm{p}}{0.5} && (2.16)\\ &= (1-\mathrm{p}) \log_2(2(1-p)) + \mathrm{p} \log_2(2\mathrm{p}) && (2.17)\\ &= 1 + (1-p) \log_2(1-p) + p \log_2 \mathrm{p}. && (2.18) \end{aligned}$$

Na figura 2.4 confrontamos os valores de $H(X)$ e $H(X : Z)$. Como era intuitivamente esperado, quanto mais próximos de 0 estão os valores p ou $1 - \mathrm{p}$, mais próximos estão os valores de $H(X : Z)$ e $H(X)$, pois maiores são as chances de que não tenha ocorrido erro (flip) no envio do bit de entrada (no caso em que $\mathrm{p} \approx 0$) ou que o erro possa ser de fato corrigido (no caso em que $1 - \mathrm{p} \approx 0$). Note que o valor mínimo de $H(X : Z)$ ocorre quando $\mathrm{p} = 1/2$, já que, para este valor de p, o canal destrói toda a informação sobre o bit de entrada, equivalendo seu uso

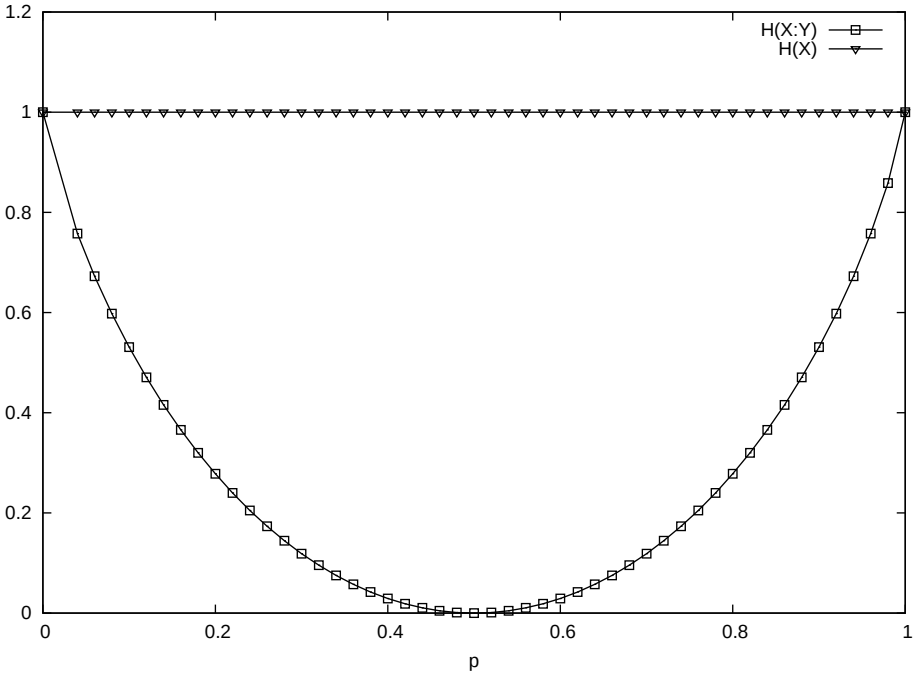

Figura 2.4: Os valores de $H(X)$ e $H(X : Z)$ aproximam-se à medida que p ou $1 - \text{p}$ aproximam-se de 0.

ao lançamento de uma moeda ideal.

O "diagrama de Venn da entropia"resume bem a relação entre as entropias de Shannon das variáveis X e Y e a informação mútua $H(X : Y)$ (ver figura 2.5).

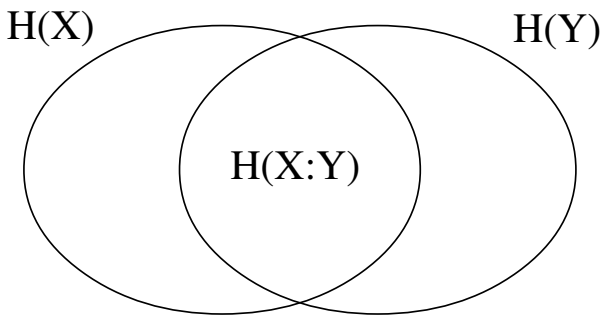

Figura 2.5: Relação entre as entropias de Shannon das variáveis X e Y e a informação mútua $H(X : Y)$.

No presente capítulo, apresentamos através de conceitos básicos da teoria da informação clássica um conjunto de critérios que permitem avaliar possíveis correlações entre variáveis aleatórias distintas. Estes critérios, dados por entropias, serão particularmente úteis na formulação do problema do cálculo da informação acessível que faremos no capítulo 3.

Capítulo 3

Teoria da informação quântica

3.1 Mecânica quântica

Nesta seção, definiremos precisamente os sistemas físicos que dão origem ao problema do cálculo da informação acessível, i.e., os sistemas quânticos. Existem diferentes conjuntos de postulados para a mecânica quântica. Nesta dissertação, adotaremos uma formulação puramente matemática dada por operadores densidade e citaremos somente os postulados relacionados ao problema do cálculo da informação acessível.

A notação padrão utilizada em mecânica quântica para a representação dos conceitos básicos da álgebra linear que serão utilizados no presente trabalho é exibida na tabela 3.1. Esta representação é chamada de notação de **Dirac** [Nielsen & Chuang (2003)].

Notação	**Descrição**
z^*	Conjugado do complexo z
$\|\psi\rangle$	Vetor. Também chamado de **ket**
$\langle\psi\|$	Vetor dual de $\|\psi\rangle$. Também chamado **bra**
$\langle\varphi\|\psi\rangle$	Produto escalar entre $\|\varphi\rangle$ e $\|\psi\rangle$
$\|\varphi\rangle\langle\psi\|$	Produto matricial entre $\langle\varphi\|$ e $\|\psi\rangle$
A^*	Complexo conjugado da matriz A
A^T	Transposta da matriz A
$A^\dagger$	Conjugado hermitiano, ou matriz adjunta de A, $A^\dagger = (A^T)^*$
$\langle\varphi\| A \|\psi\rangle$	Produto escalar entre $\|\varphi\rangle$ e $A\|\psi\rangle$

Tabela 3.1: Resumo da notação padrão utilizada em mecânica quântica para conceitos de álgebra linear.

Postulado 1. *Associado a qualquer sistema físico, existe um espaço de Hilbert conhecido como espaço de estados do sistema. O sistema é completamente descrito pelo seu* ***estado*** ρ*, um operador positivo com traço unitário definido no espaço de estados do sistema.*

Os estados de posto unitário, i.e., $\rho = |\psi\rangle\langle\psi|$, formam uma classe muito importante de estados e são freqüentemente denominados **estados puros** em oposição aos estados de maior posto, geralmente denominados **estados mistos**.

Um conceito fundamental em teoria da informação quântica é a de medidas quânticas. Grosso modo, uma **medida quântica** é qualquer processo físico que possa ser usado sobre um sistema quântico para gerar uma distribuição de probabilidades para um conjunto de saídas (resultados da medida). As medidas quânticas são formalmente definidas pelo seguinte postulado:

Postulado 2. *Medidas quânticas são descritas por uma coleção de operadores de medidas* $\{M_m\}$*. Esses operadores atuam sobre o espaço de estados do sistema a ser medido. O índice* m *refere-se a um resultado possível da medida. Se o estado do sistema imediatamente antes da medida for* ρ_i*, a probabilidade do resultado* m *ocorrer será*

$$\mathrm{p}(m|i) = \mathrm{tr}(M_m^\dagger M_m \rho_i) \tag{3.1}$$

e o estado após a medida será

$$\frac{M_m \rho_i M_m^\dagger}{\mathrm{tr}(M_m^\dagger M_m \rho_i)}. \tag{3.2}$$

Os operadores de medida satisfazem a seguinte equação de completude

$$\sum_m M_m^\dagger M_m = I. \tag{3.3}$$

Pela Eq. (3.1) do postulado 2, vemos que os operadores $M_m^\dagger M_m$ são, para a distribuição de probabilidades dos resultados m, mais relevantes que os operadores M_m. Portanto, quando nosso interesse está restrito aos resultados da medida e sua

distribuição de probabilidades, é suficiente o conhecimento dos operadores $E_m = M_m^\dagger M_m$, os quais são positivos e, pela definição dos operadores M_m, satisfazem a relação de completude. É com esta motivação que postulamos:

Postulado 3 (Medidas POVM). *Um POVM é qualquer conjunto $\{E_m\}$ formado por operadores positivos que satisfazem a relação de completude, i.e., $\sum_m E_m = I$. Se o estado do sistema imediatamente antes da medida for ρ_i, então, na medida descrita pelo POVM $\{E_m\}$, a probabilidade de obtermos um resultado m é* $\mathrm{p}(m|i) = \mathrm{tr}(E_m \rho_i)$.

Muitos dos problemas propostos pela teoria da informação quântica baseiam-se em sistemas quânticos cujos estados não são inteiramente conhecidos. Geralmente, o que se tem é uma distribuição de probabilidades associada aos possíveis estados do sistema. Se um sistema tem probabilidade $\mathrm{p}(i)$ de estar no estado ρ_i, então podemos representá-lo pelo conjunto $\{\rho_i, \mathrm{p}(i)\}$. O conjunto $\{\rho_i, \mathrm{p}(i)\}$ formado pelos possíveis estados de um sistema quântico e suas respectivas probabilidades recebe o nome de **ensemble de estados do sistema**. A todo sistema com ensemble $\{\rho_i, \mathrm{p}(i)\}$ está associado um operador ρ denominado **operador densidade** dado por

$$\rho \equiv \sum_i \mathrm{p}(i)\rho_i. \tag{3.4}$$

Os operadores densidade são particularmente úteis no cálculo das probabilidades dos resultados de uma medida, pois se um sistema possui ensemble $\{\rho_i, \mathrm{p}(i)\}$ e operador densidade ρ e, além disso, utilizarmos um conjunto de operadores de medida POVM $\{E_m\}$, então a probabilidade de obtermos um resultado m é dada

por

$$\begin{aligned} \mathrm{p}(m) &= \sum_i \mathrm{p}(i,m) && (3.5)\\ &= \sum_i \mathrm{p}(i)\ \mathrm{p}(m|i) && (3.6)\\ &= \sum_i \mathrm{p}(i)\ \mathrm{tr}(E_m\rho_i) && (3.7)\\ &= \mathrm{tr}(E_m \sum_i \mathrm{p}(i)\rho_i) && (3.8)\\ &= \mathrm{tr}(E_m\rho) && (3.9) \end{aligned}$$

Uma conseqüência deste resultado é que sistemas com o mesmo operador densidade possuem a mesma distribuição de probabilidades associada aos resultados das medidas.

3.2 Informação quântica

A introdução de novos tipos de informação, como os estados quânticos, amplia as possibilidades dos processos dinâmicos de tratamento da informação (compressão, descompressão, codificação, decodificação, transmissão, código de correção de erros) e torna a **teoria da informação quântica** mais rica do que a teoria da informação clássica.

Uma diferença fundamental entre a teoria da informação clássica e a teoria da informação quântica é a capacidade de distinção de diferentes itens de informação. A princípio, somos sempre capazes de distinguir os diferentes símbolos de um alfabeto clássico. É claro que algumas vezes, na prática, distinguir um "a" mau grafado de um "o" pode ser um trabalho árduo, mas não existe nenhuma característica inerente à informação clássica que impeça a distinção. Por outro lado, a informação quântica, i.e., os estados quânticos não são sempre distinguíveis. Isto fica evidente pelo seguinte resultado:

Proposição 1. *Estados não-ortogonais não podem ser distinguidos com certeza.*

Demonstração. Provaremos por absurdo que se ρ_0 e ρ_1 são estados não-ortogonais,

i.e., $\mathrm{tr}(\rho_0\rho_1^\dagger) \neq 0$, então não existe nenhuma medida capaz de distingui-los com certeza. Suponha que tal medida seja possível. Como estamos interessados apenas no resultado da medida, os operadores POVM serão suficientes. O conjunto POVM $\{E_0, E_1\}$ é capaz de distinguir com certeza ρ_0 e ρ_1 se, e somente se,

$$\mathrm{tr}(E_0\rho_0) = 1 \text{ e } \mathrm{tr}(E_1\rho_1) = 1, \tag{3.10}$$

pois neste caso, a partir do resultado 0 ou 1 da medida, podemos concluir com certeza que, antes da medida, o estado do sistema era ρ_0 ou ρ_1, respectivamente.

Como $E_i \geq 0$ e $\sum_i E_i = I$, segue que $\mathrm{tr}(\sum E_i\rho_0) = \mathrm{tr}(\rho_0) = 1$ (Ver postulado 1); e como $\mathrm{tr}(E_0\rho_0) = 1$, devemos ter $\mathrm{tr}(E_1\rho_0) = 0$. Suponha que se faça a decomposição $\sqrt{\rho_1} = \alpha\sqrt{\rho_0} + \beta\sqrt{\rho}$, em que $\sqrt{\rho}$ é ortonormal a $\sqrt{\rho_0}$, $\left\|\sqrt{\rho_1}\right\|^2 = \mathrm{tr}(\rho_1) = |\alpha|^2 + |\beta|^2 = 1$, e $|\beta| < 1$, já que ρ_0 e ρ_1 são não-ortogonais. Então, $\mathrm{tr}(E_1\sqrt{\rho_1}) = \beta\mathrm{tr}(E_1\sqrt{\rho})$, o que contradiz a Eq. (3.10), pois

$$\mathrm{tr}(E_1\rho_1) = \mathrm{tr}(E_1(|\alpha|^2\rho_0 + |\beta|^2\sqrt{\rho}\sqrt{\rho}^\dagger)) = |\beta|^2\,\mathrm{tr}(E_1\sqrt{\rho}\sqrt{\rho}^\dagger) \leq |\beta|^2 < 1, \tag{3.11}$$

em que a primeira desigualdade segue da observação de que

$$\mathrm{tr}(E_1\sqrt{\rho}\sqrt{\rho}^\dagger) \leq \mathrm{tr}(\sum_i E_i\sqrt{\rho}\sqrt{\rho}^\dagger) = \mathrm{tr}(\sqrt{\rho}\sqrt{\rho}^\dagger) = \|\sqrt{\rho}\|^2 = 1. \tag{3.12}$$

□

A proposição 1 evidencia a impossibilidade de distinguir com certeza estados quânticos arbitrários, mas não nos impede de distingui-los desde que aceitemos alguma incerteza na decisão.

Outra diferença entre a informação clássica e a informação quântica que interfere na distinção de estados quânticos é a impossibilidade de duplicação de ítens arbitrários de informação, este resultado é conhecido como **teorema da não-clonagem** ([Dieks (1982)], [Wootters & Zurek (1982)]). Grosso modo, o teorema da não-clonagem estabelece que não é possível construir um aparato físico capaz de

clonar estados quânticos arbitrários. Se a construção de tal aparato fosse possível, então uma estratégia para a distinção de estados quânticos seria gerar um número arbitrariamente grande de cópias dos estados que se deseja distinguir e, usando em cada medição diferentes operadores de medidas, caracterizá-los completamente.

O problema de distinguir estados quânticos arbitrários pode ser formulado de diversas maneiras, cada uma utilizando diferentes critérios qualitativos para a capacidade de distinção dos estados. Alguns dos possíveis critérios são: a probabilidade de erro quântico, a fidelidade quântica, "*quantum Rényi Overlaps*", a informação quântica de Kullback e a informação acessível [Fuchs (1995)]. O critério da **informação acessível** é um dos mais utilizados e pode ser introduzido através de um jogo envolvendo dois parceiros, Alice e Bob.

Sejam X, uma fonte de informação, i.e., uma variável aleatória que assume valores no conjunto $\{x : x = 0, 1, \ldots, n\}$ com distribuição de probabilidades $\mathrm{p}(X = x) = \mathrm{p}_x$, e $\{\rho_0, \rho_1, \ldots, \rho_n\}$, um conjunto de estados quânticos. Suponha que Alice, baseada no resultado x da variável X, prepara um determinado estado quântico ρ_x selecionado no conjunto $\{\rho_0, \ldots, \rho_n\}$ e o envia a Bob. O objetivo de Bob é determinar o valor de X da melhor forma possível. Para isso, Bob faz uma medida quântica no estado por ele recebido e tenta adivinhar, com base no resultado Y da medida, qual foi o estado dado a ele e, conseqüentemente, o valor da variável X (veja figura 3.1). Bob tem conhecimento tanto da distribuição de probabilidades da variável aleatória X quanto do conjunto de estados quânticos $\{\rho_x\}$, portanto para ele o estado do sistema é matematicamente representado por seu operador densidade $\rho = \sum_x \mathrm{p}_x \rho_x$.

Uma boa medida da informação que Bob adquiriu sobre X é a informação mútua entre as variáveis aleatórias X e Y (Ver definição 5),

$$H(X : Y) = \sum_{x,y} \mathrm{p}_x \mathrm{p}(y|x) \log_2 \frac{\mathrm{p}(y|x)}{\mathrm{p}_y} \tag{3.13}$$

O interesse de Bob é maximizar a informação mútua entre as variáveis X e Y e, assim, maximizar sua capacidade de reconstrução da entrada X a partir da saída

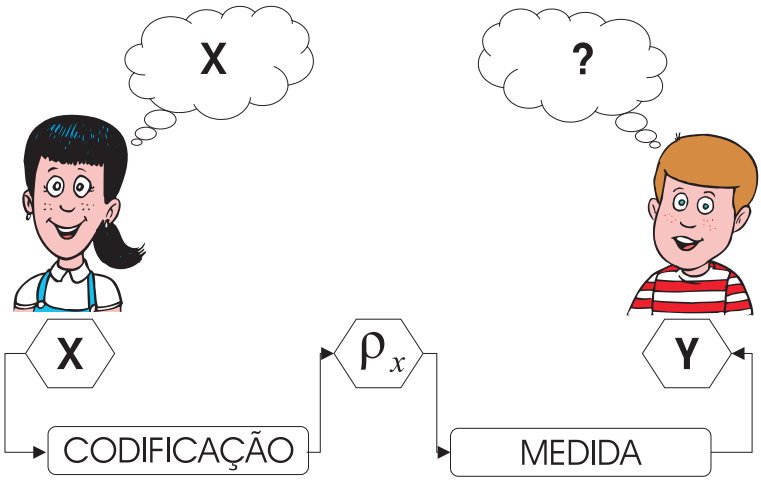

Figura 3.1: Motivação do problema do cálculo da informação acessível. A partir do resultado y de uma medida sobre o estado ρ_x, Bob deve determinar a entrada x escolhida por Alice.

Y. O formalismo de medidas POVM é suficiente para representar as diferentes estratégias (medidas) que Bob pode utilizar para maximizar sua capacidade de reconstrução do valor de X. Para uma medida definida pelo POVM $\{E_y\}$, a distribuição de probabilidade da variável resultado Y é $\mathrm{p}(Y = y) = \mathrm{p}_y = \mathrm{tr}(E_y\rho)$ e $\mathrm{p}(Y = y|X = x) = \mathrm{p}(y|x) = \mathrm{tr}(E_y\rho_x)$.

Portanto,

$$H(X : Y) = \sum_{x,y} \mathrm{p}_x \mathrm{tr}(E_y\rho_x) \log_2 \frac{\mathrm{tr}(E_y\rho_x)}{\mathrm{tr}(E_y\rho)} \tag{3.14}$$

Para tornar claro quais são os parâmetros da informação mútua dada pela Eq. (3.14), defina

$$I(\{\rho_x, \mathrm{p}_x\} : \{E_y\}) \equiv H(X : Y) = \sum_{x,y} \mathrm{p}_x \mathrm{tr}(E_y\rho_x) \log_2 \frac{\mathrm{tr}(E_y\rho_x)}{\mathrm{tr}(E_y\rho)}. \tag{3.15}$$

A **informação acessível** é definida como sendo o máximo da informação mútua $I(\{\rho_x, \mathrm{p}_x\} : \{E_y\})$ sobre todos os conjuntos POVM possíveis.

Definição 7 (Informação Acessível). *Seja $\mathcal{R} = \{\rho_x, \mathrm{p}_x\}$ um ensemble de estados quânticos. A informação acessível de $\mathcal{R}$ é a solução do seguinte problema de maximização:*

$$I_{acc}(\mathcal{R}) \equiv \max_{\{E_y\} \in \mathcal{P}_\mathcal{R}} I(\mathcal{R} : \{E_y\}), \tag{3.16}$$

onde $\mathcal{P}_{\mathcal{R}}$ é o conjunto formado por todos os POVM's cujos elementos atuam no mesmo espaço de Hilbert que os operadores de $\mathcal{R}$.

Apesar de existirem soluções para ensembles particulares baseadas em restrições sobre o ensemble $\mathcal{R}$ como, por exemplo, ensembles com estruturas algébricas [Sasaki *et al* (1999)]), no caso geral, o cálculo da informação acessível é ainda um problema aberto. Uma abordagem alternativa bastante explorada é a determinação de limites para a informação acessível [Fuchs (1995)]. Talvez os mais famosos destes sejam o limite superior para a informação acessível conhecido como limite de Holevo (conjecturado por Gordon [Gordon (1964)] e provado por Holevo [Holevo (1973)]) e o limite inferior de Jozsa-Robb-Wooters [Jozsa *et al* (1994)]. Antes de apresentá-los, definiremos a **entropia de von Neumann** e a **sub-entropia** de operadores densidade.

Definição 8 (Entropia de von Neumann). *Sejam σ um operador densidade e $\{\lambda_x\}$ o conjunto de seus autovalores. A entropia de von Neumann de σ, denotada por $S(\sigma)$, é definida por*

$$S(\sigma) \equiv -\sum_x \lambda_x \log_2 \lambda_x. \tag{3.17}$$

Definição 9 (Sub-entropia). *Sejam σ um operador densidade e $\{\lambda_x\}$ o conjunto de seus autovalores. A sub-entropia de σ, denotada por $\varphi(\sigma)$, é definida por*

$$Q(\sigma) \equiv -\sum_x \left(\prod_{y \neq x} \frac{\lambda_x}{\lambda_x - \lambda_y} \right) \lambda_x \log_2 \lambda_x. \tag{3.18}$$

De posse destas definições, podemos apresentar os limites de Holevo e Jozsa-Robb-Wooters para a informação acessível.

Sejam $\mathcal{R} = \{\rho_x, \mathrm{p}_x\}$ um ensemble com operador densidade ρ e $\mathcal{P}_{\mathcal{R}}$ o conjunto formado por todos os POVM's que atuam no mesmo espaço de Hilbert que os operadores de $\mathcal{R}$.

Teorema 3 (Limite de Holevo). *Se $\chi(\mathcal{R}) = S(\rho) - \sum_x \mathrm{p}_x S(\rho_x)$, então*

$$I(\mathcal{R} : \{E_y\}) \leq \chi(\mathcal{R}) \quad \forall \{E_y\} \in \mathcal{P}_{\mathcal{R}}. \tag{3.19}$$

Teorema 4 (Limite de Jozsa-Robb-Wooters). *Se* $\varphi(\mathcal{R}) = Q(\rho) - \sum_x \mathrm{p}_x Q(\rho_x)$, *então*

$$I(\mathcal{R} : \{E_y\}) \geq \varphi(\mathcal{R}) \quad \forall \{E_y\} \in \mathcal{P}_{\mathcal{R}}. \tag{3.20}$$

Concluímos portanto destes limites que $I_{acc}(\mathcal{R})$, a informação acessível de um ensemble $\mathcal{R}$, satisfaz necessariamente as seguintes desigualdades:

$$\varphi(\mathcal{R}) \leq I_{acc}(\mathcal{R}) \leq \chi(\mathcal{R}). \tag{3.21}$$

O desenvolvimento de métodos numéricos que viabilizem o cálculo da informação acessível são, portanto, uma alternativa consistente. Os resultados dados a seguir garantem a existência de soluções para o problema em questão e simplificam o espaço de busca destas soluções.

A existência de solução é garantida pelo seguinte resultado devido a Davies [Davies (1978)].

Teorema 5 (Davies). *Sejam* $\mathcal{H}$ *um espaço de Hilbert de dimensão finita* d *e* $\mathcal{R}$ *um ensemble em* $\mathcal{H}$. *Então a informação acessível de* $\mathcal{R}$ *é alcançada por um POVM* $\{E_y : y = 0, \ldots, n-1\}$ *tal que*

$$E_y = |\psi_y\rangle\langle\psi_y| \text{ para algum vetor } |\psi_y\rangle \text{ com } \||\psi_y\rangle\| \leq 1. \tag{3.22}$$

Além disso, a constante n *satisfaz* $d \leq n \leq d^2$.

Um resultado derivado do teorema de Davies e que se aplica a ensembles reais ajuda a restringir o conjunto de valores possíveis para a cardinalidade do POVM ótimo [Sasaki *et al* (1999)].

Corolário 1 (Davies para ensembles reais). *Seja* $\mathcal{R}$ *um ensemble de estados reais em* d *dimensões. Então a informação acessível pode ser alcançada por um POVM* $\{E_y\}$ *com* n *elementos da forma* $E_y = |\psi_y\rangle\langle\psi_y|$ *onde* $d \leq n \leq d(d+1)/2$.

Neste capítulo, expusemos as bases da mecânica quântica necessárias ao entendimento e à formulação do problema do cálculo da informação acessível. Os te-

oremas apresentados que impossibilitam a perfeita distinção de estados quânticos não-ortogonais e a clonagem de estados arbitrários são relevantes, pois impõem severas restrições a capacidade de distinção e, portanto, motivam sua quantificação. Os limites teóricos exibidos serão utilizados como parâmetros qualitativos nos experimentos numéricos do capítulo 5. Finalmente, o teorema de Davies e seu derivado real permitem uma melhor estruturação do problema de otimização definido pelo cálculo da informação acessível.

Capítulo 4

Otimização global e aritmética intervalar

4.1 Branch and bound

O método BB (**branch and bound**) é uma técnica de otimização bem conhecida que produz cotas superiores para o máximo global e gera, portanto, alguma informação sobre a qualidade dos máximos locais. O BB é um **método determinístico** de otimização, pois, para uma tolerância previamente especificada, garantidamente encontra todos os máximos globais em um tempo finito [Neumaier (2004)].

Um problema de otimização global pode ser assim descrito: dadas uma função contínua $f : \mathbb{R}^n \to \mathbb{R}$ e $S \subset \mathbb{R}^n$ a região em que buscamos o(s) ponto(s) x^* onde o valor máximo ocorre, encontre o máximo global $f^* = \max\{f(x) : x \in S\}$ e o conjunto de todos os maximizadores globais de f, $X^*(f) = \{x^* \in S : f(x^*) = f^*\}$.

Um método BB alterna entre duas etapas principais: **decomposição** (**branching**), que faz uma subdivisão recursiva do conjunto S; e **estimação** (**bounding**), que faz o cálculo de cotas inferiores e superiores para o maior valor de f numa sub-região de S.

Em cada passo desse procedimento, tem-se uma partição de S em subconjuntos S_γ ($\gamma \in \Gamma$), uma cota superior $\bar{f}(S_\gamma)$ para $\max\limits_{x \in S_\gamma} f(x)$, e uma cota inferior $\underline{f}$ para $\max\limits_{x \in S} f(x)$, representando o maior valor de f encontrado até o momento. Obviamente, subconjuntos S_γ, para os quais $\bar{f}(S_\gamma) < \underline{f}$, não podem conter um máximo global e, portanto, são descartados. Se depois de possíveis melhoramentos de $\underline{f}$,

algum subconjunto S_γ é mantido com $\bar{f}(S_\gamma) > \underline{f}$, então a partição é refinada e o procedimento se repete. Existem muitas variações desse esquema, cada uma com resultados de convergência para o máximo global sob condições diferentes ([Benson (1982)], [Horst & Tuy (1987)], [Pintér (1988)]).

Os vários métodos de otimização global diferem-se geralmente nos métodos aplicados para definir os subconjuntos S_γ e computar as cotas $\bar{f}(S_\gamma)$ e $\underline{f}$. Nas seções seguintes, utilizaremos as técnicas da aritmética intervalar para gerar estimativas (cotas) válidas para o problema do cálculo da informação acessível.

4.2 Aritmética Intervalar

Um dos problemas fundamentais na teoria e no uso dos métodos numéricos é o controle dos erros devidos à representação dos números reais em um sistema em **ponto flutuante** [Goldberg (1991)]. Em termos absolutos, a maioria dos números reais são muito grandes (**overflow**) ou muito pequenos (**underflow**) para serem representados neste sistema finito. Além disso, a maioria dos números reais estará entre dois números em ponto flutuante e um deles deverá ser escolhido para representá-los. Neste caso, o erro cometido é denominado **erro de arredondamento**.

Uma maneira simples e muito utilizada para estimar o erro em um cálculo em ponto flutuante é repetir o cálculo, usando mais precisão, e comparar os resultados. No entanto, dependendo da sensibilidade do cálculo aos erros de arredondamento, essa prática pode ser muito enganosa. A **aritmética intervalar** é um método muito mais eficiente para o controle dos erros e um passo em direção à transformação do computador em uma ferramenta matemática segura.

A aritmética intervalar (ou **análise intervalar**) teve como primeira aplicação o controle de erros de arredondamento provenientes das operações realizadas em um computador. O seu desenvolvimento moderno é marcado pela tese de doutorado de Moore [Moore (1962)].

A aritmética intervalar opera com intervalos, ao invés de números reais. Cada número real x é representado por um par de números em ponto flutuante,

$\underline{x}$ e $\bar{x}$, que definem um intervalo $\mathbf{x} = [\underline{x}, \bar{x}]$ de números reais, tal que $\underline{x} \leq x \leq \bar{x}$. Esta representação fornece uma estimativa para o valor de x, dada pelo centro do intervalo $\mathbf{x}$, e, além disso, uma medida qualitativa dessa estimativa, dada pelo tamanho do intervalo.

As funções e operações definidas sobre os números reais podem ser estendidas para os intervalos. Para qualquer função ou operação $f : \mathbb{R}^n \rightarrow \mathbb{R}$, a aritmética intervalar define uma função $F(\mathbf{x}_1, \ldots, \mathbf{x}_n)$, onde as entradas $\mathbf{x}_i$ são intervalos. A nova função F retorna um intervalo - preferencialmente o menor - que contém todos os valores possíveis de $f(x_1, \ldots, x_n)$, para x_i variando em $\mathbf{x}_i$ e $i = 1, \ldots, n$.

Os números a e b de um intervalo $[a, b]$ de números reais podem não ser representados em um dado computador. Nesse caso, arredonda-se a para o maior número em ponto flutuante menor que a e arredonda-se b para o menor número em ponto flutuante maior que b. Dessa forma, o intervalo obtido ainda contém $[a, b]$. Esse procedimento é chamado de **arredondamento externo**. Todas as máquinas que suportam o padrão IEEE, como a maioria dos PC's e as estações de trabalho, permitem arredondamento externo.

As aplicações da aritmética intervalar não se restringem à limitação dos erros de arredondamento. A aritmética intervalar aplicada a problemas de otimização global permite a determinação de intervalos arbitrariamente pequenos que **garantidamente** contenham o máximo (ou mínimo) global, algo que métodos estocásticos ou heurísticas são incapazes de realizar. A primeira aplicação da aritmética intervalar em otimização foi feita por Robinson [Robinson (1973)].

Os algoritmos tradicionais de otimização avaliam a função objetivo somente em um número finito de pontos. Entretanto, não fornecem informações sobre oscilação da função entre os pontos avaliados. Um dos mais importantes recursos da aritmética intervalar é fornecer limites para a variação da função objetivo em conjunto contínuo de pontos, inclusive aqueles sem representação em ponto flutuante.

4.2.1 Definições e operações

Sejam $\underline{x}$ e $\bar{x}$ elementos de $\mathbb{R}$ tais que $\underline{x} \leq \bar{x}$. O conjunto $\mathbf{x} = [\underline{x}, \bar{x}]$ é o intervalo (fechado) $\{x \in \mathbb{R} : \underline{x} \leq x \leq \bar{x}\}$. Denotaremos o conjunto de todos os intervalos reais por $\mathbb{IR}$.

O **centro** $c(\mathbf{x})$ do intervalo $\mathbf{x}$ é dado por

$$c(\mathbf{x}) = \frac{\underline{x} + \bar{x}}{2}. \tag{4.1}$$

O **raio** $rad(\mathbf{x})$ de $\mathbf{x}$ é definido como a metade da distância entre os extremos de $\mathbf{x}$. Ou seja,

$$rad(\mathbf{x}) = \frac{\bar{x} - \underline{x}}{2}. \tag{4.2}$$

São também úteis as definições

$$\inf(\mathbf{x}) = \underline{x} \text{ e } \sup(\mathbf{x}) = \overline{x}. \tag{4.3}$$

As operações aritméticas podem ser estendidas para o conjunto $\mathbb{IR}$.

Definição 10 (Operações aritméticas). *Para* $\mathbf{x}, \mathbf{y} \in \mathbb{IR}$ *define-se:*

$$\mathbf{x} \circ \mathbf{y} = [\min_{\substack{x \in \mathbf{x} \\ y \in \mathbf{y}}} (x \circ y), \max_{\substack{x \in \mathbf{x} \\ y \in \mathbf{y}}} (x \circ y)]. \tag{4.4}$$

onde $\circ$ *é qualquer operação aritmética em* $\{+, -, *, /\}$.

Usando a definição 10, obtém-se:

$$\mathbf{x} + \mathbf{y} = [\underline{x} + \underline{y}, \bar{x} + \bar{y}], \tag{4.5}$$

$$\mathbf{x} - \mathbf{y} = [\underline{x} - \bar{y}, \bar{x} + \underline{y}], \tag{4.6}$$

$$\mathbf{x} * \mathbf{y} = [\min\left\{\underline{x}\underline{y}, \bar{x}\underline{y}, \underline{x}\bar{y}, \bar{x}\bar{y}\right\}, \max\left\{\underline{x}\underline{y}, \bar{x}\underline{y}, \underline{x}\bar{y}, \bar{x}\bar{y}\right\}], \tag{4.7}$$

$$\frac{1}{\mathbf{y}} = \left[\frac{1}{\bar{y}}, \frac{1}{\underline{y}}\right] \text{ (se } 0 \notin \mathbf{y}) \text{ e} \tag{4.8}$$

$$\frac{\mathbf{x}}{\mathbf{y}} = \mathbf{x} * \frac{1}{\mathbf{y}} \text{ (se } 0 \notin \mathbf{y}). \tag{4.9}$$

Para $n = 1, 2, \ldots,$ define-se:

$$\mathbf{x}^n = \begin{cases} [1,1], & \text{se } n = 0; \\ [\underline{x}^n, \bar{x}^n], & \text{se } \underline{x} \geq 0 \text{ ou } n \text{ é ímpar}; \\ [\bar{x}^n, \underline{x}^n], & \text{se } \bar{x} \leq 0 \text{ e } n \text{ é par}; \\ [0, \max(\underline{x}^n, \bar{x}^n)], & \text{se } \underline{x} \leq 0 \leq \bar{x} \text{ e } n \text{ é par}. \end{cases} \tag{4.10}$$

É importante salientar que algumas propriedades algébricas dos números reais não se aplicam aos intervalos. Por exemplo, $0 \in \mathbf{x} - \mathbf{x}$ ao invés de $0 = \mathbf{x} - \mathbf{x}$. Além disso, temos

Proposição 2 (Subdistributividade). *Sejam* $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{IR}$*, então*

$$(\mathbf{x} + \mathbf{y}) * \mathbf{z} \subseteq \mathbf{x} * \mathbf{z} + \mathbf{y} * \mathbf{z}, \tag{4.11}$$

ao invés da lei de distributividade.

4.2.2 Extensões intervalares de funções

As funções definidas para os números reais podem ser estendidas para o conjunto $\mathbb{IR}$. Formalmente,

Definição 11 (Extensões intervalares). *Sejam* $f : D_f \subset \mathbb{R}^n \to \mathbb{R}$ *e* $\mathbf{x} \in \mathbb{IR}^n$*. Defina* $f(\mathbf{x}) \equiv \{f(x) : x \in \mathbf{x} \cap D_f\}$*. Uma extensão de* f *em* $\mathbf{x}$ *é qualquer função* $F : \mathbb{IR}^n \to \mathbb{IR}$ *tal que* $F(\mathbf{x}) \supset f(\mathbf{x})$*.*

A definição 11 admite, portanto, mais de uma extensão intervalar para uma função real f. A precisão nos limites gerados pela aritmética intervalar está condicionado ao número de vezes que uma determinada variável aparece na definição da função. Por exemplo,

$$F_1(\mathbf{x}) = \mathbf{x}^2 - \mathbf{x} \text{ e } F_2(\mathbf{x}) = (\mathbf{x} - 1/2)^2 - 1/4 \tag{4.12}$$

são extensões intervalares para

$$f(x) = x^2 - x \ (x \in \mathbf{x} = [0, 2] \subset \mathbb{R}), \tag{4.13}$$

apesar de não produzirem o mesmo resultado, pois

$$F_1([0, 2]) = [-2, 4] \text{ e } F_2([0, 2]) = [-1/4, 2]. \tag{4.14}$$

O resultado gerado por F_2, do ponto de vista qualitativo, é superior ao gerado por F_1, já que $F_2([0, 2]) = f([0, 2])$ e isto se deve a repetição da variável X na definição de F_1. Este fenômeno de **superestimação** devido a repetição da variável é conhecido como **interdependência** e, em geral, não pode ser evitado.

Definição 12 (Extensão intervalar ótima). *Diremos que $F : \mathbb{IR}^n \to \mathbb{IR}$ é uma extensão intervalar ótima para $f : D_f \subset \mathbb{R}^n \to \mathbb{R}$ em $X \in \mathbb{IR}^n$ se $F(\mathbf{x}) = f(\mathbf{x})$.*

O principal uso das extensões intervalares é a estimação da imagem das funções consideradas. Se $f : \mathbb{R} \to \mathbb{R}$ é uma função monótona crescente, então a extensão ótima para f em $\mathbf{x} = [\underline{x}, \bar{x}]$ é dada por $F(\mathbf{x}) = [f(\underline{x}), f(\bar{x})]$. De forma inteiramente análoga, se f é monótona decrescente, então a extensão ótima de f em $\mathbf{x}$ é $F(\mathbf{x}) = [f(\bar{x}), f(\underline{x})]$.

Uma estratégia para a estimação da imagem de funções que não são monótonas é dividir o seu domínio em intervalos menores de tal forma que, restritas a estes intervalos, elas gozem de monotonicidade. Por exemplo, através do conhecimento tanto do comportamento quanto do ponto de máximo e^{-1} de $h(x) = x \log_2(x)$ em $\mathbf{x} = [0, 1]$ (ver figura 4.1), podemos construir uma extensão ótima para h em $\mathbf{x}$ e, assim, contornar os efeitos da interdependência. A imagem da função $h = x \log_2(x)$ em $\mathbf{y} = [0.2, 0.6]$ pode ser obtida dividindo-se o intervalo $\mathbf{y}$ nos intervalos $\mathbf{y}_1 = [0.2, e^{-1}]$ e $\mathbf{y}_2 = [e^{-1}, 0.6]$. Em $\mathbf{y}_1$, a função $h(y)$ é monótona crescente, portanto $h(\mathbf{y}_1) = [h(0.2), h(e^{-1})]$. Já em $\mathbf{y}_2$, a função $h(y)$ é monótona decrescente, então $h(\mathbf{y}_2) = [h(e^{-1}), h(0.6)]$. Uma vez que $h(0.6) < 0.45 < h(0.2)$, concluímos que $h(\mathbf{y}_1 \cup \mathbf{y}_2) = h(\mathbf{y}) = [h(0.2), h(e^{-1})]$.

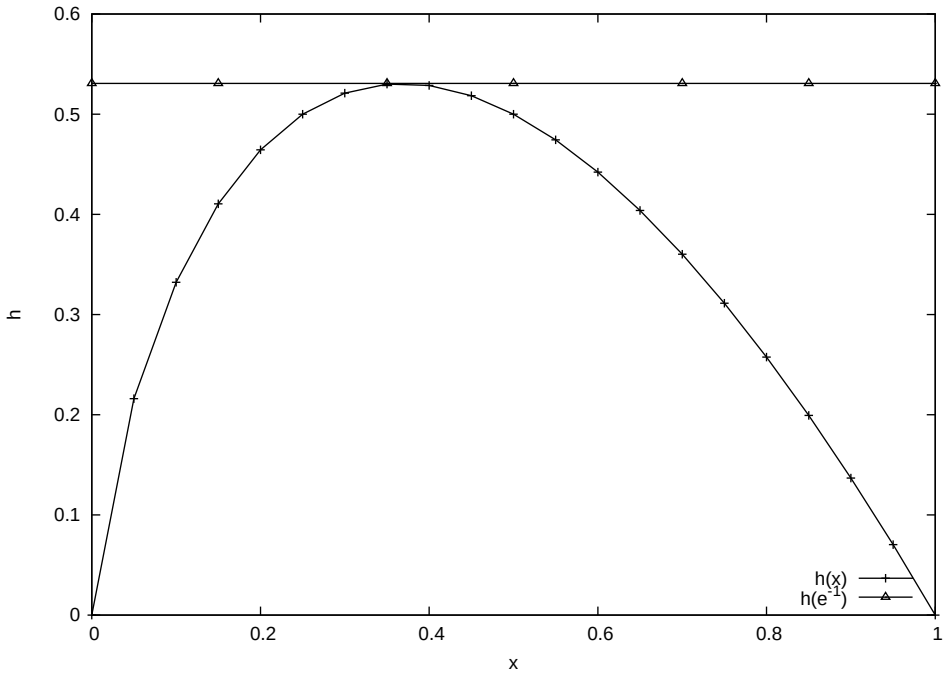

Figura 4.1: A função $h(x) = x \log_2(x)$ e seu máximo $h(e^{-1})$.

4.2.3 Vetores e matrizes intervalares

Um **vetor intervalar** é um vetor cujos elementos são intervalos. De forma similar, uma **matriz intervalar** é uma matriz cujos elementos são intervalos e o espaço de todas as matrizes $m \times n$ é denotado por $\mathbb{IR}^{m \times n}$.

Da mesma forma que a aritmética dos vetores e matrizes reais é uma extensão da aritmética em $\mathbb{R}$, a aritmética dos vetores e matrizes intervalares é baseada na aritmética de $\mathbb{IR}$. Dados $\mathbf{x} \in \mathbb{IR}^n$ e $\mathbf{A} \in \mathbb{IR}^{m \times n}$, definimos o centro de $\mathbf{x}$ pelo vetor real $c(\mathbf{x}) \equiv (c(\mathbf{x}_1), \ldots, c(\mathbf{x_n}))$ e o centro de $\mathbf{A}$ pela matriz real $c(\mathbf{A}) \equiv [c(\mathbf{A}_{ij})]_{m \times n}$. O raio de um vetor (matriz) intervalar é definido como o maior raio das componentes do vetor (matriz).

4.2.4 Método de Newton intervalar

Nesta seção, para determinar (aproximadamente) as soluções de sistemas não-lineares, apresentaremos uma importante ferramenta, o método de Newton intervalar.

A versão do método de Newton que apresentaremos baseia-se no teorema do

valor médio e no método de Gauss-Seidel para solução de sistemas lineares.

Teorema 6 (Teorema do valor médio). *Sejam X um conjunto real convexo e $a, b \in X$. Se $f : X \to \mathbb{R}^n$ é uma função diferenciável, então existe $\xi \in X$ tal que*

$$f(b) = f(a) + \nabla f(\xi)(b - a). \tag{4.15}$$

O objetivo é determinar a solução $x^* \in \mathbf{x}^{(k)} \in \mathbb{IR}^n$ do sistema não-linear $f(x) = 0$, onde $f : \mathbb{R}^n \to \mathbb{R}^n$ é uma função diferenciável em $\mathbf{x}^{(k)}$. Para tal, fixamos uma caixa (vetor intervalar) $\mathbf{x}^{(k)}$ que contenha uma solução do sistema e, utilizando o teorema do valor médio, recaímos em um sistema linear intervalar. Neste ponto, utilizamos uma iteração do método de Gauss-Seidel para obter uma nova caixa $\mathbf{x}^{(k+1)}$ que também contenha a solução x^* e tal que $rad(\mathbf{x}^{(k+1)}) \leq rad(\mathbf{x}^{(k)})$. O processo é, então, repetido até que critério de parada seja satisfeito.

Em detalhes, se $x^*, x^{(k)} \in \mathbf{x}^{(k)}$, então, pelo teorema do valor médio,

$$f(x^*) = f(x^{(k)}) + \nabla f(\xi)(x^* - x^{(k)}) = 0, \text{ com } \xi \in \mathbf{x}^{(k)}. \tag{4.16}$$

Isto implica

$$\nabla f(\xi)(x^* - x^{(k)}) = -f(x^{(k)}), \text{ com } \xi \in \mathbf{x}^{(k)}. \tag{4.17}$$

A partir da Eq. 4.17 e usando a aritmética intervalar, obtemos o sistema intervalar

$$\nabla f(\mathbf{x}^{(k)})(x - x^{(k)}) = -f(x^{(k)}). \tag{4.18}$$

Para simplificação da notação, defina $\mathbf{A} = \nabla f(\mathbf{x}^{(k)})$ e $b = f(x^{(k)})$ e, portanto, passamos ao sistema

$$\mathbf{A}(x - x^{(k)}) = -b. \tag{4.19}$$

Fixaremos a matriz $C = (c(\mathbf{A}))^{-1}$ para o precondicionamento do sistema.

A iteração do método de Gauss-Seidel para a solução de sistemas lineares

fornece a caixa $\mathbf{x}^{(k+1)}$ tal que $x^* \in \mathbf{x}^{(k+1)}$ e

$$\begin{aligned} \Delta^{(l)} &= \mathbf{x}^{(l)} - x^{(k)}, & (4.20) \\ GS(\mathbf{x}^{(k)})_i &= x_i^{(k)} - \frac{C_i b + \sum_{j=1}^{i-1} C_i \mathbf{A}_j \Delta_j^{(k+1)} + \sum_{j=i+1}^{n} C_i \mathbf{A}_j \Delta_j^{(k)}}{C_i \mathbf{A}_i}, & (4.21) \\ \mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} \cap GS(\mathbf{x}^{(k)}), & (4.22) \end{aligned}$$

onde e por C_i e $\mathbf{A}_j$ denotamos, respectivamente, a i-ésima e a j-ésima linhas de C e $\mathbf{A}$.

Os teoremas seguintes destacam as propriedades que permitem a resolução do sistema definido na Eq. (4.18) (para demonstrações consulte Hansen (1993)).

Teorema 7. *Se existe uma raiz x^* de ∇f em $\mathbf{x}$, então $x^* \in GS(\mathbf{x})$.*

Teorema 8. *Se $\mathbf{x} \cap GS(\mathbf{x}) = \emptyset$, então não existe raiz de ∇f em $\mathbf{x}$.*

Teorema 9. *Se $GS(\mathbf{x})$ está no interior de $\mathbf{x}$, então existe uma única raiz de ∇f em $\mathbf{x}$.*

4.3 O Algoritmo de otimização

Apresentaremos agora um algoritmo de otimização global baseado em um método BB que utiliza técnicas da aritmética intervalar. Esse algoritmo é baseado no algoritmo de Hansen [Hansen (1993)].

O problema é

$$\max f(x) \text{ sujeito a } \begin{cases} x \in \mathbf{x}, \\ a(x) = 0. \end{cases} \quad (4.23)$$

onde $f : \mathbb{R}^n \to \mathbb{R}$ e $a : \mathbb{R}^n \to \mathbb{R}^m$ são funções de classe C^2 e $\mathbf{x}$ é uma caixa no $\mathbb{R}^n$. Sejam f^* o maior valor de f em $\mathbf{x}$ e x^* um ponto onde esse valor é atingido, ou seja, $f^* = f(x^*)$.

Se x^* está no interior de $\mathbf{x}$, então $\nabla f(x^*) = 0$. Entretanto, o gradiente também se anula nos mínimos locais, nos máximos locais e em pontos que não são nem mínimos nem máximos locais. Usaremos o método de Newton intervalar,

descrito na seção 4.2.4, para encontrar os pontos que anulam o gradiente e que são soluções do sistema não linear $a(x) = 0$. Antes de aplicá-lo, faremos dois testes usando o gradiente e a Hessiana de f.

Consideremos uma subcaixa $\mathbf{b}$ de $\mathbf{x}$. O primeiro teste é a verificação da existência de pontos em $\mathbf{b}$ nos quais o gradiente da função f se anule. Se existir algum $i = 1, \ldots, n$ tal que $0 \notin \nabla f_i(\mathbf{b})$, então ∇f não se anula em nenhum ponto de $\mathbf{b}$ e, portanto, podemos descartar a caixa $\mathbf{b}$.

No segundo teste, avaliamos a Hessiana de f em $\mathbf{b}$. Se x^* está no interior de $\mathbf{x}$, então a Hessiana $\nabla^2 f$ de f é semidefinida negativa em x^*. Uma condição necessária para isso é que, para $i = 1, \ldots, n$ $\nabla^2 f_{ii}(x^*) \leq 0$, onde $\nabla^2 f_{ii}(x^*)$ são os elementos da diagonal de $\nabla^2 f(x*)$.

Consideremos, então, uma subcaixa $\mathbf{b}$ de $\mathbf{x}$. Se existir algum $i = 1, \ldots, n$ tal que $\nabla^2 f_{ii}(\mathbf{b}) > 0$, então $\nabla^2 f_{ii}(x) > 0$ para todo $x \in \mathbf{b}$ e, portanto, $\nabla^2 f_{ii}(x)$ não pode ser semidefinida negativa em nenhum ponto de $\mathbf{b}$. Por esse motivo, podemos descartar a caixa $\mathbf{b}$.

Supondo que a subcaixa $\mathbf{b} \subset \mathbf{x}$ não tenha sido descartada, aplicamos uma iteração do método de Newton intervalar visando a solução do sistema quanto $a(x) = 0$.

Se ao final do procedimento a subcaixa $\mathbf{b}$ não satisfizer os critérios abaixo:

$$rad(\mathbf{b}) \leq \varepsilon_X,\ rad(f(\mathbf{b})) \leq \varepsilon_f \text{ e } rad(a(\mathbf{b})) \leq \varepsilon_a; \tag{4.24}$$

onde $\varepsilon_{\mathbf{x}}$, ε_f, ε_a são, respectivamente, as tolerâncias dadas para o tamanho das caixas resultantes no final do algoritmo, para o tamanho dos intervalos que contém f^* e $a(X)$, então dividimos a caixa $\mathbf{b}$ em duas novas caixas e repetimos o procedimento.

A seguir, descrevemos os passos do algoritmo (ver Algoritmo 1). Os parâmetros X e m são, respectivamente, inicializados com os valores $\mathbf{x}$ e $c(\mathbf{x})$.

Sejam $\mathbf{c}_1, \ldots, \mathbf{c}_p$ as caixas resultantes. Calcule os limites $\underline{f} = \max\limits_{1 \leq i \leq p} \underline{f}(\mathbf{c}_i)$ e

Algorithm 1 Algoritmo de otimização (Versão recursiva).

1: **procedure** BB(X, m) ▷
2: **if** $(0 \notin a(X) \mid \sup(F(X)) < F(m) \mid 0 \notin \nabla F(X) \mid \nabla^2 F(X) > 0)$ **then**
3: return
4: **end if**
5: **if** $F(c(X)) > F(m)$ **then**
6: $m = c(X)$
7: **end if**
8: $X = X \cap GS_a$
9: **if** $rad(X) \leq \varepsilon_X$ & $rad(F(X)) \leq \varepsilon_F$ & $rad(a(x)) \leq \varepsilon_a$ **then**
10: X é solução
11: **else**
12: $X_A = [\underline{X}, c(X)]$
13: $BB(X_A, m)$
14: $X_B = [c(X), \overline{X}]$
15: $BB(X_B, m)$
16: **end if**
17: **end procedure**

$\bar{f} = \min_{1 \leq i \leq p} \bar{f}(\mathbf{c}_i)$. Então, temos que $\underline{f} \leq f^* \leq \bar{f}$ e $rad(\mathbf{c}_i) \leq \varepsilon_{\mathbf{x}}$ para $i = 1, \ldots, p$.

Capítulo 5

Resultados computacionais

Nesta seção, exploraremos um caso particular do problema do cálculo da informação acessível e, utilizando os limites teóricos conhecidos, procuraremos validar a metodologia baseada no BB aliado à aritmética intervalar.

O problema do cálculo da informação acessível em sua forma mais geral pode ser assim descrito: dado um ensemble de estados quânticos $\mathcal{R} = \{\rho_x, \mathrm{p}_x\}$ com operador densidade ρ e o conjunto $\mathcal{P}_\mathcal{R}$ formado por todos os POVM's que atuam no mesmo espaço que os operadores $\rho_x \in \mathcal{R}$, maximize a função $f_\mathcal{R} : \mathcal{P}_\mathcal{R} \to \mathbb{C}$ dada por

$$f_\mathcal{R}(\{E_y\}) = \sum_{x,y} \mathrm{p}_x \mathrm{tr}(E_y \rho_x) \log_2 \frac{\mathrm{tr}(E_y \rho_x)}{\mathrm{tr}(E_y \rho)}. \tag{5.1}$$

Em nossos experimentos numéricos estipularemos limites para a informação acessível do ensemble real $\mathcal{R} = \{\rho_0, \rho_1\}$ formado por operadores com posto unitário que atuam em $\mathbb{R}^2$ e que possuem distribuição de probabilidade $\mathrm{p}(0) = \mathrm{p}$ e $\mathrm{p}(1) = 1 - \mathrm{p}$. Ou seja, $\rho_0 = |\psi_0\rangle\langle\psi_0|$ e $\rho_1 = |\psi_1\rangle\langle\psi_1|$ com $|\psi_0\rangle, |\psi_1\rangle \in S^1$ (esfera de raio unitário em $\mathbb{R}^2$). Os vetores $|\psi_0\rangle$ e $|\psi_1\rangle$ admitem a seguinte parametrização:

$$|\psi_0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ e } |\psi_1\rangle = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} \text{ para } \theta \in [0, \pi/2]. \tag{5.2}$$

Neste caso, para cada $\theta \in [0, \pi/2]$ e $\mathrm{p} \in (0, 1)$, teremos um ensemble $\mathcal{R} = \mathcal{R}(\theta, \mathrm{p})$

com operador densidade ρ dado por

$$\begin{aligned} \rho &= \mathrm{p}\rho_0 + (1-\mathrm{p})\rho_1 && (5.3)\\ &= \begin{pmatrix} \mathrm{p} + (1-\mathrm{p})\cos^2(\theta) & (1-\mathrm{p})\cos(\theta)\sin(\theta) \\ (1-\mathrm{p})\cos(\theta)\sin(\theta) & (1-\mathrm{p})\sin^2(\theta) \end{pmatrix}. && (5.4) \end{aligned}$$

Pelo corolário 1 (Davies para ensembles reais), a cardinalidade n do POVM E^* que maximiza a informação acessível do ensemble $\mathcal{R}$ é tal que $2 \leq n \leq 3$ e, além disso, E^* é formado por operadores E_y da forma:

$$E_y = \alpha_y |\phi_y\rangle\langle\phi_y|, \text{ com } |\phi_y\rangle = \begin{pmatrix} \cos(\phi_y) \\ \sin(\phi_y) \end{pmatrix} \text{ e } \alpha_y \in [0,1]. \tag{5.5}$$

A função objetivo $f_{\mathcal{R}}$ passa, então, a ser expressa por

$$\begin{aligned} f_{\mathcal{R}}(\alpha,\phi) &= \sum_{x,y} \mathrm{p}(x)\ \mathrm{tr}(\alpha_y\,|\phi_y\rangle\,\langle\phi_y|\psi_x\rangle\,\langle\psi_x|)\log_2 \frac{\mathrm{tr}(\alpha_y\,|\phi_y\rangle\,\langle\phi_y|\psi_x\rangle\,\langle\psi_x|)}{\mathrm{tr}(\alpha_y|\phi_y\rangle\langle\phi_y|\rho)} && (5.6)\\ &= \frac{1}{\ln 2}\sum_{x,y} \mathrm{p}(x)\alpha_y\langle\phi_y|\psi_x\rangle^2 \ln \frac{\langle\phi_y|\psi_x\rangle^2}{\langle\phi_y|\,\rho\,|\phi_y\rangle}. && (5.7) \end{aligned}$$

O problema do cálculo da informação acessível assume a forma:

$$\max_{\alpha,\,\phi} f_{\mathcal{R}}(\alpha,\phi) \text{ sujeito a } \begin{cases} 2 \leq n \leq 3; \\ \alpha = (\alpha_1,\ldots,\alpha_n) \in [0,1]^n; \\ \phi = (\phi_1,\ldots\,\phi_n) \in [0,2\pi]^n; \\ \displaystyle\sum_{y=1}^{n} \alpha_y|\phi_y\rangle\langle\phi_y| = I_2 \text{ (matriz identidade de } \mathbb{R}^2). \end{cases} \tag{5.8}$$

As restrições do problema podem ser simplificadas. De fato, sabendo que a cardinalidade de E^* é igual a 2, i.e., $n = 2$ [Shor (2000)], a última restrição do problema de otimização é dada explicitamente por

$$\sum_{i=1}^{2} \alpha_i \begin{pmatrix} \cos^2(\phi_i) & \sin(\phi_i)\cos(\phi_i) \\ \sin(\phi_i)\cos(\phi_i) & \sin^2(\phi_i) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{5.9}$$

Aplicando o operador traço em ambos os lados da igualdade obtemos:

$$\alpha_1 + \alpha_2 = 2. \tag{5.10}$$

Uma vez que $\alpha_i \in [0,1]$, concluímos que $\alpha_1 = \alpha_2 = 1$. Agora, usando esta informação e as identidades

$$\cos^2(x) = \frac{1+\cos(2x)}{2}, \sin(x)\cos(x) = \frac{\sin(2x)}{2} \text{ e } \sin^2(x) = \frac{1-\cos(2x)}{2},$$

podemos reescrever a equação matricial Eq. (5.9) através do sistema

$$\begin{cases} \cos(2\phi_1) + \cos(2\phi_2) = 0 \\ \sin(2\phi_1) + \sin(2\phi_2) = 0 \end{cases} \tag{5.11}$$

cuja solução é

$$\phi_2 = \phi_1 + \frac{\pi}{2}. \tag{5.12}$$

Isto produz uma drástica simplificação do problema de otimização, pois passamos a um problema de otimização cuja restrição é dada por um intervalo.

Formalmente, o problema do cálculo da informação acessível do ensemble $\mathcal{R} = \mathcal{R}(\theta, \mathrm{p})$ tratado em nosso experimento é dado por

$$\max_{\phi \in [0,\pi/2]} F(\phi), \tag{5.13}$$

onde

$$F(\phi) = \frac{1}{\ln(2)} \sum_{x,y=0}^{1} p(x)\langle\phi_y|\psi_x\rangle^2 \ln \frac{\langle\phi_y|\psi_x\rangle^2}{\langle\phi_y|\,\rho\,|\phi_y\rangle}, \ \phi_0 = \phi \text{ e } \phi_1 = \phi + \pi/2. \tag{5.14}$$

Devido ao logaritmo na definição da função F, tanto o gradiente quanto a

Hessiana da função F não estão definidos em todos os pontos $\phi \in [0, \pi/2]$. No entanto, este não é um problema crítico para a metodologia baseada no BB com a aritmética intervalar, pois basta não excluirmos as caixas que contêm estes pontos utilizando os critérios de nulidade do gradiente e não positividade da Hessiana. Graças à capacidade da aritmética intervalar de gerar intervalos arbitrariamente estreitos que contenham o máximo da função F, até mesmo estas caixas, caso não atinjam o máximo da função F, serão excluídas durante a execução do algoritmo. No entanto, é claro que quanto maior for o número de critérios que permitam a exclusão das caixas, maior será a velocidade e convergência do método.

A figura 5.1 exibe o comportamento da função $F(\phi)$ para p = 0.5 e $\theta = \pi/4$.

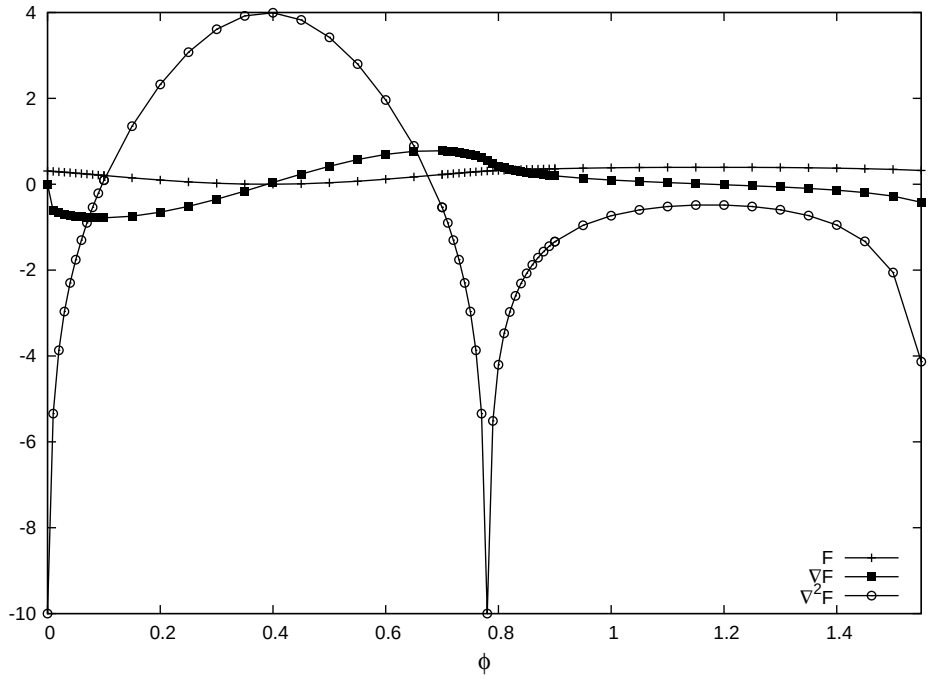

Figura 5.1: Comportamento da função objetivo $F(\phi)$ para e p = 0.5 e $\theta = \pi/4$.

Para p = 1/2 e θ variando de 0 a $\pi/2$, a diferença entre os limites teóricos $\chi = \chi(\theta)$ e $\varphi = \varphi(\theta)$ (Teoremas 3 e 4) para a informação acessível do ensemble $\mathcal{R}(\mathrm{p} = 1/2, \theta)$ tem valor médio

$$VM_T = \frac{2}{\pi} \int_0^{\pi/2} \chi(\theta) - \varphi(\theta)\, d\theta \approx 0.31966 \text{ (ver figura 5.2).} \qquad (5.15)$$

Devido à simplicidade da restrição do caso particular que estamos abordando,

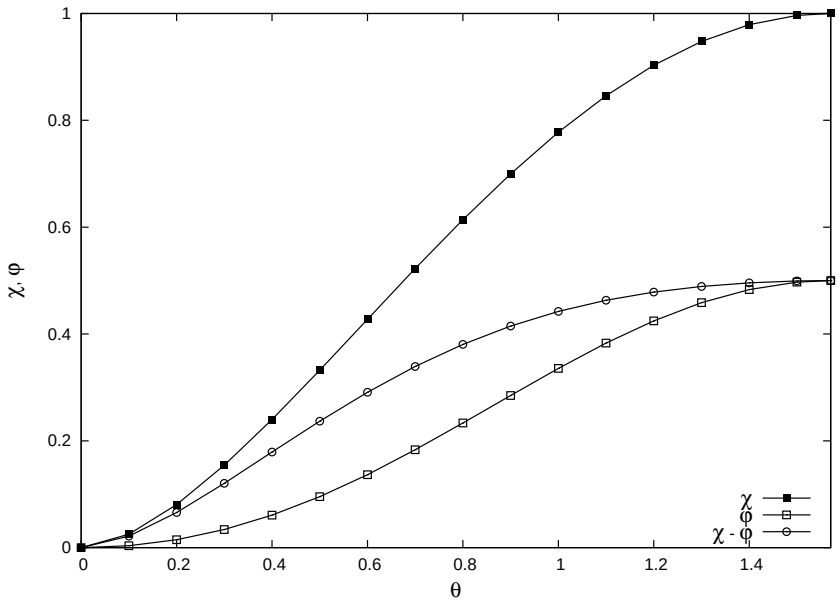

Figura 5.2: O limite superior de Holevo $\chi(\theta)$ e o limite inferior de Jozsa-Robb-Wootters $\varphi(\theta)$ para a informação acessível do ensemble eqüiprovável formado pelos estados puros ρ_0 e ρ_1.

o algoritmo que utilizaremos será uma versão simplificada do que foi apresentado no capítulo anterior (ver Algoritmo 2). A implementação foi feita utilizando o pacote **Intlab** [Hargreaves (2002)] do software Matlab®.

Algorithm 2 Algoritmo de otimização (Versão recursiva simplificada).

procedure BB(X, m) ▷
  if $(\sup(F(X)) < F(m) \mid 0 \notin \nabla F(X) \mid \nabla^2 F(X) > 0)$ **then**
    return
  end if
  if $F(mid(X)) > F(m)$ **then**
    $m = mid(X)$
  end if
  if $(rad(X) \leq \varepsilon_X$ & $rad(F(X)) \leq \varepsilon_F)$ **then**
    X é solução
  else
    $X_A = [\underline{X}, mid(X)]$
    $BB(X_A, m)$
    $X_B = [mid(X), \overline{X}]$
    $BB(X_B, m)$
  end if
end procedure

Para ϕ variando de 0 a $\pi/2$, p = 0.5 e $\theta = \pi/4$, os limites encontrados para o máximo de $F(\phi)$ usando a metodologia sugerida com critérios de parada $\varepsilon_x = 10^{-5}$

e $\varepsilon_F = 10^{-2}$ foram $L_i = 0.3842$ e $L_s = 0.4043$ (ver figura 5.3).

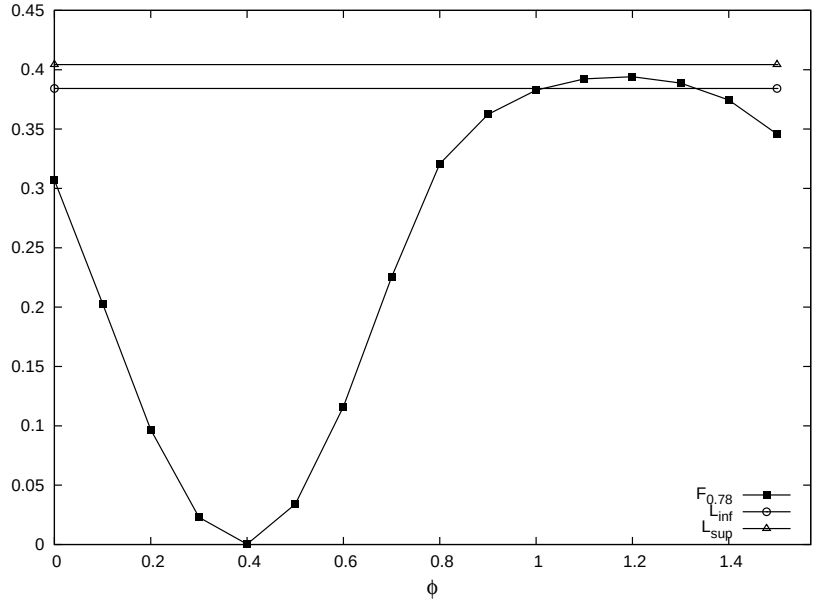

Figura 5.3: Limites numéricos obtidos pelo método BB para máximo da função $F(\phi)$ para $\theta = 1.0$ e p $= 1/2$.

Na figura 5.4, comparamos os limites teóricos e os limites numéricos obtidos para p $= 1/2$ e θ variando de 0 a $\pi/2$. O valor médio da diferença entre os limites numéricos satisfaz a desigualdade

$$VM_N \leq 2 \times 10^{-2} < 6\% \text{ de } VM_T. \tag{5.16}$$

No que tange a acurácia, os resultados obtidos são, portanto, consideravelmente superiores aos limites teóricos apresentados.

Na figura 5.5, exibimos o tempo gasto pelo algoritmo e o número de caixas resultantes. Todos os tempos foram verificados utilizando um Pentium III 800MHz com 368MB de Ram rodando Windows XP. Como esperado, uma vez que a exclusão de subcaixas se baseia na comparação dos limites gerados para F em cada subcaixa, o tempo gasto pelo algoritmo e o número de caixas resultantes aumentam à medida que o valor de θ se aproxima de 0, pois, neste caso, a função F pouco varia (ver figura 5.6).

A análise do custo do método BB apresentado não pode ser feita diretamente, pois não há informação a priori sobre os limites que serão gerados na fase

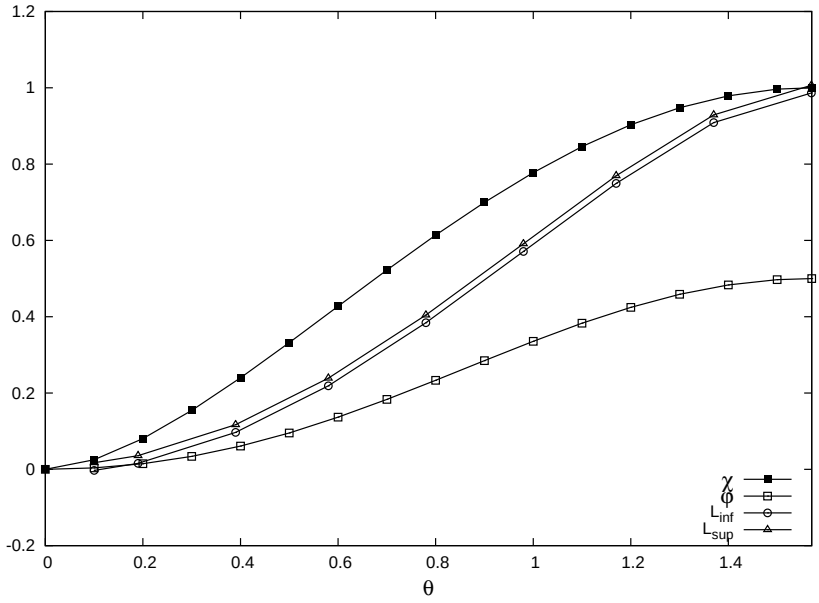

Figura 5.4: O limite superior de Holevo $\chi(\theta)$, o limite inferior de Jozsa-Robb-Wootters $\varphi(\theta)$ e os limites numéricos obtidos pelo método BB para a informação acessível do ensemble eqüiprovável formado pelos estados puros ρ_0 e ρ_1.

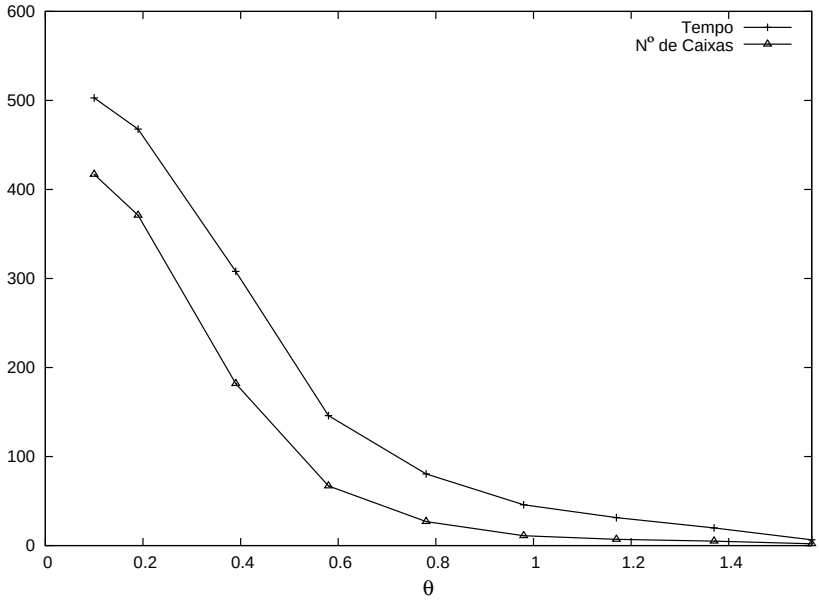

Figura 5.5: O tempo em segundos para o término da execução do algoritmo e o número de caixas resultantes.

de **bounding** pelas técnicas da aritmética intervalar. No entanto, no caso mais simples, i.e., quando as restrições $rad(X) \leq \varepsilon_x$ e $rad(f(X)) \leq \varepsilon_F$ são equivalentes, teremos $rad(\mathbf{x})/\varepsilon_x$ iterações. Convém salientar a simplicidade do algoritmo como motivação para sua aplicação e, além disso, a possibilidade do uso da computação paralela para aceleração do método. Basicamente, poderíamos analisar separa-

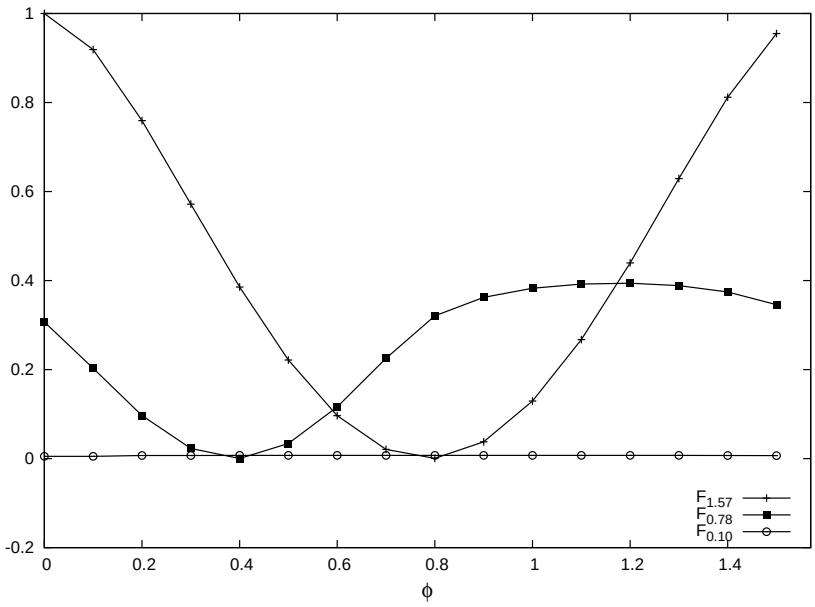

Figura 5.6: A variação dos valores de F_θ diminuem à medida que θ tende a 0.

damente diferentes regiões do intervalo inicial, comparar os resultados e, então, selecionar os mais promissores. Além disso, o uso de métodos híbridos que selecionem pontos de máximo locais, ao invés de selecionar o ponto médio de $\mathbf{x}$, podem também acelerar a execução do algoritmo.

Capítulo 6

Conclusão

Este trabalho teve como objetivo a aplicação do potencial dos métodos de otimização global ao problema do cálculo da informação acessível. Procuramos enunciar o problema no campo da teoria da informação, buscando tornar compreensível a escolha do critério da informação mútua como parâmetro coerente para o cálculo da informação acessível.

Os experimentos numéricos realizados fundamentaram-se no método de otimização **branch and bound** (BB) aliado à aritmética intervalar. Sua escolha nos pareceu razoável, dado o caráter determinístico do BB e a capacidade das técnicas da aritmética intervalar para a geração de limites garantidamente válidos para funções.

Os resultados numéricos obtidos quando confrontados com os resultados teóricos de maior projeção, o limite superior de Holevo e o limite inferior de Josza-Robb-Wootters, obtiveram maior acurácia, estreitando consideravelmente o intervalo possível para os valores da informação acessível.

Em pesquisas futuras, procuraremos aplicar o conhecimento adquirido no desenvolvimento de uma metodologia para a abordagem do problema geral do cálculo da informação acessível. Outra contribuição possível é o fornecimento de soluções para um maior número de casos particulares para que, a partir delas, seja possível a dedução de soluções analíticas. O desenvolvimento e o uso de algoritmos BB híbridos, i.e., que explorem as propriedades da aritmética intervalar aliados à

computação paralela e que selecionem pontos de máximo locais, ao invés de pontos pré-determinados, nos parecem ser alternativas viáveis para a redução do tempo de computação.

Referências Bibliográficas

Benson, H. P., 1982. On the convergence of two branch and bound algorithms for nonconvex programming problems. *Journal of optimization theory and applications* 36, 129–134.

Davies, E. B., 1978. Information and quantum measurement. *IEEE Transactions on Information Theory* IT-24 (5), 596–599.

Dieks, D., 1982. Communication by EPR devices. *Phys. Lett. A* 92 (6), 271–272.

Fuchs, C. A., 1995. *Distinguishability and accessible information in quantum theory*. Ph.D. thesis, University of New Mexico, Department of Physics and Astronomy.

Goldberg, D., March 1991. What every computer scientist should know about floating-point arithmetic. *ACM Computing Surveys* 23 (1), 5–48.

Gordon, J. P., 1964. *Noise at optical frequencies; information theory*. Quantum Electronics and Coherent Light; Proceedings of the International School of Physics Enrico Fermi, Course XXXI. Academic Press New York.

Hansen, E. R., 1993. *Global Optimization using Interval Analysis*. Springer-Verlag, Berlin.

Hargreaves, G. I., 2002. Interval Analysis in MATLAB. *Numerical Analysis Report* 416.

Holevo, A. S., 1973. Bounds for the information transmitted by a quantum communication channel. *Problems of Information Transmission* 9, 177–183.

Horst, R., Tuy, H., 1987. On the convergence of global methods in multiextremal optimization. *Journal of optimization theory and applications* 54, 253–271.

Jozsa, R., Robb, D., Wootters, W. K., 1994. Lower bound for accessible information in quantum mechanics. *Physical Review A* 49, 668–677.

Moore, R. E., 1962. *Interval arithmetic and automatic error analysis in digital computing*. Ph.D. thesis, Stanford University.

Neumaier, A., 2004. Complete search in continuous global optimization and constraint satisfaction. In: Iserles, A. (Ed.), *Acta Numerica 2004*. Cambridge University Press, Cambridge.

Nielsen, M. A., Chuang, I. I., 2003. *Quantum computation and quantum information*. Bookman.

Pintér, J., 1988. Branch and bound algorithms for solving global optimization problems with lipschitzian structure. *Optimization* 19, 101–110.

Robinson, S. M., 1973. Computable error bounds for nonlinear programming. *Math. Programming* 5.

Sasaki, M., Barnett, S. M., Jozsa, R., Osaki, M., Hirota, 1999. Accessible Information and Optimal Strategies for Real Quantum Sources. *Phys. Rev. A* 59, 3325–3335.

Shanon, C. E., 1948. A mathematical theory of communication. *The Bell Sistem Tech* 27, 379–429,623–656.

Shor, P. W., 2000. Quantum Information Theory: Results and Open Problems. *GAFA2000*, 816–838.

Wootters, W. K., Zurek, W. H., October 1982. A single quantum cannot be cloned. *Nature* 299, 802–803.