

# TESE

submetida por

**ÉRICA MARQUES DA SILVA**

para obter o título de

**MESTRE**

pelo Centro Brasileiro de Pesquisas Físicas - CBPF

(Especialidade: Instrumentação Científica)

## **Inversão de Fourier Utilizando o Método do Máximo de Entropia**

Data da defesa: 25 de Fevereiro de 2005

### **Composição da Banca:**

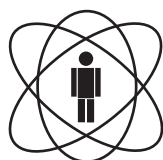
Presidente: Márcio Portes de Albuquerque (CBPF)

Examinador: Ivan dos Santos Oliveira Júnior (CBPF)

Examinador: Rubén Jesus Sanchez Rodriguez (UENF)

Examinador: Marcelo Portes de Albuquerque (CBPF - convidado especial)

Examinador: Ademarlaudo França Barbosa (CBPF - suplente)



**CBPF**

<http://www.cbpf.br>



**Ministério  
da Ciência  
e Tecnologia**

# TESE DE MESTRADO

## INVERSÃO DE FOURIER UTILIZANDO O MÉTODO DO MÁXIMO DE ENTROPIA

ÉRICA MARQUES DA SILVA

MESTRADO EM INSTRUMENTAÇÃO CIENTÍFICA

CENTRO BRASILEIRO DE PESQUISAS FÍSICAS

RIO DE JANEIRO, JANEIRO DE 2005

*Dedico este trabalho ao meu filho Gustavo, meu esposo Valdecir e o meu grande amigo Romeu. Obrigado pelo amor e por estarem sempre do meu lado transmitindo apoio e carinho durante minha vida.*

# Agradecimentos

Em primeiro lugar agradeço a *Deus*.

Neste trabalho contei com a colaboração e incentivo de várias pessoas, sendo todas fundamentais na realização deste trabalho tanto no sentido de me prestarem informações quanto no incentivo pessoal.

Em primeiro lugar gostaria de agradecer aos meus orientadores, Márcio Portes de Albuquerque e Marcelo Portes de Albuquerque pela convivência amiga, pelos ensinamentos e críticas, ao longo destes anos. Através de suas valiosas orientações dei meus primeiros passos na pesquisa científica e com sua confiança e incentivo tive a oportunidade de amadurecer profissionalmente e como pessoa.

Ao Engenheiro Ismar Tomaz Jabur, responsável pela Coordenação de Atividades Técnicas (CAT-CBPF); aos amigos e colegas da CAT, Eliene, Rogério, Fernando, Denise, Jaime, Sandro, Fagner, Eduardo, Marita, Deyse, Nilton, Ismar, Gabriel, Aline, Eugênio e Cristiano.

Aos professores que sempre me incentivaram: Ivan Oliveira, Geraldo Cernicchiaro, Ademar-laudo Barbosa, Elís Sinnecker, Roberto Sarthour, Flávio Garcia e Alberto Passos Guimarães. Aos Professores de graduação Ademar B. Guimarães e Lourival de Souza Lima agradeço pelo empenho, dedicação e ensinamentos transmitidos durante a graduação.

Agradeço também a Ir. Cristina Maria Pastore, diretora da Faculdade de Filosofia Ciências e Letras “Santa Marcelina”, que com sua competência e dedicação ajudou e incentivou para o meu sucesso.

Ao meu grande amigo Romeu Abrahão Pereira, cuja sabedoria, amizade, sugestões discussões e apoios sempre disponíveis, me inspiraram e me guiaram ao longo desta jornada. Meu sincero reconhecimento e gratidão.

Um agradecimento especial aos meus amigos que muito me ajudaram na construção desta tese, Eugênio, Aline, Brayan, Érick, Luiz Lima, Luciano, Marcelo Perantoni, Rogério, Fagner, Eduardo e Sarah.

Aos meus amigos de curso, Sarah, Luciano, Rogério e Luiz Lima (Lula), Marcelo Perantoni, Francisco, Ana Amélia, Valberto e Rúben.

Aos meus grandes amigos que fiz nesta etapa da minha vida Ailin, Rodrigo, Fabiana, Eni, Simone, Beethovem e em especial a duas grandes amigas Sarah (minha irmã) e Célia. Vocês são demais!

Agradeço a minha família, pela amizade, atenção e incentivo, e principalmente pelo apoio e amor constantes.

Aos meus pais Abelar e Adenir, que sempre me conduziram na busca de novos caminhos para que eu alcançasse sempre um maior desenvolvimento intelectual e emocional. Ao meu irmão Éric pelo amor, carinho e companheirismo e por sua felicidade em me ajudar em momentos difíceis.

Agradeço ao meu esposo Valdecir, pelo amor, confiança e pela compreensão dos momentos nos quais estive ausente, ao meu filho Gustavo que mesmo tão pequeno, entendeu minha ausência e sempre me recebeu com um belo sorriso.

Aos meus amigos Tatiana, Lucieny, Alexandre, Ana Paula, Mirela e Lys, que por consequência do caminho que escolhemos, acabamos nos distanciando, mas para uma verdadeira amizade distância não é problema.

Por fim, agradeço ainda à Myriam, Fernando, Cecília, Rosemary, Rosana, Ivani, Aparecida, Lucimar, Dulce, Rogério, Bruno, Diego, Lucelena, Jaqueline, Lucas, Alexssander e Carmem . . . enfim a todos os meus amigos que souberam me entender e ajudar nos momentos em que precisei.

# Resumo

Esta tese aborda a utilização do método do máximo de entropia (MME) no cálculo da transformada de Fourier. Em experimentos onde as medidas são realizadas no domínio da frequência a inversão de Fourier clássica é a ferramenta mais utilizada na reconstrução do sinal no espaço direto. Em tais situações, o procedimento de medida introduz no sinal reconstruído diversas oscilações devido a discretização do espectro, a incerteza experimental e a limitação do instrumento para medir em todo o domínio frequencial. O MME produz um resultado melhor nestas circunstâncias devido a utilização de um enfoque bayesiano. Esta técnica nos permite melhorar a estimação do sinal, uma vez que introduz no cálculo um conhecimento *a priori*, i.e. insere a expectativa prévia da medida por meio de um modelo teórico dos resultados que podem ser obtidos. A inversão de Fourier à máximo de entropia foi implementada em um programa em MATLAB. Sua eficiência foi avaliada por meio da simulação de um espectro incompleto contendo erros experimentais conhecidos. Quatro estudos de casos foram propostos e os resultados analisados detalhadamente, evidenciando as propriedades fundamentais e formas de utilização do MME.

# Abstract

Fourier transforms plays an important role in signal analysis in experimental physics. In this thesis we are interested in Fourier transform by maximum entropy approaches. In scientific experiments where measurements are carried through in the frequency domain, classic Fourier inversion is a commonly used tool to reconstruct signals on direct space. In such situations, the measurement procedure leads to oscillations on the reconstructed signal due to spectrum discretization, experimental errors and instrument limitations in the frequency domain. Bayesian approaches can achieve better results with maximum entropy methods (MEM) in these circumstances. This technique improves the signal estimation, introducing in the calculation a priori knowledge, i.e. it puts in a previous expectations of the measurement through a theoretical model of the expected results. A MATLAB code was developed to calculate the MEM Fourier inversion. Its efficiency was evaluated by simulations of discontinuous spectrums with known experimental errors. Four case studies were proposed and its results were analyzed, showing MEM fundamental properties and its use.

# Sumário

<b>Introdução</b>	<b>1</b>
<b>1 Introdução a Proabilidade</b>	<b>4</b>
1.1 Probabilidade . . . . .	4
1.2 Variáveis aleatórias . . . . .	5
1.2.1 Variáveis Aleatórias Discretas . . . . .	5
1.2.2 Variáveis Aleatórias Contínuas . . . . .	5
1.3 Esperança Matemática . . . . .	6
1.4 Independência Estatística . . . . .	7
1.5 Probabilidade Conjunta . . . . .	7
1.6 Probabilidade Condicional . . . . .	8
1.7 A Lei da Probabilidade Total e a Regra de Bayes . . . . .	9
1.8 Verossimilhança . . . . .	10
1.9 Entropia . . . . .	13
<b>2 Fundamentos de Processamento de Sinais Aplicados ao Máximo de Entropia</b>	<b>16</b>
2.1 Sinais discretos no tempo . . . . .	16
2.2 O processo de amostragem . . . . .	17
2.3 Transformada discreta de Fourier (DFT) . . . . .	17
2.3.1 A Transformada de Fourier Discreta como uma Transformação Linear	18
2.4 Amostragem de sinais no domínio da frequência e reconstrução . . . . .	19
2.4.1 Amostragem em frequência . . . . .	20
2.4.2 Análise do sinal $x_s$ por Transformada de Fourier Discreta Inversa . . .	22
2.4.3 Limitação Experimental . . . . .	24
2.5 Conclusão . . . . .	26
<b>3 Máximo de Entropia</b>	<b>27</b>
3.1 Problemas inversos . . . . .	27
3.2 Análise de dados em medidas experimentais . . . . .	29
3.3 Conhecimento prévio e estratégia de análise de dados . . . . .	29
3.4 Análise de Dados Experimentais . . . . .	31
3.4.1 Transformação de Fourier . . . . .	31
3.4.2 Modelos para o Experimento: . . . . .	32
3.5 Introdução ao Máximo de Entropia . . . . .	33
3.5.1 Entropia de uma distribuição . . . . .	34
Maximização da Entropia . . . . .	37
3.6 Modelo uniforme ou conhecimento anterior? . . . . .	38



3.7	Discussão das soluções obtidas com o máximo de entropia . . . . .	39
3.8	O parâmetro Alfa . . . . .	39
3.9	Algoritmo do Máximo de Entropia . . . . .	40
3.9.1	Algoritmo do MEM . . . . .	41
3.10	Exemplo Simples : sistema linear com duas equações e duas incógnitas . . . .	41
<b>4</b>	<b>Aplicação do MME e Resultados</b>	<b>44</b>
4.1	Implementação Computacional do MME . . . . .	44
4.1.1	Interface Gráfica . . . . .	46
4.2	Análise do método de Inversão de Fourier a Máximo de Entropia . . . . .	48
4.2.1	Simulações e Resultados . . . . .	51
4.3	Conclusão . . . . .	55
	<b>Conclusão</b>	<b>60</b>
	<b>Anexos</b>	<b>62</b>
	<b>A Métodos Bayesianos e Máximo de Entropia: História e Filosofia</b>	<b>63</b>
	<b>B Inversão de Fourier a Máximo de Entropia</b>	<b>75</b>
	<b>Bibliografia</b>	<b>77</b>

# Lista de Tabelas

1.1	Valores da Probabilidade conjunta de $P(x, y)$ . . . . .	8
3.1	Número de possíveis distribuições de 10 cangurus. Aqui mostramos que sem informações suplementares sobre os cangurus, a configuração [2 3 2 3] é a mais provável. . . . .	36

# Lista de Figuras

1.1	A figura apresenta o gráfico de uma medida genérica $\bar{d}^{obs}$ calculado por meio de $n$ observações de um fenômeno físico. Como exemplo temos dois possíveis valores teóricos que podem ser preditos por meio da teoria do experimento. A verossimilhança nos dá a probabilidade de cada valor teórico $d^{th}$ em relação ao valor médio experimental $\bar{d}^{obs}$ . . . . .	11
1.2	O modelo teórico $D^{th}$ , cujo os valores se encontram dentro das barras de erros experimentais é aquele que a teoria descreve melhor o comportamento físico do experimento. . . . .	12
2.1	Exemplo do sinal $x_s$ contínuo e não periódico representado nos domínios da frequência e do tempo. . . . .	20
2.2	Sinal $X_{Sa}(f)$ é uma aproximação de $X_S(f)$ devido a discretização com passo $\Delta f$ . . . . .	21
2.3	Representação do sinal $X_{Sa}(f)$ no domínio do tempo - $x_{sp}(t)$ . . . . .	21
2.4	Exemplo da $DFT^{-1}$ do sinal $x_s$ . . . . .	22
2.5	(a) Sinal $x_s(n)$ com baixa resolução espectral ( $M_1 = 30, \Delta f_1 = \sigma/3$ ) estimado com diferentes valores para a resolução temporal ( $N_1 = 30, N_2 = 100, N_3 = 1000$ ) (b) Sinal $x_s(n)$ com alta resolução espectral ( $M_2 = 100, \Delta f_2 = \sigma/10$ ) estimado com diferentes valores para a resolução temporal ( $N_1 = 30, N_2 = 100, N_3 = 1000$ ) . . . . .	23
2.6	Análise espectral de um sinal simulando uma limitação experimental por meio da multiplicação por uma porta. . . . .	25
3.1	Diferença entre problemas diretos e problemas inversos. Em problemas diretos conhecemos a formulação analítica e a solução. Em problemas inversos a solução é dada por meio de métodos algorítmicos. . . . .	28
3.2	A teoria da experiência permite prever os dados medidos experimentalmente. O tratamento de dados é realizado no espaço de medidas nos levando ao espaço das conclusões. . . . .	30
3.3	(a) Distribuição de probabilidade de uma medida antes da realização do experimento (conhecimento a <i>priori</i> ). (b) Concordância das medidas realizadas com a teoria (verossimilhança). O Teorema de Bayes nos diz que se deve multiplicar (a) e (b) para obter a probabilidade a posteriori (c). Isto nos mostra como alterar a conclusão inicial a partir da realização da experiência. Pode-se observar, no gráfico (c), que a estimativa a <i>priori</i> não tem importância sobre o resultado final se as medidas são boas. . . . .	31
3.4	A entropia é a medida de desordem de um sistema fechado. A entropia é máxima numa configuração equiprovável deste sistema. Em configurações mais ordenadas, a entropia toma valores menores. Na realidade, existem diversas configurações mas as probabilidades não as mesmas e portanto, a mais provável é a que possui maior entropia. . . . .	35
3.5	Três possíveis distribuições em 4 categorias de 10 cangurus. As distribuições (a) e (b) correspondem a uma configuração [1 4 3 2] e a (c) à configuração [2 3 2 3]. . . . .	36

3.6	Exemplo da utilização do método do máximo de entropia por um sistema linear com duas incógnitas. Os cálculos são realizados a partir de um “modelo”: $\mathbf{m} = (1, 1)$ , os gráficos de superfícies representa o $\chi^2(\mathbf{f})$ e $S(\mathbf{f}, \mathbf{m})$ . As curvas em traços fortes são as trajetórias das soluções aproximadas; a solução final a para a qual $\chi^2 - 1 < \xi$ . Todos os valores contidos na elipse são soluções aceitáveis. A solução obtida é $\hat{\mathbf{f}}_{1,1} = (4.34, 3.02)$ . A cada iteração, a nova solução aproximada é a que possui maior entropia . . . . .	43
4.1	Diagrama da função <b>idftme</b> construída no MATLAB, com os parâmetros de entrada [ <b>DObs</b> , <b>SigDObs</b> , <b>modelo</b> , <b>R</b> , <b>LimConv</b> , <b>NIterMax</b> ] e os respectivos parâmetros de saída [ <b>S</b> , $\chi^2$ , <b>f</b> , <i>FlagFim</i> ]. . . . .	44
4.2	O fluxograma apresenta os passos computacionais para o cálculo da inversão de Fourier pelo método do máximo de entropia. . . . .	45
4.3	A interface gráfica foi elaborada para permitir ao usuário uma melhor visualização dos resultados alcançados pelo programa <b>idftme</b> . . . . .	47
4.4	Diagrama de blocos da simulação para a avaliação do programa <b>idftme</b> . . . . .	48
4.5	Representação do sinal $x_s$ no tempo e em frequência normalizada. . . . .	49
4.6	A relação sinal ruído em função da barra de erro $p$ mostra que o resultado do MME é melhor .	53
4.7	Relação sinal ruído em função da barra de erro $p$ . A solução MME fornece sempre o melhor resultado. . . . .	54
4.8	Inversão de Fourier com a utilização do máximo de entropia. As curvas (a) e (b) possuem barra de erro de 0.5%; as curvas (c) e (d) com barras de erro 5%. O cálculo é realizado conservando as 25 primeiras componentes de Fourier. O MME atribui valores às componentes que não foram fornecidas (altas frequências $k > 0.19$ ). O erro da medida é um grau de liberdade utilizado por MME para aproximar a solução final do modelo inicial uniforme.(Frequência normalizada = $k/128$ ). . . . .	56
4.9	A figura apresenta as estimações de $\hat{x}_s(n)$ se a componente de Fourier que descreve o cosseno encontra-se fora do jogo de dados. A solução do MME calculada com um modelo uniforme está de acordo com o jogo de dados. O MME estima apenas as componentes da gaussiana. .	57
4.10	Reconstrução da curva utilizando a função $x_s$ como modelo não uniforme. Qualquer que seja a amplitude da barra de erro, a solução do MME é perfeita e a entropia final é máxima. Evitamos fornecer as componentes que correspondem ao cosseno. Se o modelo contém informação que está em acordo com o conjunto de dados, o MME não tem razões para retirá-la.	58
4.11	Reconstrução da curva utilizando a gaussiana como modelo não uniforme. Quando o conjunto de dados não possui a componente de Fourier do cosseno, o resultado é perfeito e a entropia é máxima $S = 0$ . No caso contrário, o modelo não está completamente de acordo com o conjunto de dados $S = -1, 24$ . A entropia permite verificar a imprecisão do modelo.	59

# Introdução

Nos últimos anos, uma parte das atividades do grupo de processamento de sinais (LPS) do CBPF esteve dedicada ao desenvolvimento de técnicas de análise de sinais baseadas nos novos conceitos de entropia propostas pelo Grupo de Física Estatística deste Centro. Mais recentemente o LPS começou a estudar técnicas de reconstrução de sinais obtidos em instrumentos científicos que realizam medidas no domínio de Fourier (e.g. espectrômetros, difratômetros, tomógrafos por raios X, etc). Nestes instrumentos, técnicas entrópicas de reconstrução de sinais podem ser aplicadas, o que motivou o estudo da resolução espectral de instrumentos para a reconstrução de sinais medidos no espaço de Fourier.

Em instrumentação científica, encontramos frequentemente diversos problemas na realização de medidas de dados experimentais. Nos casos onde estes dados são componentes de frequência, é muito comum o uso da inversão de Fourier clássica a fim de estimar o sinal no espaço recíproco. Em tais situações os instrumentos de medida introduzem uma discretização do espectro, além de terem limitações técnicas para realizar medidas em todo o espectro. Dessa forma, o problema passa a estar relacionado a síntese de Fourier e a estimação de uma função a partir de um conhecimento parcial do seu domínio.

Quando nos referimos a síntese de Fourier, estamos nos referindo a utilização da Transformada de Fourier Discreta Inversa (**idft**). No entanto, a utilização da **idft** em dados experimentais tem dois inconvenientes básicos: a limitação dos dados experimentais e a eliminação das barras de erros experimentais no seu cálculo. O primeiro inconveniente faz com que as componentes de Fourier que não foram medidas, sejam colocadas em zero e, na maioria das vezes, isto é feito sem nos darmos conta de sua conseqüência. J. W. Gibbs [1] mostrou, em 1899, que essa limitação do espectro insere diversas oscilações no sinal estimado, e tais oscilações não são próprias do fenômeno em estudo e sim da transformação, ocasionando inferências equivocadas no sinal reconstruído. O segundo faz com que utilizemos os dados experimentais como se estes fossem precisos.

Diversos autores propuseram técnicas para melhorar a estimação do sinal quando se utiliza a **idft**. Existem algumas propostas para diminuir as distorções introduzidas pela técnica na reconstrução do sinal. Tentativas como a multiplicação do espectro por uma função Janela  $W$  (e.g. Janela de Hanning, Hamming, Kaiser, entre outras) buscam minimizar este efeito, porém, acabam por introduzir outras oscilações devido à convolução do sinal no espaço direto com a transformada de Fourier da janela. Devemos estar cientes que as oscilações não podem ser totalmente eliminadas, e o que se procura fazer é minimizá-las ao máximo para que estas não interfiram no processo de análise e conclusão dos resultados.

Esta tese aborda uma forma alternativa de reconstrução de sinais utilizando o método de inversão de Fourier à Máximo de Entropia (**idftme**). Este método é baseado no Teorema de Bayes que nos permite, por meio de probabilidades condicionais, descrever a realização de uma experiência como sendo um processo de aprendizagem. As probabilidades condicionais são decompostas nas funções da verossimilhança e da entropia. A entropia é calculada por meio de um modelo matemático que descreve os resultados esperados (informação *a priori*) do experimento. Dois conceitos de entropia são utilizados: a entropia clássica de Shannon e a entropia generalizada proposta por Shannon-Jaynes.

Um algoritmo iterativo para maximização de funções é utilizado a fim de ajustar um modelo teórico aos dados experimentais <sup>1</sup>. O resultado final é aquele que está de acordo com os dados experimentais e tem a maior entropia. A entropia nos dá uma informação importante, ela nos indica quantitativamente o quanto nos afastamos do resultado esperado, i.e. qual o novo conhecimento que a experiência nos trouxe. Esta técnica permite também a utilização da incerteza experimental no processo de reconstrução do sinal. Além disso, para os casos onde não temos como determinar o modelo inicial, podemos utilizar um modelo uniforme, i.e. quando não temos nenhuma informação *a priori* sobre o resultado esperado da reconstrução. Para os casos onde temos dúvida entre vários modelos ela nos permite selecionar àquele que é o mais provável. Cabe lembrar que este processo de refinamento de um modelo requer um cuidado especial, haja visto que, uma escolha errada implica em conclusões equivocadas sobre o experimento.

Nesse sentido, desenvolvemos uma ferramenta para a reconstrução de sinais que calcula a inversão de Fourier a máximo de entropia. Dados experimentais podem ser apresentados ao programa em conjunto com um modelo teórico do resultado esperado e estes são invertidos para o espaço recíproco. Para esta transformação implementamos computacionalmente o Método de Newton-Raphson a fim de maximizar a entropia e ajustar o modelo aos dados experimentais. A técnica foi implementada em MATLAB devido a esta ser uma linguagem de programação apropriada para cálculos computacionais envolvendo matrizes e/ou vetores. A programação em MATLAB possibilita um bom desempenho do algoritmo, sub-rotinas compactas e fácil visualização dos resultados.

A seguir apresentamos breves comentários sobre a estrutura escolhida para esta tese:

**Capítulo 1 - Introdução a probabilidade:** Este capítulo possui o objetivo de introduzir as bases teóricas do trabalho, abordando conceitos de probabilidade, tais como: probabilidades condicionais, Teorema de Bayes, entropia, entropia relativa, entre outros.

**Capítulo 2 - Fundamentos de processamento de sinais aplicados ao máximo de entropia:** Este capítulo aborda os conceitos da transformada de Fourier, colocando em evidência as limitações da **idft** na reconstrução de um sinal.

---

<sup>1</sup>Este tipo de abordagem é denominado de “Problemas Inversos”, i.e. aqueles que tentam estimar um determinado parâmetro maximizando a informação desejada e minimizando, por exemplo, a influência do erro experimental. Para os casos onde conhecemos a formulação analítica podemos calcular a solução diretamente. Entretanto, em problema inverso, conhecemos a formulação analítica, porém não podemos calcular a solução direta, sendo necessário a utilização de métodos algorítmicos não analíticos para resolvê-los.

**Capítulo 3 - Máximo de Entropia:** Este capítulo abrange o princípio do máximo de entropia apresentando desde de seu conceito histórico até a formulação matemática do problema inverso em questão. O capítulo discute também, de uma forma didática, a aplicação do método do máximo de entropia na reconstrução de sinais e a escolha do modelo mais provável. Por fim, o capítulo apresenta um algoritmo para a reconstrução de sinais por meio da transformada de Fourier a máximo de entropia (**idftme**).

**Capítulo 4 - Aplicação e Resultados:** Este capítulo apresenta a implementação da técnica do máximo de entropia em um programa em MATLAB. Esta técnica é avaliada por meio de simulações para diversos estudos de casos. O capítulo apresenta detalhadamente discussões dos resultados com o objetivo de verificar a eficiência do método do máximo de entropia e evidenciar suas propriedades fundamentais.

**Capítulo 5 - Conclusão:** Conclusão geral do trabalho com propostas de aplicações e sugestões para sua continuidade.

Ao final dois anexos são apresentados. O primeiro faz uma revisão histórica e filosófica dos métodos bayesianos e do máximo de entropia. O segundo apresenta o código que calcula a inversão de Fourier à máximo de entropia desenvolvido em MATLAB.

# Capítulo 1

## Introdução a Probabilidade

Praticamente qualquer sistema (físico ou não) está sujeito a complicadas influências que não podem ser inteiramente conhecidas. Estas influências, tipicamente associadas a um grande número de graus de liberdade envolvidos, impedem prever com precisão arbitrária o estado do sistema em cada instante. Assim, o resultado de uma experiência programada para medir um dado observável tem uma componente que flutua ao se repetir à experiência, mesmo sendo esta preparada sob condições praticamente idênticas. Apesar dessa imprevisibilidade, que caracteriza o fenômeno como sendo aleatório, quando a experiência é repetida um grande número de vezes aparece regularidades. Isto permite a formulação de leis matemáticas onde os conceitos de probabilidade, são fundamentais.

### 1.1 Probabilidade

A Teoria da probabilidade constitui uma importante área da Matemática, dotada de sólida base axiomática e que experimentou um vertiginoso crescimento desde a segunda metade do século XX. Oferece uma modelagem para uma classe de fenômenos sujeitos a incerteza, nos quais a idéia de regularidade estatística aparece como ponto importante.

Um experimento aleatório é tal que não é possível afirmar *a priori* o resultado que ocorrerá, podendo o resultado ser diferente mesmo ao se repetir o ensaio em condições praticamente inalteradas. Os resultados podem parecer erráticos nas primeiras tentativas, entretanto, após um grande número de repetições, aparecem regularidades.

Quando lidamos com fenômenos aleatórios, podemos, porém conhecer, em geral, o conjunto dos possíveis resultados a serem observados ao realizar uma dada experiência. A partir de um modelo, podemos também atribuir aos resultados ou conjuntos de resultados possíveis, números que representem suas chances de ocorrência. Estes números, não negativos e somando um para todos os possíveis resultados excludentes, são denominados probabilidades. O modelo pode ser construído seja a partir da frequência de ocorrência observada em um grande número de experimentos passados (Lei dos grandes números) ou teoricamente *a priori*.



## 1.2 Variáveis aleatórias

Em um experimento aleatório muitas vezes o resultado é expresso de forma numérica, isto é, declarando os valores supostos no experimento por um certo número de quantidades variáveis, isto acontece sempre que o experimento se acha diretamente relacionado com a contagem ou medição de variáveis quantitativas [Mil03].

Uma variável aleatória (VA) é uma quantidade variável que expressa o resultado de um determinado experimento aleatório.

### 1.2.1 Variáveis Aleatórias Discretas

Seja  $x$  uma variável aleatória que pode tomar um número finito  $m$  de valores diferentes no conjunto  $X = \{\nu_1, \nu_2, \dots, \nu_m\}$ . Designa-se por  $p_i$  a probabilidade de  $x$  tomar o valor  $\nu_i$ :

$$p_i = Pr [x = \nu_i], i = 1, \dots, m. \quad (1.1)$$

As probabilidades  $p_i$  satisfazem as seguintes condições:

$$p_i \geq 0 \quad e \quad \sum_{i=1}^m p_i = 1 \quad (1.2)$$

Por vezes é conveniente exprimir o conjunto de probabilidades  $\{p_1, p_2, \dots, p_m\}$  em termos da *função de massa de probabilidade* (ou distribuição de probabilidades discreta)  $P(x)$ , que satisfaz as seguintes condições:

$$P(x) \geq 0 \quad e \quad \sum_{i=1}^m P(x) = 1 \quad (1.3)$$

### 1.2.2 Variáveis Aleatórias Contínuas

Quando uma variável aleatória  $x$  pode tomar valores no domínio contínuo (infinitos valores), não faz sentido falar da probabilidade de  $x$  ser igual a determinado valor, por exemplo,  $Pr [x = 2.15]$ , pois a probabilidade de um valor em particular é sempre nula (ou quase sempre). Assim, faz sentido falar da probabilidade de  $x$  tomar valores num determinado intervalo  $[a, b]$ ; em vez de termos a função de massa de probabilidade  $P(x)$ , temos a *função de densidade de probabilidade*  $p(x)$ . Esta função tem a propriedade de

$$Pr [x \in [a, b]] = \int_a^b p(x) dx \quad (1.4)$$

O nome *densidade* vem da analogia à densidade de massa. Se considerarmos um intervalo pequeno  $[a, a + \Delta x]$  sobre o qual a função  $p(x)$  é essencialmente constante, tendo o valor  $p(a)$ , vemos que  $p(a) = Pr [x \in [a, b]] / \Delta x$ . Ou seja, a densidade de probabilidade em  $x = a$  é a massa de probabilidade  $p(a) = Pr [x \in [a, b]]$  por unidade de distância. Assim, a função densidade de probabilidade satisfaz as condições [Mil03].

$$p(x) \geq 0 \quad e \quad \int_{-\infty}^{+\infty} p(x) dx = 1 \quad (1.5)$$

A esperança matemática, a média e a variância de uma variável aleatória contínua são respectivamente definidas por

$$E[f(x)] = \int_{-\infty}^{+\infty} f(x)p(x)dx \quad (1.6)$$

$$\mu_x = E[x] = \int_{-\infty}^{+\infty} x p(x)dx \quad (1.7)$$

$$var(x) = \sigma^2 = E[(x - \sigma)^2] = \int_{-\infty}^{+\infty} (x - \sigma)^2 p(x)dx \quad (1.8)$$

### 1.3 Esperança Matemática

A esperança, média ou valor médio da v.a.  $x$  é definido por

$$E[x] = \mu_x = \sum_{x \in X} xP(x) = \sum_{i=1}^m \nu_i p_i \quad (1.9)$$

Considerando a função de massa de probabilidade como um conjunto de pontos de massa, sendo  $p_i$  a massa concentrada em  $x = \nu_i$ , então o valor esperado  $\mu_x$  é o centro de massa. Alternativamente, pode-se interpretar  $\mu_x$  como a média aritmética dos valores de um conjunto significativo de amostras de  $x$ . Genericamente, se  $f(x)$  é uma função de  $x$ , o valor esperado de  $f$  é definido por

$$E[f(x)] = \sum_{x \in X} f(x)P(x) \quad (1.10)$$

A esperança matemática apresenta a propriedade de linearidade

$$E[\alpha_1 f_1(x) + \alpha_2 f_2(x)] = \alpha_1 E[f_1(x)] + \alpha_2 E[f_2(x)] \quad (1.11)$$

Sendo  $\alpha_1$  e  $\alpha_2$  constantes arbitrárias.

Dois casos especiais de valores esperados são o segundo momento (momento de 2ª ordem) e a variância:

$$E[x^2] = \sum_{x \in X} x^2 P(x) \quad (1.12)$$

$$var(x) = \sigma^2 = E[(x - \mu)^2] = \sum_{x \in X} (x - \mu)^2 P(x), \quad (1.13)$$

Onde  $\sigma_x$  é o *desvio-padrão* de  $x$ . A variância nunca toma valores negativos e só pode ser nula se a função de massa está centrada num único ponto.

O desvio-padrão é uma medida de dispersão dos valores de  $x$  em torno da média. O seu nome sugere que é a quantidade típica esperado que uma saída aleatória de  $x$  se desvie, ou difira de  $\mu$ .

Expandindo a Eq(1.13), obtemos

$$\sigma^2 = E[x^2] - \mu_x^2 \quad (1.14)$$

Note que, ao contrário da média (valor esperado de  $x$ ), a variância *não* é linear. Em particular, se  $y = \alpha x$ , onde  $\alpha$  é uma constante, então  $\text{var}[y] = \alpha^2 \text{var}[x]$ . Além do mais, a variância da soma de duas variáveis aleatórias, não é igual à soma das variâncias, de um modo geral. No entanto, veremos mais à frente que as variâncias se somam quando as variáveis aleatórias em causa são estatisticamente independentes.

No caso de  $x$  ser uma variável aleatória binária, que toma valores  $\nu_1 = 0$  e  $\nu_2 = 1$ , pode-se obter fórmulas simples para  $\mu$  e  $\sigma$ . Sendo  $p = \text{Pr}[x = 1]$ , temos

$$\mu = p \quad e \quad \sigma = \sqrt{p(1-p)} \quad (1.15)$$

## 1.4 Independência Estatística

As variáveis  $x$  e  $y$  dizem-se estatisticamente independentes se é somente se

$$P(x, y) = P_x(x)P_y(y) \quad (1.16)$$

Pode-se entender a independência estatística do seguinte modo. Suponha que  $p_i = \text{Pr}[x = \nu_i]$  é uma fração de tempo que  $x = \nu_i$  e que  $q_j = \text{Pr}[y = w_j]$  é a fração de tempo que  $y = w_j$ . Considere as situações em que  $x = \nu_i$ . Se continuar a ser verdade que a fração de tempo em que  $y = w_j$  tem o mesmo valor  $q_j$ , então se conclui que conhecer o valor de  $x$  não trouxe informação adicional sobre os possíveis valores de  $y$ ; neste sentido  $y$  é independente de  $x$ . Finalmente, se  $x$  e  $y$  são estatisticamente independentes, a fração de tempo que um específico par de valores  $(\nu_i, w_j)$  ocorre vem igual ao produto das duas frações  $p_i q_j = P(\nu_i)P(w_j)$  como será explorado mais adiante [Mil03].

## 1.5 Probabilidade Conjunta

Sejam  $x$  e  $y$  duas v.a. que tomam valores em  $X = \nu_1, \nu_2, \dots, \nu_m$  e  $Y = w_1, w_2, \dots, w_m$ , respectivamente. Considerando o par  $(x, y)$  como um vetor do espaço  $\mathbb{R}^2$ , para cada possível par de valores  $(\nu_i, w_j)$  temos a *probabilidade conjunta*  $p_{ij} = \text{Pr}[x = \nu_i, y = w_j]$ . Estas  $mn$  probabilidades conjuntas são representadas através da *função de massa de probabilidade conjunta*  $P(x, y)$ , tal que

$$P(x, y) \geq 0 \quad e \quad \sum_{x \in X} \sum_{y \in Y} P(x, y) \quad (1.17)$$

A função de massa de probabilidade conjunta representa completamente o par de variável  $(x, y)$ ; ou seja, toda a informação das v.a.  $x$  e  $y$ , individualmente ou em conjunto, pode ser extraída de  $P(x, y)$ . Em particular, extraem-se as *distribuições marginais*<sup>1</sup>

$$P_x(x) = \sum_{y \in Y} P_{xy}(x, y) \quad (1.18)$$

<sup>1</sup>Dada uma função de probabilidade conjunta  $P(x, y)$ , a distribuição de probabilidade de  $x$  isolado é obtida pela soma dos valores de  $P(x, y)$  ao longo de  $y$ . Do mesmo modo, a distribuição de probabilidade de  $y$  isolado é dada pela soma dos valores de  $P(x, y)$  ao longo de  $x$ . Assim a probabilidade marginal de  $x$  é a soma das probabilidades conjuntas de  $x$  em relação a  $y$ , da mesma forma calculamos a probabilidade marginal de  $y$

$$P_y(x) = \sum_{x \in X} P_{xy}(x, y) \quad (1.19)$$

Para tornar a compreensão das Equações (1.18) e (1.19) vamos analisar o seguinte exemplo [TT91]: Dada a uma determinada distribuição conjunta  $P(X, Y)$  representada da seguinte forma:

Y \ X	1	2	3	4
1	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$
2	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{32}$
3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
4	$\frac{1}{4}$	0	0	0

Tabela 1.1: Valores da Probabilidade conjunta de  $P(x, y)$ .

Com base na Tabela (1.1), encontramos os seguintes valores para  $P_x(x)$  e  $P_y(y)$ :  
 $P_x(x) = (\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8})$  e  $P_y(y) = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$

Casualmente usam-se índices, como na Eq(1.19), para realçar o fato de  $P(x)$  ter significado diferente de  $P(y)$ . É comum escrever simplesmente  $P(x)$  e  $P(y)$  quando o contexto torna claro que se trata de duas funções diferentes - e não a mesma função com diferentes valores de argumento. Usa-se a mesma regra para a média  $\mu$ , o desvio-padrão  $\sigma$  e a variância  $\sigma^2$ , como na Eq.1.14, para realçar o fato de serem medidas da v.a.  $x$ .

## 1.6 Probabilidade Condicional

Quando duas v.a. são estatisticamente dependentes, sabendo o valor de uma delas permitenos obter uma melhor estimativa do valor da outra. Este conhecimento vem expresso pela seguinte definição de *probabilidade condicional* de  $x$  dado  $y$  [Mil03]:

$$Pr[x = \nu_i | y = w_j] = \frac{Pr[x = \nu_i, y = w_j]}{Pr[y = w_j]} \quad (1.20)$$

Ou em termos das funções de *massa de probabilidade*,

$$P(x | y) = \frac{P(x, y)}{P(y)} \quad (1.21)$$

Note que, se  $x$  e  $y$  são estatisticamente independentes então  $P(x | y) = P(x)$ . Ou seja, quando  $x$  e  $y$  são independentes, conhecer o valor de  $y$  não nos fornece mais informação acerca de  $x$  além da que já tínhamos através da sua distribuição marginal  $P(x)$ .

Considere o exemplo de duas v.a. binárias  $x$  e  $y$  em que ambas tomam os valores 0 ou 1. Suponha que são produzidas aleatoriamente, em elevado número,  $n$  amostras dos pares  $(x, y)$ . Seja  $n_{ij}$  o número de pares  $(x = i, y = j)$ , ou seja,  $n_{00}$  é o número de vezes que saiu o par  $(0, 0)$ ,  $n_{10}$  é o número de vezes que saiu o par  $(1, 0)$ , e assim sucessivamente, tal que  $n_{00} + n_{10} + n_{01} + n_{11} = n$ . Se considerarmos apenas os pares em que  $y = 1$  - isto é, os pares  $(0, 1)$  e  $(1, 1)$  - então a fração dos casos em  $x$  também vem igual a 1 é

$$\frac{n_{11}}{n_{01} + n_{11}} = \frac{n_{11}/n}{(n_{01} + n_{11})/n} \quad (1.22)$$

Intuitivamente, o valor desta fração é o que gostaríamos de obter para  $P(x | y)$  quando  $y = 1$  e  $n$  é muito grande. De fato, é o que se obtém, pois  $n_{11}/n$  é aproximadamente igual a  $P(x, y)$  e  $(n_{01} + n_{11})/n$  é aproximadamente igual a  $P(y)$  para valores elevados de  $n$ .

## 1.7 A Lei da Probabilidade Total e a Regra de Bayes

A *Lei da Probabilidade Total* diz que se um acontecimento  $A$  ocorrer em  $m$  condições diferentes  $A_1, A_2, \dots, A_m$  e estes  $m$  sub-acontecimentos forem mutuamente exclusivos - ou seja, não ocorrem simultaneamente - então a probabilidade de ocorrer  $A$  é a soma das probabilidades dos sub-acontecimentos  $A_i$ . Em particular, a v.a.  $y$  pode tomar o valor  $y$  em  $m$  condições diferentes - com  $x = \nu_1, x = \nu_2, \dots, x = \nu_m$ . Porque estas condições são simultaneamente exclusivas, deduz-se a *Lei da Probabilidade Total* que  $P(y)$  é a soma das probabilidades conjuntas  $P(x, y)$  sobre todos os valores possíveis de  $x$ . Formalmente tem-se que

$$P(y) = \sum_{x \in X} P(x, y) \quad (1.23)$$

Como, através da definição de probabilidade condicional  $P(y | x)$ , tem-se

$$P(x, y) = P(y | x)P(x) \quad (1.24)$$

Então, reescrevendo a Eq(1.23), a  $P(y)$ , vem igual a

$$P(y) = \sum_{x \in X} P(y | x)P(x) \quad (1.25)$$

Substituindo na Eq(1.21) as probabilidades  $P(x, y)$  e  $P(y)$ , definidas respectivamente nas Eqs.(1.23) e (1.25), vem

$$P(x | y) = \frac{P(y | x)P(x)}{\sum_{x \in X} P(y | x)P(x)} \quad (1.26)$$

Em outras palavras, tem-se

$$posteriori = \frac{likelihood \times priori}{evidencia} \quad (1.27)$$

A Eq(1.26) é designada *Regra de Bayes*. Note que o denominador é, de acordo com a Eq(1.25) igual a  $P(y)$ . Desta forma percebemos que todos os termos do lado direito da equação são condicionados por  $x$ . Se considerarmos  $x$  uma *v.a.* significativa, então podemos dizer que a forma da distribuição  $P(x | y)$  depende apenas do numerador  $P(y | x)P(x)$ ; o denominador é o fator de normalização, por vezes designado por *evidência*, para garantir que a soma de  $P(x | y)$  seja igual a um.

A interpretação mais freqüente da regra de *Bayes* é inverter ligações estatísticas, tornando  $P(y | x)$  em  $P(x | y)$ . Considere que  $x$  é uma “causa” e  $y$  um “efeito” da causa  $x$ . Assumindo que a causa  $x$  está presente, é fácil determinar a probabilidade do efeito  $y$  ser observado; a função de probabilidade condicional  $P(y | x)$  - *função de verossimilhança* (*likelihood* em inglês) - representa esta probabilidade explicitamente. Ao contrário, se observarmos o efeito  $y$ , pode não ser tão fácil determinar a causa  $x$ , pois haverá diferentes causas, podendo cada uma delas produzir o mesmo efeito observado. No entanto, a regra de *Bayes* torna fácil a determinação de  $P(x | y)$ , considerando que são conhecidas  $P(y | x)$  e  $P(x)$ , sendo que esta última é designada por *probabilidade a priori* e que exprime a probabilidade de  $x$  antes de observarmos qualquer valor de  $y$ . Ou seja, a regra de *Bayes* mostra como a distribuição de probabilidade de  $x$  se altera desde *distribuição a priori*  $P(x)$ , antes de se observar  $y$ , até a *distribuição posterior*  $P(x | y)$ , depois de se observar o valor de  $y$ .

## 1.8 Verossimilhança

Do ponto de vista bayesiano o princípio da verossimilhança é uma consequência direta do teorema de Bayes, desta forma a verossimilhança pode ser considerada como probabilidade condicional.

Em física experimental a verossimilhança é utilizada para inferir sobre uma medida a partir da distribuição de probabilidade de um conhecimento anterior. Por meio da verossimilhança, é possível responder a seguinte pergunta: Qual a probabilidade de que o evento (medida) que já ocorreu (que possui uma determinada função de densidade de probabilidade) volte a ocorrer?

### *Medidas Experimentais*

Fenômenos físicos são observados por meio da realização de um experimento que fornece um conjunto de dados representando um comportamento em função de uma grandeza acessível. Na prática, o que fazemos é calcular o valor médio  $\bar{d}$  de  $n$  medidas e seu desvio padrão, sendo representado por  $\bar{d} \pm \sigma$ .

Na observação de um fenômeno físico raramente estamos interessados na medida por si só, o que freqüentemente nos interessa é testar uma teoria a fim de compará-la com outras medidas de outros experimentos e usar este parâmetro para ajudar a prever outros experimentos. Desta forma o valor do erro experimental torna-se crucial para a análise do resultado. O valor numérico do erro experimental nos dá uma indicação do quão longe estamos da resposta correta para o experimento.

Para compreendermos melhor a verossimilhança vamos analisar o seguinte exemplo. Em uma experiência, medimos várias vezes um mesmo valor,  $d^{obs}$  e calculamos o  $\bar{d}^{obs} \pm \sigma_{d^{obs}}$ , seguindo uma distribuição de probabilidade gaussiana.

A teoria do experimento nos permite calcular analiticamente o  $d^{th}$ , isto é o valor teórico para  $d$ . A verossimilhança pode ser calculada entre o valor observado e o teórico por meio de,

$$P(\bar{d}^{obs} | d^{th}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(\bar{d}^{obs} - d^{th})^2}{\sigma^2}} \quad (1.28)$$

A Figura 1.1 apresenta um gráfico de dois possíveis valores teóricos para um valor observado experimentalmente  $\bar{d}^{obs}$ . Nesta figura podemos observar que entre os dois valores teóricos, o que se encontra mais próximo da medida é aquele que tem a maior verossimilhança.

$$P(\bar{d}^{obs} | d_1^{th}) > P(\bar{d}^{obs} | d_2^{th}) \quad (1.29)$$

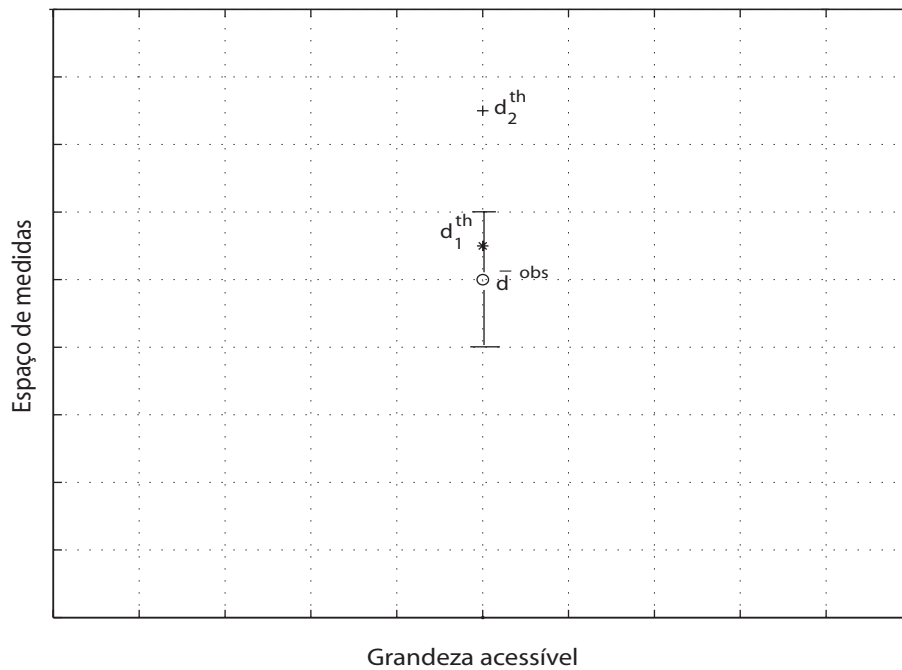


Figura 1.1: A figura apresenta o gráfico de uma medida genérica  $\bar{d}^{obs}$  calculado por meio de  $n$  observações de um fenômeno físico. Como exemplo temos dois possíveis valores teóricos que podem ser preditos por meio da teoria do experimento. A verossimilhança nos dá a probabilidade de cada valor teórico  $d^{th}$  em relação ao valor médio experimental  $\bar{d}^{obs}$ .

Agora vamos supor que foi realizado um experimento, no qual foram feitas  $l$  medidas, com seus respectivos erros experimentais, representadas por  $D^{obs} = \{\bar{d}_1^{obs}, \bar{d}_2^{obs}, \bar{d}_3^{obs}, \dots, \bar{d}_l^{obs}\}$  e

$\sigma = \{\sigma_1, \sigma_2, \sigma_3, \dots, \sigma_l\}$ . Podem existir vários modelos teóricos que descrevem o comportamento de um experimento  $D$ , representadas por  $D_m^{th} = \{d_{m_1}^{th}, d_{m_2}^{th}, d_{m_3}^{th}, \dots, d_{m_l}^{th}\}$ . Onde  $m$  representa os vários modelos teóricos que descrevem o experimento.

A Figura 1.2 apresenta um exemplo de um conjunto de dados observados  $D^{obs}$  em um experimento e os dois possíveis modelos teóricos  $D_m^{th}$  para os dados

Para o cálculo da verossimilhança de um conjunto de medidas que seguem distribuição de probabilidade do tipo gaussiana usamos a expressão

$$P(D | f) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_{d_{obs}^i}} e^{-\frac{1}{2}\chi^2} \quad (1.30)$$

onde

$$\chi^2 = \sum_{i=1}^n \frac{(d_i^{obs} - d_m^{th})^2}{\sigma_{d_{obs}^i}^2} \quad (1.31)$$

e

$$P(\bar{D}^{obs} | D_1^{th}) > P(\bar{D}^{obs} | D_2^{th}) \quad (1.32)$$

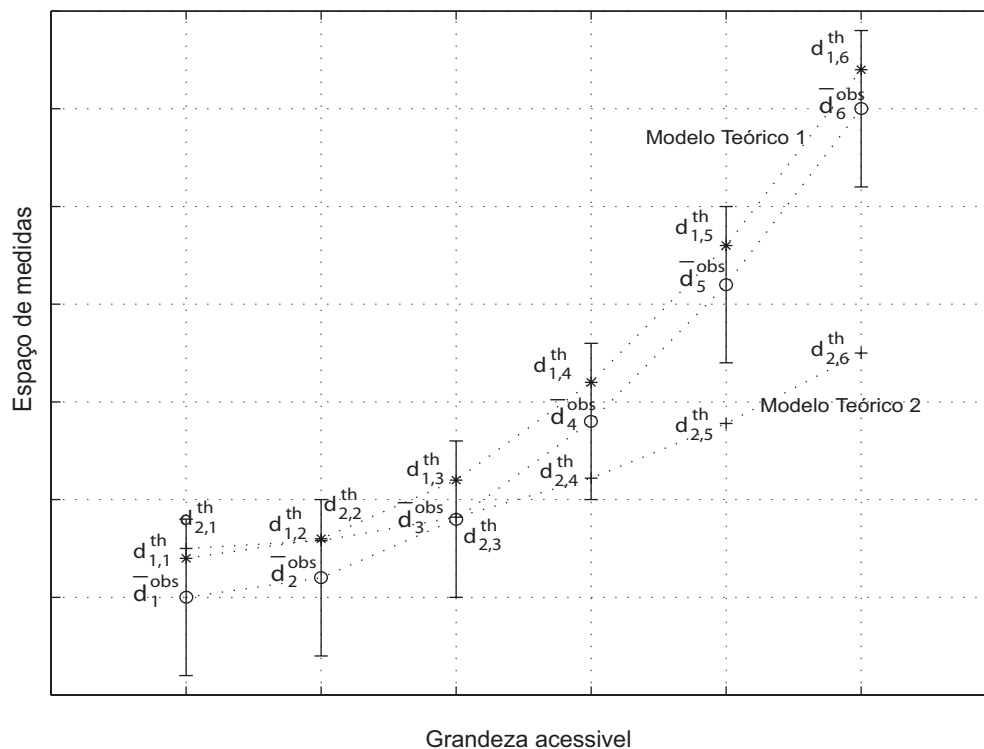


Figura 1.2: O modelo teórico  $D^{th}$ , cujo os valores se encontram dentro das barras de erros experimentais é aquele que a teoria descreve melhor o comportamento físico do experimento.



## 1.9 Entropia

Originária de estudos de termodinâmica, onde foi introduzida para caracterizar o grau de desordem de um sistema, a noção de entropia já foi objeto de muitas controvérsias e distintas formulações. O conceito de entropia foi introduzido há mais de 150 anos, mas somente a partir de meados do Século XX é que se difundiram suas aplicações em diversas áreas do conhecimento. Na raiz deste movimento, estiveram os trabalhos de Shannon (1948), que introduziu um conceito de entropia em teoria da informação e uma medida para quantificá-la, e os estudos de Jaynes [Jay57] e Kullback (1959), que propuseram *princípios de otimização da entropia* segundo formulações distintas. Atualmente, diferentes áreas, como termodinâmica, probabilidade, estatística, pesquisa operacional, reconhecimento de padrões, dentre outras, vêm usando e desenvolvendo princípios de otimização da entropia.

O conceito de entropia de Shannon refere-se a incerteza de uma distribuição de probabilidade e a medida que propôs destinava-se a quantificar essa incerteza. Formalmente, o princípio de Jaynes envolve a busca pela distribuição de probabilidade que maximiza a medida de Shannon, dado um conjunto de restrições lineares. Estas restrições informam características da distribuição procurada, como por exemplo, sua média e variância. O princípio Kullback por sua vez, envolve a busca pela distribuição de probabilidade mais próxima de uma outra distribuição *a priori*, por meio da minimização de uma medida de divergência entre ambas, dado um mesmo conjunto de restrições. Tanto a medida de Shannon como a de Kullback são funções intrinsecamente não-lineares de probabilidade. Assim os princípios de Jaynes e Kullback reduzem-se a problemas de programação não-linear cuja solução demanda um algoritmo de busca iterativa.

### Medida de Entropia

Shannon (1948) derivou uma medida para quantificar o grau de incerteza de uma distribuição de probabilidade. A medida de entropia de Shannon denominada como  $S$  possui uma expressão formal para distribuições discretas de probabilidade dada por:

$$S(\mathbf{p}) = - \sum_{i=1}^n p_i \ln p_i \quad (1.33)$$

onde  $p_i$  é probabilidade referente a cada evento e o logaritmo na base 2 representa a unidade de medida de informação. Esta medida é sempre não-negativa e assume seu valor máximo  $S(\mathbf{p}) = 0$  quando a distribuição uniforme e seu mínimo ocorre quando há ausência de incerteza, (i.e., uma  $p_i = 1$  e as demais iguais a zero).

Por meio de um método axiomático, Shannon derivou essa medida de modo que ela refletisse certas características desejadas. As características de  $S$  mais relevantes são:

1.  $S(p_1, p_2, \dots, p_n)$  é uma função duas vezes diferenciável de  $p_1, p_2, \dots, p_n$ .
2.  $S(p_1, p_2, \dots, p_n)$  é simétrica em relação à permutação de  $p_1, p_2, \dots, p_n$ .
3.  $S(1/N, \dots, 1/N)$  é uma função monotonicamente crescente de  $N$ .
4.  $S$  é uma função estritamente côncava de  $p_1, p_2, \dots, p_n$ .

A primeira característica permite a aplicação de técnicas para maximização de funções diferenciáveis. A segunda significa que o valor final de  $S$  não se altera quando a ordem das probabilidades  $p_i$  sofrem uma inversão. A terceira característica significa que a entropia da distribuição uniforme (máxima entropia) cresce quanto maior for o número de resultados possíveis  $N$ . Por fim, a quarta característica é de especial relevância, pois esta garante a  $S$  um único máximo (global), mesmo sujeita a restrições lineares.

### Medida de Entropia Cruzada

A entropia relativa é a medida de uma distância estatística entre duas distribuições definidas sobre um mesmo espaço composto de todos eventos possíveis. Em estatística, isto significa o valor esperado do logaritmo da relação entre as probabilidades. A entropia relativa é definida como sendo

$$D_{KL}(p : p') = \sum_{i=1}^K p_i \cdot \log \frac{p_i}{p'_i} \quad (1.34)$$

É também conhecida como *Distância Kullback-Leiber*, *Entropia Kullback-Leiber* ou *Divergência I*. Na definição acima, assumimos por convenção (baseada em argumentos de continuidade) que  $0 \cdot \log \frac{0}{p'} = 0$  e  $p \cdot \log \frac{p}{0} = \infty$ .

A entropia relativa é sempre não negativa, satisfazendo a desigualdade de Gibbs:

$$\sum_{i=1}^K p_i \cdot \log \frac{1}{p_i} \leq \sum_{i=1}^K p_i \cdot \log \frac{1}{p'_i} \quad (1.35)$$

considerando  $p_i$  uma distribuição de probabilidades qualquer e  $p'_i$  uma outra distribuição que satisfaz a condição a seguir

$$\sum_{i=1}^K p'_i \leq 1 \quad (1.36)$$

Apesar da entropia relativa ser chamada de distância Kullback-Leiber, esta não pode ser considerada uma distância verdadeira entre duas distribuições. A entropia relativa não apresenta simetria,

$$\sum_{i=1}^K p_i \cdot \log \frac{p_i}{p'_i} \neq \sum_{i=1}^K p'_i \cdot \log \frac{p'_i}{p_i} \quad (1.37)$$

É, portanto não pode ser considerada uma distância métrica. Mesmo assim, é usual pensar na entropia relativa como uma medida de “distância” entre duas distribuições estatísticas [Esq03].

Assim como a entropia de Shannon a entropia relativa também apresenta características que são fundamentais para sua aplicação. Dentre elas destacam-se:

1.  $K(\mathbf{p} : \mathbf{q})$  é uma função duas vezes diferenciável de  $p_1, p_2, \dots, p_N$ ;
2.  $K(\mathbf{p} : \mathbf{q})$  é simétrica em relação à permutação dos pares  $(p_1, q_1), \dots, (p_N, q_N)$ ;

3.  $S$  é uma função estritamente convexa de  $p_1, p_2, \dots, p_N$ ;
4.  $K(\mathbf{p} : \mathbf{q}) \geq 0$  (não negatividade);
5.  $K(\mathbf{p} : \mathbf{q}) = 0$  se e somente se  $\mathbf{p} = \mathbf{q}$ ;

As três primeiras características possuem significados semelhantes a entropia de Shannon. As duas últimas características são próprias de distâncias métricas (no entanto, a entropia de Kullback não é uma métrica pois não atende às propriedades de simetria e desigualdade do triângulo que toda medida de distância tende a apresentar para ser métrica). Da mesma forma que a medida de Shannon, a medida de Kullback se presta a estudos de entropia em que as distribuições se refiram a proporções e não a probabilidades.

## Capítulo 2

# Fundamentos de Processamento de Sinais Aplicados ao Máximo de Entropia

A observação de fenômenos físicos na maioria das vezes se faz através de experimentos. Cada experiência tem sua particularidade, como, por exemplo, o tipo de instrumento de medida que deve ser utilizado.

Alguns instrumentos como magnetômetros, espectômetros e difratômetros são utilizados para medidas de propriedades magnéticas e de estruturas cristalinas de diversos materiais. Estes instrumentos nos fornecem como resultado um sinal característico, sendo que todas as informações relativas ao fenômeno em estudo estão contidas neste sinal. Em alguns casos, nos depararmos com um sinal ruidoso e incompleto, isto acontece por diversos motivos como por exemplo, limitação do instrumento utilizado para a realização da experiência. Diante disto é necessário a realização de um tratamento deste sinal.

Esta etapa de tratamento do sinal é feita com técnicas de processamento digital de sinais (PDS). Uma das técnicas de fazer isso é a transformação do sinal de sua representação base para uma representação em outro domínio, com o intuito de explicitar características relevantes do mesmo.

Este capítulo apresenta a reconstrução de sinais por meio da transformação de Fourier. Ao final os problemas da utilização da *IDFT* são abordados detalhadamente.

### 2.1 Sinais discretos no tempo

Em PDS os sinais são representados por seqüências de números, chamados amostras. O valor de uma amostra de um típico sinal discreto ou seqüência no tempo é denominado  $x(n)$ , com o argumento  $n$  sendo um inteiro que varia entre  $-\infty$  à  $\infty$ . Devemos ressaltar que  $x(n)$  é definido somente para valores inteiros de  $n$  e é indefinido para qualquer valor não inteiro de  $n$  [S98][JGP92].

Em algumas aplicações uma seqüência discreta no tempo é gerada pela amostragem periódica de um sinal contínuo no tempo. O espaçamento  $T$  entre duas amostras consecutivas é chamado de intervalo de amostragem ou período de amostragem.

## 2.2 O processo de amostragem

Uma seqüência em tempo discreto é obtida pela amostragem uniforme de um sinal contínuo no tempo. A relação entre esses dois sinais é dada pela equação:

$$x_s(n) = x_a(t)_{t=nT}, \quad n = \dots, -2, -1, 0, 1, 2, \dots \quad (2.1)$$

onde a variável  $t$  do sinal contínuo está relacionada com a variável  $n$  do sinal discreto somente nos instantes discretos  $t_n$  dado por:

$$t_n = nT = \frac{n}{F_T} = \frac{2\pi n}{\Omega_T} \quad (2.2)$$

onde  $F_T = 1/T$  e representa a freqüência de amostragem e  $\Omega_T = 2\pi F_T$  é denominada freqüência angular.

Um sinal periódico contínuo no tempo deve ser amostrado com freqüência superior a maior freqüência existente no sinal para não gerar erros de amostragem no sinal discreto. Esses erros aparecem sobre a forma de freqüências que não estavam presentes no sinal, chamadas de freqüências de sobreposição *aliasing*. Através do teorema de Nyquist podemos provar que a freqüência de amostragem deve ser no mínimo maior que duas vezes a freqüência máxima do sinal [JGP92].

## 2.3 Transformada discreta de Fourier (DFT)

Para compreendermos a Transformada de Fourier Discreta (*DFT*), vamos partir de um sinal aperiódico  $x_s(n)$  de energia finita amostrado no domínio da freqüência. Em geral, as amostras em freqüência são igualmente espaçadas  $X_S(2\pi k/N)$ ,  $k = 0, 1, \dots, N-1$ , e não representam o sinal  $x_s(n)$  original exceto quando o  $x_s(n)$  tem a duração infinita. Sendo assim, as amostras em freqüência  $X_S(2\pi k/N)$ ,  $k = 0, 1, \dots, N-1$ , correspondem a um  $x_{sp}(n)$  periódico com período  $N$ , desta forma o  $x_{sp}(n)$  é considerado uma versão de  $x_s(n)$ .

Em si tratando de seqüências discretas de comprimento finito  $0 \leq n \leq N-1$  como é o caso do sinal  $x_s(n)$ , a representação no domínio da freqüência é chamada de Transformada de Fourier Discreta (*DFT*) e esta é representada da seguinte forma:

$$X_S(k) = \sum_{n=0}^{N-1} x_s(n) e^{-j2\pi kn/N} \quad k = 0, 1, \dots, N-1 \quad (2.3)$$

A transformada discreta inversa de Fourier (*DFT*<sup>-1</sup>) é dada por,

$$x_s(n) = \frac{1}{N} \sum_{k=0}^{N-1} X_S(k) e^{j2\pi kn/N}, \quad n = 0, 1, \dots, N-1 \quad (2.4)$$

Podemos observar pelas Equações 2.3 e 2.4 que são necessárias aproximadamente  $N^2$  multiplicações complexas e  $N(N-1)$  adições complexas para o cálculo das transformadas  $DFT$  e  $DFT^{-1}$ . No entanto, métodos mais elegantes desenvolvidos reduzem para  $N(\log_2 N)$  operações. Estas técnicas são normalmente chamadas de *Fast Fourier Transform* (FFT)[JGP92].

### 2.3.1 A Transformada de Fourier Discreta como uma Transformação Linear

As fórmulas da  $DFT$  e  $DFT^{-1}$  dadas pelas Equações 2.3 e 2.4 podem ser expressas como:

$$X_S(k) = \sum_{n=0}^{N-1} x_s(n) W_N^{kn}, \quad k = 0, 1, \dots, N-1 \quad (2.5)$$

A transformada discreta inversa de Fourier ( $DFT^{-1}$ ) é dada por,

$$x_s(n) = \frac{1}{N} \sum_{k=0}^{N-1} X_S(k) W_N^{kn}, \quad n = 0, 1, \dots, N-1 \quad (2.6)$$

onde por definição,

$$W_N = e^{-j2\pi/N}, \quad (2.7)$$

Assim podemos perceber que o cálculo de cada ponto da  $DFT$  pode ser realizado por meio de  $N$  multiplicações e  $N(N-1)$  adições complexas. Então os  $N$  valores da  $DFT$  podem ser calculados por um total de  $N^2$  multiplicações complexas e  $N(N-1)$  adições complexas.

É interessante ver a  $DFT$  e a  $DFT^{-1}$  como uma transformação linear na seqüência  $x_s(n)$  e  $X_S(k)$ , respectivamente. Vamos definir o vetor  $\mathbf{x}_N$  de tamanho  $N$ , com mesmo comprimento da seqüência do sinal de  $x_s(n) = 0, 1, \dots, N-1$ , com mesmo comprimento e um vetor  $\mathbf{X}_N$  também de tamanho  $N$  de amostras em freqüência e uma matriz  $\mathbf{W}_N$  de comprimento  $N \times N$ , como:

$$\mathbf{x}_N = \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix} \quad \mathbf{X}_N = \begin{bmatrix} X(0) \\ X(1) \\ \vdots \\ X(N-1) \end{bmatrix} \quad \mathbf{W}_N = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & W_N & W_N^2 & \dots & W_N^{N-1} \\ 1 & W_N^2 & W_N^4 & \dots & W_N^{2(N-1)} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & W_N^{N-1} & W_N^{2(N-1)} & \dots & W_N^{(N-1)(N-1)} \end{bmatrix},$$

Com estas definições, a  $DFT$  de  $N$  pode ser expressa em forma de matriz como:

$$\mathbf{X}_N = \mathbf{W}_N \mathbf{x}_N \quad (2.8)$$

onde  $\mathbf{W}_N$  é a matriz de transformação linear. Observamos que  $\mathbf{W}_N$  é uma matriz simétrica<sup>1</sup>. Se nós assumirmos que a matriz inversa de  $\mathbf{W}_N$  existe, então a Eq.2.8 pode ser invertida realizando uma pré-multiplicação de ambos os termos da equação por  $\mathbf{W}_N^{-1}$ . Assim obtemos

<sup>1</sup>Chama-se matriz simétrica toda a matriz quadrada  $A$  de ordem  $n$ , tal que:  $A^T = A$

$$\mathbf{x}_N = \mathbf{W}_N^{-1} \mathbf{X}_N \quad (2.9)$$

Da mesma forma podemos expressar a  $DFT^{-1}$  na forma de matriz,

$$\mathbf{x}_N = \frac{1}{N} \mathbf{W}_N^* \mathbf{X}_N \quad (2.10)$$

onde  $\mathbf{W}_N^*$  representa o complexo conjugado da matriz  $\mathbf{W}_N$ . Comparando Eq.2.9 com Eq.2.10, podemos concluir que

$$\mathbf{W}_N^{-1} = \frac{1}{N} \mathbf{W}_N^* \quad (2.11)$$

que por sua vez, implica em

$$\mathbf{W}_N \mathbf{W}_N^* = N \mathbf{I}_N \quad (2.12)$$

onde  $\mathbf{I}_N$  é uma matriz identidade de tamanho  $N \times N$ . Conseqüentemente, a matriz  $\mathbf{W}_N$  na transformação é uma matriz ortogonal(unitária). Além disso, sua inversa existe e é dada como  $\mathbf{W}_N^*/N$ .

## 2.4 Amostragem de sinais no domínio da frequência e reconstrução no tempo discreto

A análise de um sinal no domínio da frequência é freqüentemente realizada digitalmente por ser mais conveniente. Uma função contínua  $X_S(f)$  não é apropriada para ser analisada por meio de um computador. Nesta seção partimos da representação de um sinal no seu domínio freqüencial e analisamos as conseqüências da utilização da Transformada de Fourier Discreta Inversa ( $DFT^{-1}$ ) para a representação deste sinal no domínio do tempo.

A fim de exemplificar a análise de sinais em frequência, definimos um sinal  $X_S(f)$  contínuo, não periódico e de faixa frequência limitada, dado pela seguinte equação:

$$X_S(f) = e^{-\frac{f^2}{2\sigma^2}} \quad (2.13)$$

A representação deste sinal no domínio do tempo é dada pela sua Transformada de Fourier Inversa ( $TF^{-1}$ )

$$x_s(t) = \int X_S(f) e^{-2\pi jft} df \quad (2.14)$$

$$x_s(t) = \sigma e^{-\frac{1}{2}t^2\sigma^2} \quad (2.15)$$

Conforme apresentado na Figura 2.1, um sinal contínuo e não periódico no domínio da frequência é um sinal contínuo e não periódico no domínio do tempo.

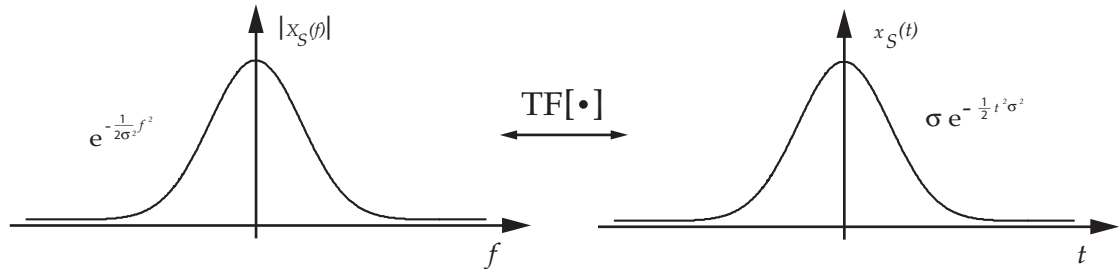


Figura 2.1: Exemplo do sinal  $x_s$  contínuo e não periódico representado nos domínios da frequência e do tempo.

### 2.4.1 Amostragem em frequência

Ao amostrarmos  $X_S(f)$  periodicamente em frequência com um espaçamento  $\Delta f$ , construímos o sinal  $X_{Sa}(f)$ . Este sinal é a multiplicação do espectro  $X_S(f)$  por uma função  $\delta_{\Delta f}(f)$  conhecida como Pente de Dirac, dada por

$$\delta_{\Delta f}(f) = \sum_{k=-\infty}^{\infty} \delta(f - k\Delta f) \quad (2.16)$$

Desta forma  $X_{Sa}(f)$ , apresentado na Figura 2.2, é dado por

$$\begin{aligned} X_{Sa}(f) &= X_S(f) \delta_{\Delta f}(f) \\ &= X_S(f) \sum_{k=-\infty}^{\infty} \delta(f - k\Delta f) \\ &= \sum_{k=-\infty}^{\infty} X_S(f) \delta(f - k\Delta f) \end{aligned} \quad (2.17)$$

Na Figura 2.2 podemos observar graficamente que o sinal  $X_{Sa}(f)$  é uma aproximação de  $X_S(f)$ , o que implica numa modificação na representação do sinal  $x_s$  no tempo. Para determinar esta modificação calculamos a  $TF^{-1}$  de  $X_{Sa}(f)$

$$x_{sp}(t) = TF^{-1}[X_{Sa}(f)] \quad (2.18)$$

$$x_{sp}(t) = TF^{-1}[X_S(f) \delta_{\Delta f}(f)] \quad (2.19)$$

Devido a propriedade de convolução do teorema da amostragem, a multiplicação de dois espectros no domínio da frequência é equivalente a convolução dos mesmos no domínio do tempo [JGP92]. Desta forma podemos expressar  $x_{sp}(t)$  como,



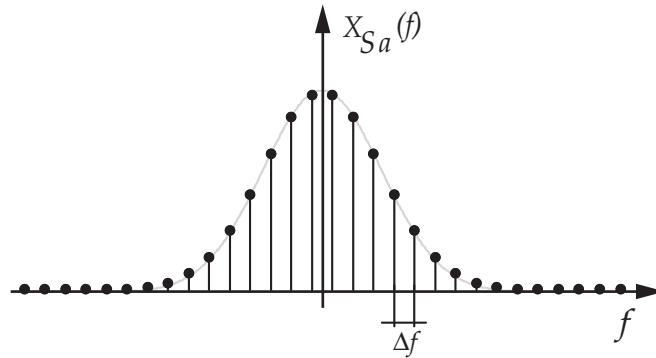


Figura 2.2: Sinal  $X_{Sa}(f)$  é uma aproximação de  $X_S(f)$  devido a discretização com passo  $\Delta f$ .

$$x_{sp}(t) = x_s(t) \otimes \frac{1}{\Delta f} \delta_{\frac{1}{\Delta f}}(t) \quad (2.20)$$

$$x_{sp}(t) = \frac{1}{\Delta f} \sum_n x_s(t) \delta(t - n/\Delta f) \quad (2.21)$$

ou

$$x_{sp}(t) = T \sum_n x_s(t) \delta(t - n/\Delta f) \quad (2.22)$$

onde  $T = \frac{1}{\Delta f}$

A Figura 2.3 mostra que a periodicidade do sinal  $x_{sp}(t)$  no domínio do tempo é uma consequência da amostragem deste sinal no domínio da frequência por  $\Delta f$ . É importante destacar que  $\Delta f$  deve ser suficientemente pequeno para não acarretar a sobreposição dos sinais no domínio do tempo (Teorema da Amostragem). Se este critério não for obedecido o sinal  $x_{sp}(t)$  será uma versão distorcida de  $x_s(t)$ , e conseqüentemente não terá toda a sua informação.

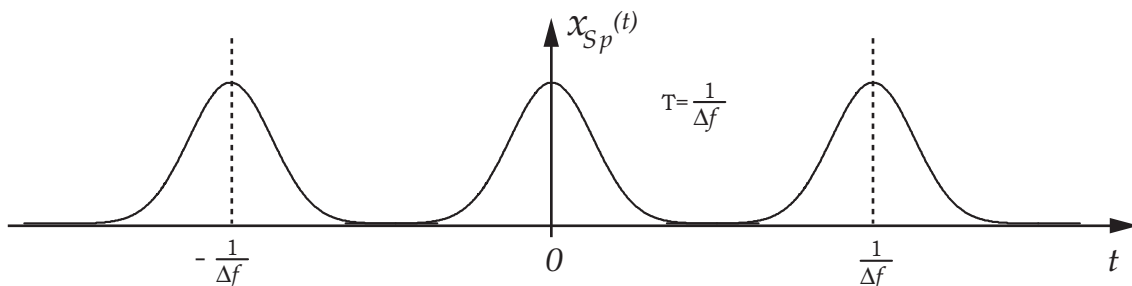


Figura 2.3: Representação do sinal  $X_{Sa}(f)$  no domínio do tempo -  $x_{sp}(t)$

### 2.4.2 Análise do sinal $x_s$ por Transformada de Fourier Discreta Inversa

A  $DFT^{-1}$  é calculada utilizando-se um número finito de amostras espectrais e, por ser realizada por um computador digital, esta transformação discretiza o sinal no domínio do tempo:

$$x_s(nT_M) = \sum_{k=-M/2}^{M/2} X_{Sa}(k\Delta f) e^{2\pi jnk/M} \quad n = 0, 1, 2, \dots, N \quad (2.23)$$

Onde o número de amostras espectrais é  $M\Delta f$  e define uma janela de observação do espectro e o passo de amostragem no tempo é  $T_M = 1/M\Delta f$ . Nesta análise a discretização do sinal no domínio do tempo transforma em periódico o sinal no domínio espectral.

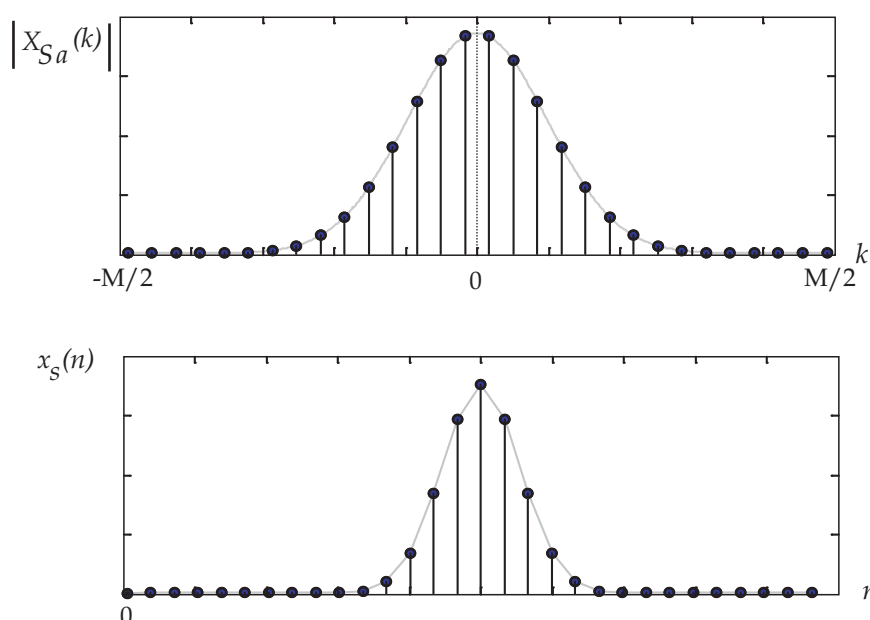


Figura 2.4: Exemplo da  $DFT^{-1}$  do sinal  $x_s$ .

Uma outra consequência da utilização da  $DFT^{-1}$ , normalmente ignorada na análise de dados, é a multiplicação do espectro por uma função Porta de largura  $M\Delta f$ .

$$X_{Sa}(f) = X_{Sa}(f) \prod_{M\Delta f(f)} \quad (2.24)$$

Esta multiplicação acarreta distorções no sinal  $x_s$  no domínio do tempo, pois  $x_s(t)$  se torna um sinal periódico, amostrado e convoluído com a transformada de Fourier de uma função Porta, tal como

$$x_{sap}(n) = (x_s(n) \otimes \frac{1}{\Delta f} \delta_{\frac{1}{\Delta f}}(n)) \otimes M \Delta f \operatorname{sinc}(\pi M \Delta f n) \quad (2.25)$$

A Figura 2.5 ilustra graficamente as conseqüências da utilização da  $DFT^{-1}$  na estimação de  $x_s$  no domínio do tempo. Dois casos são apresentados para a estimação de  $x_s(n)$ : i) a partir de um espectro com baixa resolução ( $M_1 = 30, \Delta f_1 = \sigma/3$ ) e ii) a partir de um espectro com alta resolução ( $M_2 = 100, \Delta f_2 = \sigma/10$ ). Para cada um dos dois casos, estimamos o sinal  $x_s(n)$  em três versões: baixa, média, e alta resolução temporal ( $N_1 = 30, N_2 = 100, N_3 = 1000$ ).

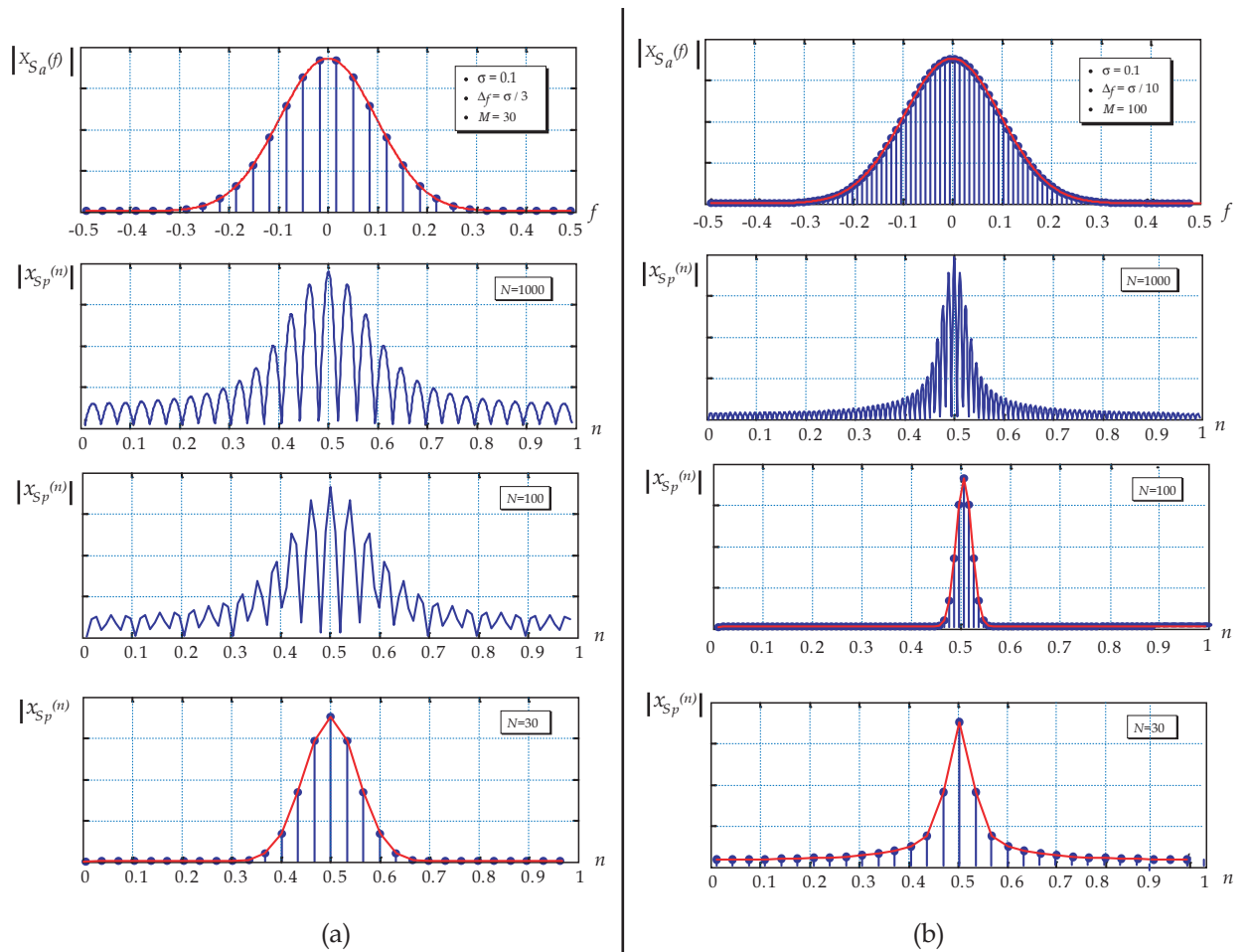


Figura 2.5: (a) Sinal  $x_s(n)$  com baixa resolução espectral ( $M_1 = 30, \Delta f_1 = \sigma/3$ ) estimado com diferentes valores para a resolução temporal ( $N_1 = 30, N_2 = 100, N_3 = 1000$ ) (b) Sinal  $x_s(n)$  com alta resolução espectral ( $M_2 = 100, \Delta f_2 = \sigma/10$ ) estimado com diferentes valores para a resolução temporal ( $N_1 = 30, N_2 = 100, N_3 = 1000$ )

Podemos observar nesta figura que a melhor estimação é feita quando  $M = N$ , pois nesta condição existe uma minimização do efeito de interferência provenientes do termo  $\operatorname{sinc}(\pi M \Delta f n)$  da Equação 2.25. Nos outros casos esta função distorce o sinal estimado.

Quando  $N > M$  o sinal estimado apresenta oscilações que se tornam tão mais importantes quanto maior for  $N$ . Estas oscilações prejudicam uma análise correta de  $x_s(n)$ . Para os casos onde  $N < M$  o sinal não apresenta oscilações, entretanto podemos observar que ele também é uma versão distorcida de  $x_s(n)$ . Estas distorções ocorrem devido ao termo  $\text{sinc}(\pi M \Delta f n)$  da Equação 2.25.

As oscilações observadas foram descritas pela primeira vez por Josiah W. Gibbs (1839-1903) e ficaram conhecidas como “Fenômeno de Gibbs”. Gibbs explicou as oscilações devido a limitações da soma da Série de Fourier quando analisamos sinais discretos e não periódicos em frequência.

É importante notar que a solução para tal problema não é somente o aumento da resolução espectral  $\Delta f$ . É importante aumentar também a janela de observação espectral  $M \Delta f$ . Se  $M \Delta f$  não for suficientemente grande não teremos uma boa definição (resolução) do sinal no domínio do tempo.

### 2.4.3 Limitação Experimental

Em algumas experiências em laboratórios de física, é comum nos depararmos com a limitação dos instrumentos. Dessa forma, os dados medidos em tais instrumentos são incompletos e ruidosos. Em casos onde as medidas são realizadas no domínio de frequência essas limitações têm que ser consideradas, pois a utilização da  $DFT^{-1}$  na análise dos dados pode gerar distorções significativas no sinal. Na maioria das vezes, sem nos darmos conta, o que se faz é atribuir zeros as componentes de frequência que não foram medidas. Esta é uma operação matemática de multiplicação do espectro por uma função Porta, tal como:

$$X_{Sobs}(f) = X_{Sa}(f) \prod_{L\Delta f}(f) \quad (2.26)$$

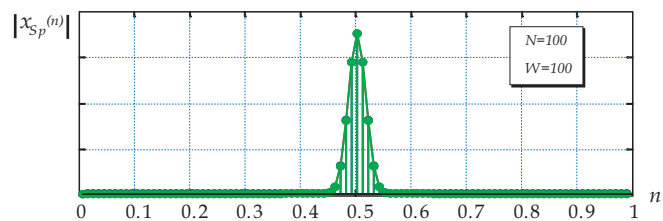
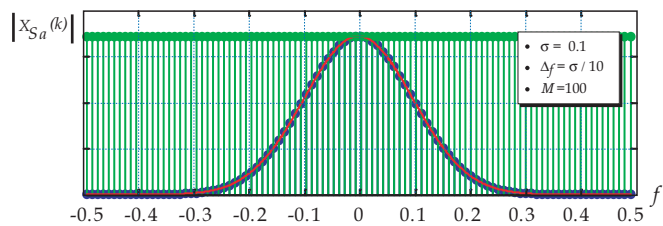
$$X_{Sobs}(n) = X_{sa}(n) \otimes L\Delta f \text{ sinc}(\pi L\Delta f n) \quad (2.27)$$

A Figura 2.6 apresenta a utilização da  $DFT^{-1}$  (caso  $N=M$ ) na análise espectral de um sinal, simulando uma limitação experimental. Neste exemplo não estamos considerando a imprecisão da medida.

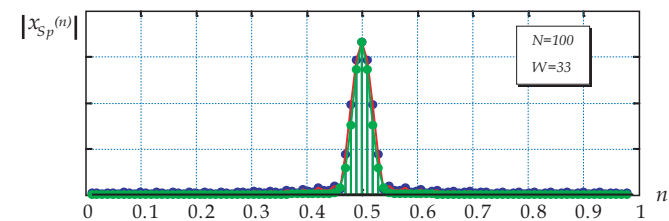
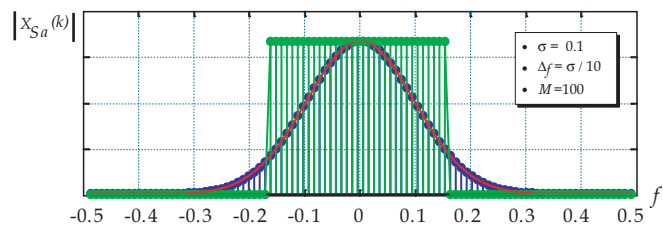
Observamos que à medida que diminuimos a largura da porta surgem oscilações no sinal no domínio tempo. Estas oscilações são conseqüências do termo  $\text{sinc}(\pi L\Delta f n)$ , e não devem ser confundidas com características e artefatos do sinal  $x_s$ . Em alguns casos estas oscilações são tão intensas que se torna impossível identificar o sinal original. Muitas formas de diminuir estas distorções foram propostas, tais como a multiplicação do espectro por uma função Janela  $W$  (como por exemplo: Janela de Hanning, Hamming, Kaiser ...) [JGP92]. As oscilações acontecem devido a descontinuidade do espectro propiciada pela multiplicação por uma função Porta. As funções Janelas tentam minimizar este efeito, mas introduzem outros pois a convolução com  $X_{sa}(n)$  será com a Transformada de Fourier de  $W$ .

$$X_{Sobs}(f) = X_{Sa}(f) W_{L\Delta f}(f) \quad (2.28)$$

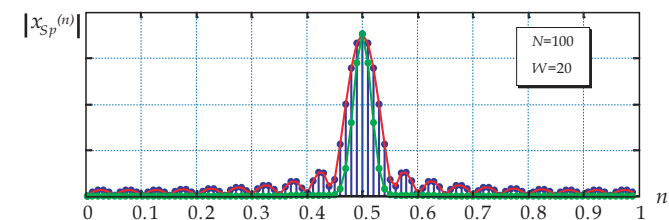
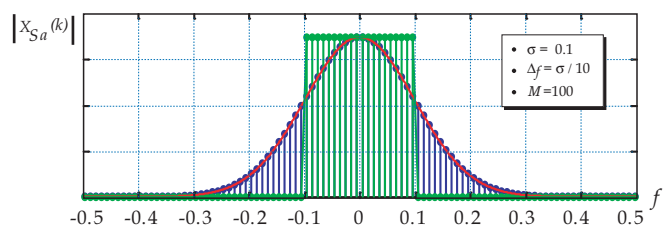
$$X_{sobs}(n) = X_{sa}(n) \otimes TF[w(n)] \quad (2.29)$$



(a)



(b)



(c)

Figura 2.6: Análise espectral de um sinal simulando uma limitação experimental por meio da multiplicação por uma porta.

## 2.5 Conclusão

Esse capítulo nos mostrou as implicações da utilização da  $DFT^{-1}$  na inversão de sinais do espaço de frequência para o espaço recíproco. Mostramos que a análise por  $DFT^{-1}$  introduz oscilações no sinal estimado podendo ser tão intensas tornando impossível a análise do sinal de interesse.

Além disso, discutimos o problema da limitação dos instrumentos de medida quando utilizamos a  $DFT^{-1}$ . É muito comum em experimentos, como, por exemplo, difratômetros e espectrômetros, que as medidas sejam incompletas e ruidosas. Para uma inferência precisa sobre o experimento é necessário fazer um tratamento neste sinal de modo a estimar os dados incompletos e eliminar os ruídos.

Além das oscilações que a  $DFT^{-1}$  causa no sinal, outro aspecto relevante na instrumentação científica está relacionado aos dados experimentais serem medidas estatísticas, i.e. possuem barras de erros. No cálculo da  $DFT^{-1}$  esta informação é desprezada.

Diante deste estudo, devemos nos questionar a respeito da qualidade da informação contida nos dados incompletos e ruidosos obtidos em instrumentos científicos. Dessa forma, a fim de calcular a Inversão de Fourier da melhor forma possível, é necessário buscar uma inferência dos dados em bases mais estatísticas. O próximo capítulo apresenta uma outra técnica de estimação do sinal pela  $DFT^{-1}$  baseada no princípio do Máximo de Entropia (MaxEnt).

---

## Capítulo 3

# Máximo de Entropia

Este capítulo aborda o estudo de problemas inversos, muito encontrado em análise de dados de instrumentos científicos. O objetivo é introduzir o método do máximo de entropia explicando porque sua utilização nos auxilia na solução de problemas encontrados quando analisamos dados. O capítulo apresenta como realizar uma boa estratégia de análise de dados quando os dados devem ser transformados de um espaço em outro, como é o caso da inversão de Fourier.

### 3.1 Problemas inversos

Problemas inversos constituem uma classe de problemas muito interessante e comum em ciências físicas e engenharia. Como exemplo, vamos considerar uma função multivariável  $f(\mathbf{x})$  pertencente ao conjunto de funções integráveis. Uma transformação de  $f(\mathbf{x})$  pode ser escrita da seguinte forma:

$$D(\mathbf{x}) = \int_D r(\mathbf{x}, \mathbf{x}') f(\mathbf{x}) d\mathbf{x} \quad (3.1)$$

onde  $r(\mathbf{x}, \mathbf{x}')$  é considerado o núcleo da integral representando a função de transformação ou a resposta do sistema.

Este tipo de formulação ocorre frequentemente em problemas de restauração ou reconstrução de sinais, como por exemplo a análise de Fourier. Nesses problemas, o objetivo é determinar  $f(\mathbf{x})$  a partir de um conjunto de medidas  $D(\mathbf{x})$  - estes são chamados problemas inversos. Problemas desse tipo acontecem frequentemente em física experimental quando procuramos medir uma grandeza  $f(\mathbf{x})$  que é inacessível de forma direta, a partir de uma outra grandeza experimental, acessível,  $D(\mathbf{x})$ .

A Figura 3.1 apresenta um diagrama da diferença entre a abordagem direta e inversa para um problema. Problemas diretos são aqueles para os quais conhecemos a formulação analítica e conseqüentemente a solução. Problemas inversos são abundantes na natureza, e sua solução é dada por meio de métodos algorítmicos. Estes problemas levam em conta os erros experimentais que dificultam a correta identificação do sistema. O termo problemas inversos é derivado da área de Projeto Ótimo, que pretende identificar uma melhor forma para se projetar um experimento. Para isto, realiza-se uma medida de parâmetros do sistema e procura-se maximizar uma informação escolhida e minimizar, por exemplo, o erro de medida.

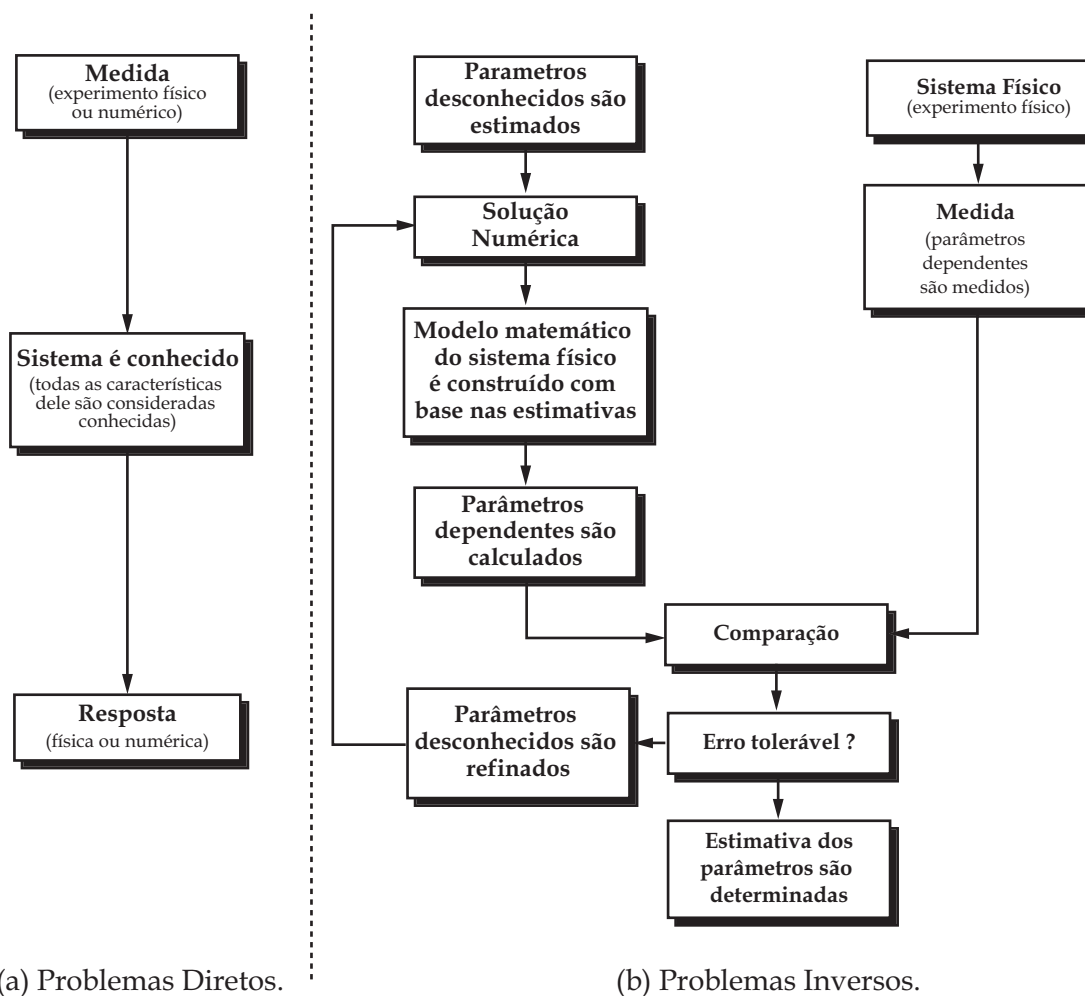


Figura 3.1: Diferença entre problemas diretos e problemas inversos. Em problemas diretos conhecemos a formulação analítica e a solução. Em problemas inversos a solução é dada por meio de métodos algorítmicos.

Uma questão principal a ser levada em conta em problemas inversos é se existe solução para a determinação da grandeza desconhecida  $f$ , a partir dos dados medidos. Para que um Problema Inverso tenha solução, este deve satisfazer a três condições, a saber: estabilidade, unicidade e existência.

Um Problema Inverso frequentemente não apresenta solução quando ele não satisfaz uma destas três condições. As dificuldades das duas primeiras condições (estabilidade e unicidade) são menos sensíveis que aquela da existência. Um Problema Inverso não irá existir desde que ele não tenha uma função  $D$  pertencente ao conjunto de observações possíveis. Uma solução única para  $f$  dependerá especificamente dos dados  $D$ . Esta interpretação quer dizer que um pequeno erro na função  $D$  poderá conduzir a erros e distorções em  $f$ . Cabe ressaltar que na prática os erros são inevitáveis, pois eles aparecem como ruídos provenientes das medidas experimentais.



Em problemas inversos, os estudos são norteados por um questionamento muito importante: “Que informação pode ser inferida com precisão a partir de dados incompletos e ruidosos?” [Lin95]. Diante deste questionamento se faz necessário encontrar uma forma de tornar os dados mais completos e menos ruidosos. Uma forma de estimar os dados incompletos é feita pelo máximo de entropia.

## 3.2 Análise de dados em medidas experimentais

A análise de dados, nos permite tirar conclusões a partir de medidas experimentais (Figura 3.2). A análise de dados pode ser vista como a determinação de  $N$  números a partir de  $K$  observações.

Um ponto elementar a respeito dos dados é que estes compreendem um conjunto finito de números como por exemplo àqueles obtidos em experiências computacionalmente controladas. Uma tabela de dados, sempre tem  $K$  números de pontos. A questão é se estas leituras são verdadeiramente independentes e se podemos converter tudo que vale a pena saber sobre uma experiência em um conjunto finito de dados  $N$ .

Neste texto “dados” significam exatamente o que se obtêm dos instrumentos de medidas. Não estamos nos referindo a números corrigidos ou calibrados. Normalmente em experimentos físicos começamos com leituras de instrumentos e fazemos uma série de correções ou calibrações nos dados, transformando-os passo a passo em nossas conclusões. Este procedimento intuitivo pode conter erros. Por exemplo, normalmente em um experimento os dados calibrados são considerados dados lidos do instrumento e isto interfere nas conclusões.

## 3.3 Conhecimento prévio e estratégia de análise de dados

A fim de exemplificar a estratégia de análise de dados, vamos considerar a determinação de médias estatísticas calculadas a partir de dados experimentais. A utilização de médias é, na prática, a maximização de uma distribuição de probabilidade. Neste conceito, encontramos a idéia de probabilidades condicionais bayesianas. Podemos apresentar o teorema de Bayes por:

$$p(\text{nova conclusão} / \text{medida}) \propto p(\text{medida} / \text{nova conclusão}) p(\text{estimação da medida, conhecimento prévio})$$

Este teorema nos indica como mudar nossas conclusões quando realizamos novas medidas. O termo  $p(\text{estimação da medida, conhecimento prévio})$  é muito interessante, pois ele representa tudo o que conhecemos da medida antes de realizá-la (a *priori*). Na verdade, para qualquer estimativa razoável, o conhecimento a *priori* não tem muita importância sobre o resultado final se as medidas experimentais são boas (Figura 3.3).

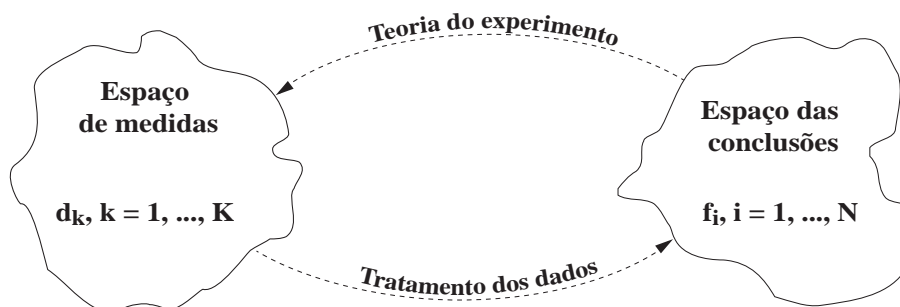


Figura 3.2: A teoria da experiência permite prever os dados medidos experimentalmente. O tratamento de dados é realizado no espaço de medidas nos levando ao espaço das conclusões.

Esta abordagem bayesiana está sempre presente na análise de dados científicos, entretanto, podemos nos perguntar por que não utilizamos sempre métodos bayesianos? Na verdade, existem casos onde não podemos calcular a probabilidade *a priori*, ou ainda situações onde os dados experimentais são tão “ruins” que interpretações *a priori* podem nos conduzir a erros. Entretanto, para a maior parte dos problemas, podemos definir um valor *a priori* que segue, por exemplo, a teoria do experimento.

Na estratégia de análise de dados, duas abordagens são comumente realizadas: ajuste dos dados experimentais e invenção de dados.

- Ajuste dos dados experimentais: em um ajuste, supomos que a curva de dados é lisa (sinal de baixa frequência) e tem uma forma conhecida. Por exemplo, podemos supor que um espectro contém duas linhas espectrais e que suas formas seguem um função Lorentziana (modelo). O procedimento de ajuste é arriscado e o que obtemos é uma resposta se essa forma esta de acordo com os dados experimentais, i.e. ou acreditamos no modelo, ou não podemos concluir sobre a experiência. Se não sabemos se o modelo é correto, tudo o que fizemos é estabelecer uma conclusão adaptada aos dados, entre milhares de outras. A construção de uma solução é fácil, a parte mais difícil é descobrir quais são os elementos desta solução que seguem inexoravelmente os dados e quais são resultados do modelo.
- Inventar dados adicionais: é uma abordagem comum e sempre escondida (passamos por ela sem nos darmos conta). Um exemplo disto é a introdução de zeros no cálculo da transformada de Fourier inversa. Esta abordagem, por soma de uma serie de Fourier incompleta, é tão ruim que algumas vezes é preciso modificar boas medidas ou mesmo reduzir a resolução para tornar o resultado mais confiável. O resultado final pode conter artefatos e são estes que estamos procurando observar. Ver (Capítulo 2 - Seção 2.4.2 e 2.4.3)

Em física experimental, muitos dados são medidos no espaço de frequência. A questão que se coloca é de como calcular, de uma maneira adequada, um sinal a partir de medidas experimentais incompletas e ruidosas de componentes de Fourier. A seguir discutimos

a utilização da inversão de Fourier, o desenvolvimento de modelos teóricos ajustados pelo método dos mínimos quadrados e o máximo de entropia.

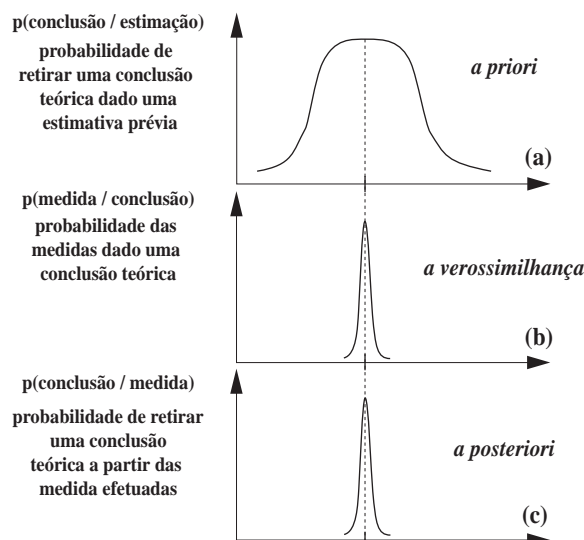


Figura 3.3: (a) Distribuição de probabilidade de uma medida antes da realização do experimento (conhecimento *a priori*). (b) Concordância das medidas realizadas com a teoria (verossimilhança). O Teorema de Bayes nos diz que se deve multiplicar (a) e (b) para obter a probabilidade *a posteriori* (c). Isto nos mostra como alterar a conclusão inicial a partir da realização da experiência. Pode-se observar, no gráfico (c), que a estimacão *a priori* não tem importância sobre o resultado final se as medidas são boas.

## 3.4 Análise de Dados Experimentais

### 3.4.1 Transformação de Fourier

Em um experimento onde são medidas componentes de Fourier, a primeira tentativa para observar o resultado no espaço recíproco é a utilização da Transformada de Fourier Discreta Inversa.

$$x_s(n) = \sum_k X_S(k) e^{2\pi jkn} \quad (3.2)$$

Na Equação 3.2 a soma deve ser realizada de  $-\infty$  a  $\infty$ . Na prática é impossível medir todas as componentes de Fourier e a soma deve ser realizada com o maior número possível de componentes para que possamos observar detalhes interessantes no sinal no espaço recíproco. A inversão de Fourier tem dois inconvenientes fundamentais: a imperfeição dos dados medidos e a omissão no cálculo das incertezas experimentais.

- Imperfeição dos dados: A inversão de Fourier, definido na Equação 3.2 é calculada corretamente se a soma sobre  $k$  inclui todas as componentes de frequência. No entanto, algumas componentes nem sempre podem ser medidas, às vezes, devido às intensidades serem fracas, ou devido ao instrumento de medida ser limitado e não poder medi-las. Entretanto, a quantidade de componentes com intensidades fracas ou inacessíveis, pode ser grande e não deveriam ser ignoradas.

Visto que o cálculo exige um número finito de componentes e, por esta razão, a soma de 3.2 é modificada para uma soma parcial e, por conseqüência, a solução do problema de inversão não é única. Um número infinito de sinais no espaço direto existe. Isto implica não só em valores com bias para o sinal calculado no espaço direto como também em erros de convergência da série.

Este tipo de abordagem implica que as componentes de Fourier, que não foram medidas, são colocadas em zero. A única justificativa para escolher o valor zero é o fato de não conhecermos estas medidas. A conseqüência de escolher zero é bem conhecida e foi apresentada na Seção 2.3 do Capítulo 2. A escolha de zeros, faz com que a função resultante no espaço direto possua oscilações provocadas pela descontinuidade no espaço recíproco. Estas oscilações são às vezes, tão importantes que devemos aplicar técnicas de filtragem a fim de reduzi-las. Para isso utilizamos funções de ponderações como Janelas, também apresentadas anteriormente, que modificam os dados medidos. Este tipo de filtragem pode melhorar a estimação do sinal, mas não utiliza a informação de incerteza do instrumento e não produz o sinal mais provável. Como decorrência, qualquer análise nos detalhes do sinal estimado deve ser feita com muito cuidado.

- Incerteza experimental: Dados  $\mathbf{D}^{obs}$  medidos em um instrumento tem uma determinada precisão, pois qualquer instrumento de medida esta sujeito a ruídos e para cada  $\mathbf{D}^{obs}$  temos um desvio padrão  $\sigma_{\mathbf{D}^{obs}}$  associado. Estas informações são completamente ignoradas no cálculo da Equação 3.2 e um dado impreciso é tratado exatamente como se este fosse preciso. Quando queremos estudar se o sinal reconstruído está, em termos estatístico, de acordo com a teoria do experimento  $\mathbf{D}^{th}$ , utilizamos a verossimilhança, ou somente o  $\chi^2$

$$\chi^2 = \frac{1}{K} \sum_k \frac{|\mathbf{D}^{obs}(k) - \mathbf{D}^{th}(k)|^2}{\sigma_{\mathbf{D}^{obs}}^2(k)} \quad (3.3)$$

$K$  é o número de observações independentes. O  $\chi^2$  deve ser previsto para não ter um valor muito maior que sua esperança formal, i.e. 1. Os dados calculados  $\mathbf{D}^{th}$  devem ter valores inferiores ou iguais aos  $\mathbf{D}^{obs}$  dentro da margem de erro determinada pelo desvio padrão. Por outro lado, se  $\mathbf{D}^{th} = \mathbf{D}^{obs}$ ,  $\chi^2 = 0$  o que é, evidentemente, sem nenhum significado. Isto reflete somente que a análise de um conjunto de dados ruidosos foi realizada como se estes estivessem perfeitos.

### 3.4.2 Modelos para o Experimento:

Outros tipos de tentativas existem para resolver o problema da inversão de Fourier de dados experimentais limitados e ruidosos. Tais soluções consistem em construir um modelo paramétrico teórico do sinal a ser reconstruído e ajustar este modelo por meio do método de mínimos quadrados.

Estes métodos baseados em modelos sofisticados são muito utilizados na reconstrução de sinais para vários tipos de dados e experimentos. Eles têm como vantagens permitir que tiremos conclusões bem confiáveis sobre o sinal, pois são baseadas em exigências físicas descritas no modelo teórico e estão em acordo com o conjunto de dados.

Por outro lado, trabalhos publicados em diversas áreas da física experimental para análise de dados, mostraram que uma modelização pode também introduzir artefatos que não são físicos na reconstrução do sinal. Estudos mostraram que mesmo as modelizações são insuficientes para descrever partes muito finas do sinal no espaço direto, pois em geral, a estimação por mínimos quadrados pode também conter um *bias* sistemático.[Alb99][BM91]

A solução que vem sendo dada para resolver tais problemas é uma abordagem bayesiana. Quando a quantidade de dados é insuficiente, isto não limita suficientemente nosso espaço de conclusão e existirá sempre um número de sinais para os quais a p(nova conclusão / medida) é grande. Devemos também decidir qual desses sinais, que estão de acordo com os dados medidos, será escolhido por meio da utilização do termo “a priori”. Na realidade, na maior parte dos casos, os sinais reconstruídos que estão em acordo com os dados medidos se distinguem uns dos outros por detalhes muitos finos. Isto não nos permite fazer uma escolha. Devemos então escolher a probabilidade anterior a fim de escolhermos um sinal liso. A escolha de um sinal suave é a condição, *sine qua non*, quando comparado com o método tradicional de atribuir zeros as componentes que não foram medidas. O máximo de entropia é um modelo que permite definir esta probabilidade anterior.

### 3.5 Introdução ao Máximo de Entropia

Em análise espectral é necessário levar os dados do espaço da frequência para o espaço recíproco, e isto significa inverter o operador de densidade espectral  $\mathbf{R}$  e este tipo de problema inverso, é formulado por,

$$\mathbf{D}^{obs} = \mathbf{R}\mathbf{f} + \vec{erro} \quad (3.4)$$

onde  $\mathbf{D}^{obs}$  é um vetor de comprimento  $K$  (número de dados),  $\mathbf{f}$  é a distribuição de interesse de comprimento  $N$ , e  $\mathbf{R}$  é uma matriz de transformação  $K \times N$ .

O fundamento do método do máximo de entropia, encontra-se na aplicação do teorema de Bayes. Quando nos referimos a estimação de dados a partir do MME, o questionamento anterior passa a ser feito da seguinte forma. Dado um conhecimento qualquer “ $\mathcal{I}$ ” prévio sobre o problema em estudo, e alguns dados observados experimentalmente  $\mathbf{D}^{obs}$ , qual a probabilidade de  $\mathbf{f}$  ser a distribuição correta?

Esta questão pode ser analisada matematicamente, usando o teorema de Bayes:

$$P(\mathbf{f} | \mathbf{D}^{obs}, \mathcal{I}) \propto P(\mathbf{D}^{obs} | \mathbf{f}, \mathcal{I}) \frac{P(\mathbf{f}, \mathcal{I})}{P(\mathbf{D}^{obs}, \mathcal{I})} \quad (3.5)$$

O método do máximo de entropia consiste em encontrar o  $\mathbf{f}$  que maximiza (Eq.3.5), ou seja, dada a informação “ $\mathcal{I}$ ” e os dados  $\mathbf{D}^{obs}$ , o objetivo é encontrar a  $\mathbf{f}$  mais provável.

O primeiro passo é encontrar uma expressão calculável para (Eq.3.5). Para isto vamos começar fazendo a seguinte observação, tendo em vista que o nosso interesse está relacionado com a determinação do  $\mathbf{f}$  mais provável, e que esta relaciona-se diretamente com  $\mathbf{D}^{obs}$  e “ $\mathcal{I}$ ”, o termo  $P(\mathbf{D}^{obs}, \mathcal{I})$  não possui nenhuma dependência com  $\mathbf{f}$ , então podemos considerar este

termo irrelevante e a assim eliminá-lo de nossos cálculos, sem influenciar no resultado. Desta forma temos

$$P(\mathbf{f} | \mathbf{D}^{obs}, \mathcal{I}) \propto P(\mathbf{D}^{obs} | \mathbf{f}, \mathcal{I})P(\mathbf{f}, \mathcal{I}) \quad (3.6)$$

é conhecida como *probabilidade posterior*

$P(\mathbf{D}^{obs} | \mathbf{f}, \mathcal{I})$  é a função verossimilhança (1), que nos fornece a probabilidade dos dados experimentais  $\mathbf{D}^{obs}$  para um determinado  $\mathbf{f}$ .

$$p(\mathbf{D}^{obs} | \mathbf{f}, \mathcal{I}) = \prod_i (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2}\chi^2(\mathbf{f})} \quad (3.7)$$

$P(\mathbf{f}, \mathcal{I})$  é a probabilidade *a priori* para uma solução particular de  $\mathbf{f}$ . O princípio fundamental do máximo de entropia avalia que esta informação *a priori* é considerada como a entropia *a priori* [Ski88].

### 3.5.1 Entropia de uma distribuição

Como já apresentado no Capítulo1, a entropia mede o grau de desordem de um sistema, ou seja, para um sistema com uma configuração uniforme, a entropia é máxima, ao contrário, para uma configuração mais ordenada, a entropia toma valores menores Fig.(3.4). A entropia na teoria da informação corresponde à *incerteza probabilística* associada a uma distribuição de probabilidade. Cada distribuição reflete um certo grau de incerteza e diferentes graus de incerteza estão associados a diferentes distribuições (embora diferentes distribuições possam refletir o mesmo grau de incerteza) [Mat02].

Uma característica importante da entropia na teoria da informação, ou incerteza probabilística, é que ela está diretamente associada ao grau de similaridade entre as probabilidades de uma distribuição. Sem pretender aprofundar nesta discussão, o fato é que a medida introduzida por Shannon para quantificar a entropia em teoria da informação também se presta a quantificar diversos conceitos de interesse em outras disciplinas, desta forma podemos ter distribuição de proporções, ao invés de distribuição de probabilidades. Sob esta perspectiva, elas servem para medir igualdade, espalhamento, similaridade, diversidade, complexidade de sistemas e outros conceitos que aparecem em diversas áreas do conhecimento, ainda que tais conceitos não tenham uma relação direta com alguma noção clássica de entropia[Mat02].

A entropia descrita por Shannon tem a seguinte formulação matemática:

$$S(p) = \sum_i p_i \ln \frac{1}{p_i} = - \sum_i p_i \ln p_i \quad (3.8)$$

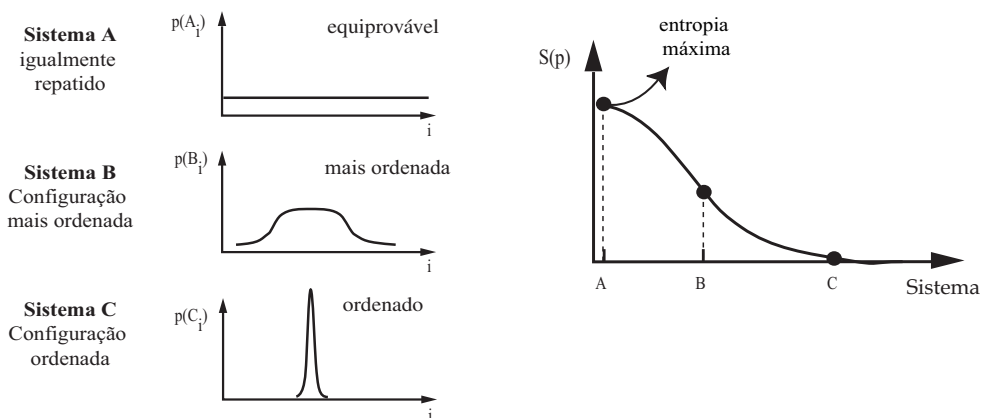


Figura 3.4: A entropia é a medida de desordem de um sistema fechado. A entropia é máxima numa configuração equiprovável deste sistema. Em configurações mais ordenadas, a entropia toma valores menores. Na realidade, existem diversas configurações mas as probabilidades não as mesmas e portanto, a mais provável é a que possui maior entropia.

Na Eq 3.8 consideramos que o espaço  $\mathbf{f}$  foi dividido em pequenas células de tamanho  $\Delta\mathbf{f}$ . A intensidade de  $\mathbf{f}$  na célula  $i$  é expressa por  $\mathbf{f}_i$ , supondo que a intensidade de  $\mathbf{f}$  seja quantificada tal que  $\mathbf{f}_i = n_i \cdot \Delta\mathbf{f}$ . A quantidade  $\Delta\mathbf{f}$  é muito pequena enquanto que os valores para  $n_i$  são todos grandes. Assim é possível representar  $\mathbf{f}$  por um conjunto de números inteiros  $n_1, n_2, \dots$ . Como apresentado pela Eq.3.8,  $p_i = n_i/N$  onde  $n_i$  é a contagem da célula  $i$  e  $N = \sum_i n_i$ . A utilização da entropia pode ser justificada de diversas maneiras. Para o nosso caso especificamente, vamos utilizar o argumento de Gull e Daniell [SS84a]

*“Se uma proporção de alguma entidade que tem uma dada propriedade for  $p$ , a estimativa mais provável da proporção em alguma subclasse que tem essa propriedade é o mesmo número  $p$  (na ausência de outra informação conectando a subclasse com a propriedade)”.*

As evidências deste enunciado foram propostas por Gull e Daniell [SS84a] na apresentação do famoso “Argumento dos macacos” e igualmente na apresentação do “Problema dos cangurus” realizada por Skilling e Gull [SS84b]

*“Se uma equipe de macacos jogar um número muito grande  $N$  de quantas aleatoriamente em  $M$  células equivalentes a priori de uma imagem, então a probabilidade de obter um arranjo particular  $(n_1, n_2, \dots, n_M)$  de uma ocupação qualquer deve ser proporcional ao fator de degenerescência  $N!/n_1!n_2! \dots n_M!$ ”.*

O “Problema dos cangurus” [SS84a][SS84b] ilustra bem a utilização da entropia neste contexto. “ Sabe-se que 50% dos cangurus que habitam uma ilha isolada têm olhos azuis (OA) e 40% destes são canhotos. A pergunta é, qual a proporção de cangurus que têm olhos azuis e que são canhotos na ausência de mais informações. Desta forma este caso não possui uma única solução, e todas as repostas entre 0% e 40% são possíveis, porém

estas não possuem a mesma probabilidade. Se limitarmos o número de cangurus em 10 denominados por  $ABC\dots IJ$ , a figura (3.5) representa 3 possíveis distribuições para os 10 cangurus, respeitando as informações conhecidas, 50% (OA) e 40% (Canhoto).

	<b>Canhoto</b>	<b>Destro</b>									
<b>OA</b>	<b>A</b>	<b>BCDE</b>	50%	<b>OA</b>	<b>B</b>	<b>AEHI</b>	50%	<b>OA</b>	<b>AB</b>	<b>EHI</b>	50%
<b>não OA</b>	<b>FGH</b>	<b>IJ</b>	50%	<b>não OA</b>	<b>CDF</b>	<b>GJ</b>	50%	<b>não OA</b>	<b>CD</b>	<b>FGJ</b>	50%
	40%	60%		40%	60%		40%	60%			
	<b>(a)</b>			<b>(b)</b>			<b>(c)</b>				

Figura 3.5: Três possíveis distribuições em 4 categorias de 10 cangurus. As distribuições (a) e (b) correspondem a uma configuração  $[1\ 4\ 3\ 2]$  e a (c) à configuração  $[2\ 3\ 2\ 3]$ .

Configuração	Distribuições possíveis				
<table border="1" style="display: inline-table; border-collapse: collapse;"> <tr><td style="padding: 5px;">0</td><td style="padding: 5px;">5</td></tr> <tr><td style="padding: 5px;">4</td><td style="padding: 5px;">1</td></tr> </table>	0	5	4	1	$\frac{10!}{5! \ 4! \ 1!} = 1260$
0	5				
4	1				
<table border="1" style="display: inline-table; border-collapse: collapse;"> <tr><td style="padding: 5px;">1</td><td style="padding: 5px;">4</td></tr> <tr><td style="padding: 5px;">3</td><td style="padding: 5px;">2</td></tr> </table>	1	4	3	2	$\frac{10!}{4! \ 3! \ 2! \ 1!} = 12600$
1	4				
3	2				
<table border="1" style="display: inline-table; border-collapse: collapse;"> <tr><td style="padding: 5px;">2</td><td style="padding: 5px;">3</td></tr> <tr><td style="padding: 5px;">2</td><td style="padding: 5px;">3</td></tr> </table>	2	3	2	3	$\frac{10!}{3! \ 3! \ 2! \ 2!} = 25200$
2	3				
2	3				
<table border="1" style="display: inline-table; border-collapse: collapse;"> <tr><td style="padding: 5px;">3</td><td style="padding: 5px;">2</td></tr> <tr><td style="padding: 5px;">1</td><td style="padding: 5px;">4</td></tr> </table>	3	2	1	4	$\frac{10!}{4! \ 3! \ 2! \ 1!} = 12600$
3	2				
1	4				
<table border="1" style="display: inline-table; border-collapse: collapse;"> <tr><td style="padding: 5px;">4</td><td style="padding: 5px;">1</td></tr> <tr><td style="padding: 5px;">0</td><td style="padding: 5px;">5</td></tr> </table>	4	1	0	5	$\frac{10!}{5! \ 4! \ 1!} = 1260$
4	1				
0	5				

Tabela 3.1: Número de possíveis distribuições de 10 cangurus. Aqui mostramos que sem informações suplementares sobre os cangurus, a configuração  $[2\ 3\ 2\ 3]$  é a mais provável.

As distribuições (a) e (b) correspondem a uma configuração  $\begin{smallmatrix} 1 & 4 \\ 3 & 2 \end{smallmatrix}$  enquanto (c) corresponde a  $\begin{smallmatrix} 2 & 3 \\ 2 & 3 \end{smallmatrix}$ . O número de diferentes distribuições que dão a mesma configuração é calculado por meio de análises combinatórias, que são apresentadas na Tabela 3.1. Na ausência de mais informações a configuração  $\begin{smallmatrix} 2 & 3 \\ 2 & 3 \end{smallmatrix}$  é a mais provável, pois corresponde às distribuições mais prováveis.



A probabilidade de obter uma determinada configuração, para uma distribuição em  $M$  células, é proporcional ao fator de degenerescência, e o número de configurações possíveis para  $\mathbf{f}$  é

$$\Omega = \frac{N!}{n_1! n_2! \dots n_M!} = \frac{N!}{\prod_j n_j!} \quad (3.9)$$

Após a aplicação da fórmula de Stirling para grandes números, obtém-se,

$$\Omega(p) = \frac{N^M}{n_1^{n_1} n_2^{n_2} \dots n_M^{n_M}} = \frac{1}{p_1^{N p_1} p_2^{N p_2} \dots p_M^{N p_M}} \quad (3.10)$$

onde  $p_i = n_i/N$

A configuração que maximiza  $\Omega$ , também maximiza o seu logaritmo, ou seja, a entropia da distribuição:

$$\ln \Omega(p) = -N \sum_{i=1}^M p_i \ln p_i = N \cdot S(p) \quad (3.11)$$

Se  $S$  é a entropia definida na Equação (3.8). Podemos escrever a Equação (3.11) sob a forma exponencial

$$\Omega(p) = \exp(N \cdot S(p)) \quad (3.12)$$

### Maximização da Entropia

Na teoria da informação, maximizar a entropia significa determinar a distribuição de probabilidade que represente o máximo de incerteza, dadas certas restrições (condições de não negatividade e soma igual a 1 para as probabilidades).

Ou seja, significa determinar uma distribuição que tenha o maior grau de similaridade de suas probabilidades com uma distribuição de referência ou modelo. Ou ainda que seja mais parecida com a distribuição uniforme (modelo uniforme:  $\mathbf{m}^t = [\mathbf{m}_1/N, \dots, \mathbf{m}_M/N]$ ) e diferindo dela apenas devido às restrições.

As distribuições modelo, por sua vez, refletem algum tipo de informação prévia sobre o fenômeno probabilístico de interesse.

O princípio de maximizar a entropia através da medida de Shannon, dado um conjunto de restrições, foi introduzido por Jaynes [Jay57] utilizando a entropia descrita por [Joh80];[YTL84]; [SS84a].

$$S(\mathbf{f}, \mathbf{m}) = \sum_{i=0}^{M-1} \mathbf{f}_i - \mathbf{m}_i - \mathbf{f}_i \ln \left( \frac{\mathbf{f}_i}{\mathbf{m}_i} \right) \quad (3.13)$$

onde  $\mathbf{f}$  é o valor do sinal calculado e  $\mathbf{m}$  é valor atribuído ao modelo inicial. Desta forma, esta análise pode ser generalizada para incluir um conhecimento anterior.

Enfim, se a função  $\mathbf{f}$  mais provável for obtida maximizando  $S(\mathbf{f}, \mathbf{m})$ ,  $p(\mathbf{f}, \mathcal{I})$  é uma função monotônica crescente de  $S$ . Desta forma, podemos escrever  $p(\mathbf{f}, \mathcal{I})$  sob a forma exponencial:  $p(\mathbf{f}, \mathcal{I}) \approx \exp(\alpha S(\mathbf{f}))$ , onde  $\alpha$  é proporcional a  $N$ . Em resumo, o modelo mais provável é aquele que possui a maior probabilidade  $p(\mathbf{f} | \mathbf{D}, \mathcal{I})$ , e o princípio do máximo de entropia consiste em maximizar:

$$p(\mathbf{f} | \mathbf{D}, \mathcal{I}) \propto p(\mathbf{D} | \mathbf{f}, \mathcal{I})p(\mathbf{f}, \mathcal{I}) \propto \exp\left(\alpha S(\mathbf{f}) - \frac{\chi^2}{2}(\mathbf{f})\right) \quad (3.14)$$

O que equivale a maximizar:

$$\phi(\mathbf{f}) = \alpha S(\mathbf{f}) - \frac{\chi^2}{2}(\mathbf{f}) \quad (3.15)$$

Nesta condição de maximização,  $p(\mathbf{f} | \mathbf{D}, \mathcal{I})$ , é conhecida como *máximo a posteriori*. Isso significa que a probabilidade de  $\mathbf{f}$  para um conjunto de dados experimentais, não somente é representada pelo acordo entre os dados observados ( $\mathbf{D}^{obs}$ ) e calculados ( $\mathbf{D}^{calc}$ ), mas também pela probabilidade intrínseca de  $\mathbf{f}$  [Mat02].

A inversão de Fourier negligencia completamente este último termo, enquanto que uma modelização paramétrica implica em outro tipo de conhecimento anterior: a modelização matemática do problema estudado.

Os princípios de otimização envolvem funções intrinsecamente não-lineares de probabilidades. Como constituem problemas de programação não linear, suas soluções demandam algoritmos de busca iterativa.

### 3.6 Modelo uniforme ou conhecimento anterior?

O modelo uniforme (ou modelo plano) é a “simplificação” do conhecimento anterior supondo que os pontos de  $\mathbf{f}$  têm a mesma probabilidade ( $\mathbf{f}^t = (\mathbf{f}_1/N, \dots, \mathbf{f}_M/N)$ ). Esta consideração é baseada no argumento de Gull e Daniel [Dan91] apresentado anteriormente. A utilização do modelo uniforme tem como vantagem evidenciar que as diferenças entre o modelo e a solução final estão realmente contidas nos dados.

Entretanto, em muitos casos, antes de realizarmos as medidas, já possuímos algum conhecimento sobre a distribuição estudada. Dessa forma, o conhecimento anterior plano utilizado no máximo de entropia tradicional (ou modelo uniforme) é uma condição inicial ruim, e deve ser evitada quando for possível.

O conhecimento anterior (ou modelo não-uniforme) pode ser o modelo paramétrico do fenômeno estudado. A vantagem de utilizarmos um modelo não uniforme é que qualquer informação ausente nos dados experimentais, mas necessários para a reconstrução, não serão inventados, mas retirado de um conhecimento teórico.

O cálculo de  $\mathbf{f}$  com um modelo não uniforme nos mostra exatamente quais pontos do espaço o conhecimento anterior é incapaz de explicar as observações experimentais. Pode-se então descobrir qual é a nova informação que a experiência traz. No caso da concorrência entre vários modelos teóricos, pode-se utilizar os dados experimentais para selecionar qual o melhor. Na verdade, o melhor modelo que foi utilizado como a *priori*, é aquele que contém o maior valor da entropia.

### 3.7 Discussão das soluções obtidas com o máximo de entropia

As principais características do máximo de entropia são

- Solução final positiva: o MME é utilizado para definir uma lei de probabilidade que não pode ser negativa;
- Solução de acordo com os dados experimentais: o MME atribui valores não nulos aos dados não medidos. Estes valores podem não ser corretos, mas são os mais prováveis levando em consideração as barras de erros das medidas.
- Precisão da estimação: o MME fornece uma reconstrução correta a partir dos dados ruidosos ou imperfeitos. No entanto, um problema importante é a avaliação da precisão desta reconstrução. Isto é imperativo para quantificações e análises muito precisas do sinal reconstruído. Para que o MME possa ser utilizado mais frequentemente, a sua confiança deve ser controlada por meio de avaliações de erros.
- *Bias* sistemático: os erros sistemáticos detectados nos resultados obtidos pelo MME são causados pelo algoritmo utilizado e não devido aos dados. Na verdade, a entropia realiza uma pressão sobre o resultado final em direção ao modelo (pressão em direção a uniformidade). Isto tem como efeito a introdução de um *bias* na reconstrução. No entanto, isto pode não ser uma restrição porque não conhecemos previamente a resposta correta relativa aos dados experimentais. Infelizmente, um *bias* existe igualmente em todos os métodos previamente descritos. Diversos, estudos vêm sendo realizados a fim de reduzir o *bias* na reconstrução pelo MME [Alb99][Moh87].

### 3.8 O parâmetro Alfa

De acordo com o Teorema de Bayes se maximizarmos  $p(\mathbf{f} | \mathbf{D}, \mathcal{I})$  em relação a  $\mathbf{f}$ , encontramos a distribuição  $\mathbf{f}$  mais provável para os dados confinados. O máximo ocorre quando

$$\nabla(\alpha S(\mathbf{f}) - \frac{\chi^2}{2}(\mathbf{f})) = 0 \quad (3.16)$$

ou seja,

$$\frac{\partial \phi(\mathbf{f})}{\partial \mathbf{f}} = 0 \quad (3.17)$$

Uma vez que já temos como calcular a probabilidade a *priori* pela entropia e a verossimilhança, resta determinar o melhor valor de  $\alpha$ . A constante  $\alpha$  é arbitrária e é escolhida frequentemente para fazer  $\chi^2$  igual ao número de  $K$  observações, desde que este valor seja previsto para uma distribuição  $\chi^2$ . Esta não é certamente uma regra definitiva e as possibilidades alternativas são discutidas por [Gul89].

Deve ficar claro que  $\alpha$ , pode ser considerado um parâmetro de regularização. Ele controla a comparação entre o ajuste do modelo  $\mathbf{m}$ , e os dados  $\mathbf{D}^{obs}$ . Se  $\alpha$  é um valor grande, a reconstrução dos dados estará de acordo com o termo de entropia por este ser dominante. Se  $\alpha$  é pequeno, a reconstrução se torna um ajuste simples dos dados pelo método dos mínimos quadrados, minimizando o termo de entropia. Enfim, o parâmetro  $\alpha$  deve ser controlado cuidadosamente.

### 3.9 Algoritmo do Máximo de Entropia

Com o objetivo de tornar o método do máximo de entropia prático foi desenvolvido um algoritmo para calcular os valores de  $\mathbf{f}$ . Desta forma o algoritmo consiste em calcular a probabilidade posterior dada pela Equação (3.14) que segue,

$$P(\mathbf{f} | \mathbf{D}^{obs}, \mathcal{I}) = \exp(\phi(\mathbf{f})) \quad (3.18)$$

O primeiro passo é determinar o valor de  $\alpha$ , que deve ser suficientemente grande, para garantir que a distribuição  $\mathbf{f}$  seja pouco ruidosa. Em seguida é preciso maximizar  $\mathbf{f}$  para encontrar a probabilidade posterior. A maximização ocorre quando

$$Q(\mathbf{f}) = \frac{\partial \phi(\mathbf{f})}{\partial \mathbf{f}} = 0 \quad (3.19)$$

Para se encontrar o valor de  $\mathbf{f}$  que maximiza  $\phi(\mathbf{f})$  utiliza-se o método de Newton-Raphson<sup>1</sup>

$$\mathbf{f}_{i+1} = \mathbf{f}_i - \frac{Q(\mathbf{f}_i)}{Q'_{\mathbf{f}_i}} \quad i = 0, 1, 2, \dots \quad (3.20)$$

fazendo uma expansão de Taylor em torno do ponto  $\mathbf{f}$

$$\mathbf{f} = \mathbf{f}_i + \delta \mathbf{f}_i \quad (3.21)$$

temos

$$Q(\mathbf{f}) = Q(\mathbf{f} + \delta \mathbf{f}_i) = Q(\mathbf{f}_i) + \delta \mathbf{f}_i Q'(\mathbf{f}_i) + \frac{1}{2} \delta \mathbf{f}_i^2 Q''(\mathbf{f}_i) + \dots \quad (3.22)$$

onde

$$Q(\mathbf{f}_i) = \frac{\partial \phi(\mathbf{f})}{\partial \mathbf{f}} = \left( \alpha \ln \left( \frac{\mathbf{m}}{\mathbf{f}} \right) + \mathbf{R}^t [\sigma^{-2}] (\mathbf{D} - \mathbf{Rf}) \right) \quad (3.23)$$

e

$$\frac{\partial Q(\mathbf{f}_i)}{\partial \mathbf{f}} = \frac{\partial^2 \phi(\mathbf{f})}{\partial \mathbf{f}^2} = (\beta [\mathbf{f}^{-1}] + \mathbf{R}^t [\sigma^{-2}] \mathbf{R}) \quad (3.24)$$

A expressão para calcular o valor de  $\delta \mathbf{f}_i$  é:

$$\delta \mathbf{f}_i = \frac{Q(\mathbf{f}_i)}{Q'(\mathbf{f}_i)} = (\beta [\mathbf{f}^{-1}] + \mathbf{R}^t [\sigma^{-2}] \mathbf{R})^{-1} \times \left( \alpha \ln \left( \frac{\mathbf{m}}{\mathbf{f}} \right) + \mathbf{R}^t [\sigma^{-2}] (\mathbf{D} - \mathbf{Rf}) \right) \quad (3.25)$$

com  $\beta \geq \alpha$ .

<sup>1</sup>O método de Newton-Raphson, é um processo iterativo para encontrar as raízes de uma função, sua eficiência consiste numa boa escolha (feita arbitrariamente) do valor inicial.

Como foi dito anteriormente métodos de otimização de funções envolvem quase sempre processos iterativos, e estes por sua vez dependem na maioria dos casos da determinação de  $\delta\mathbf{f}$  a ser dado em direção ao máximo da função. A determinação do valor de  $\delta\mathbf{f}$  é um fator importante no processo de otimização e, com o objetivo de assegurar um bom resultado é conveniente determinar um limite  $r_0$ , que define a região na qual a maximização é realizada com sucesso [SS84b].

$$(\delta_r)^2 = \delta_{\mathbf{f}}^t[\mathbf{f}^{-1}]\delta_{\mathbf{f}} \leq r_0^2 \quad (3.26)$$

O algoritmo inicia com  $\mathbf{f} = \mathbf{m}$  e  $\alpha = \infty$ , onde a cada iteração o valor de  $\alpha$  diminui para satisfazer o critério de  $\chi^2 - 1 < \xi$ , onde  $\xi$  é o limite de convergência do algoritmo. A cada iteração, o valor de  $\delta_{\mathbf{f}}$  é adicionado a  $\mathbf{f}$  maximizando  $\phi(\mathbf{f})$

### 3.9.1 Algoritmo do MEM

- Passo 1 - Medir  $K$  dados e colocar na variável  $\mathbf{D}_i \pm \sigma$ .
- Passo 2 - Definir o valor do modelo -  $\mathbf{m}$ .
- Passo 3 - Atribuir a distribuição inicial  $\mathbf{f}$  o valor de  $\mathbf{m}$  isto é,  $\mathbf{f} = \mathbf{m}$ .
- Passo 4 - Atribuir ao critério de convergência o valor  $\xi$ .
- Passo 5 - Inicializar  $\alpha$  suficientemente grande ( $K.1000$ ), onde  $K$  é o número de observações independentes
- Passo 6 - Faça  $\alpha Inc = 0.9$ .
- Passo 7 - Calcular  $\delta_{\mathbf{f}_i}$  de acordo com a Eq(3.25).
- Passo 8 - Faça  $\mathbf{f}_{i+1} = \mathbf{f}_i + \delta_{\mathbf{f}_i}$
- Passo 9 - Calcular  $\chi_{f_{i+1}}^2$
- Passo 10 - Se  $\chi_{f_i}^2 - 1 > \xi$ , faça  $\alpha = \alpha.\alpha Inc$  e volte ao passo 7.

## 3.10 Exemplo Simples : sistema linear com duas equações e duas incógnitas

Nesta seção apresentamos um exemplo da utilização do MME [Alb99][SS84b], para a solução de um sistema linear com duas equações e duas incógnitas  $\mathbf{f} = (\mathbf{f}_1, \mathbf{f}_2)$ , da seguinte forma:

$$\begin{cases} 0.56\mathbf{f}_1 + 0.83\mathbf{f}_2 = 5.32 \pm 0.48 \\ 0.83\mathbf{f}_1 - 0.56\mathbf{f}_2 = 4.24 \pm 2.00 \end{cases}$$

onde os dados medidos são,

$$\mathbf{D} = \begin{bmatrix} 5.32 \\ 4.24 \end{bmatrix} \pm \sigma = \begin{bmatrix} 0.48 \\ 2.00 \end{bmatrix}$$

a matriz de inversão é

$$\mathbf{R} = \begin{bmatrix} 0.56 & 0.83 \\ 0.83 & -0.56 \end{bmatrix},$$

e as matrizes  $[\mathbf{f}^{-1}]$  e  $\sigma^{-2}$  são definidas como:

$$[\mathbf{f}^{-1}] = \begin{bmatrix} \frac{1}{\hat{\mathbf{f}}_1} & 0 \\ 0 & \frac{1}{\hat{\mathbf{f}}_2} \end{bmatrix}, \quad \sigma^{-2} = \begin{bmatrix} \frac{1}{0.48^2} & 0 \\ 0 & \frac{1}{4} \end{bmatrix},$$

A entropia e o  $\chi^2$  são escritos da seguinte forma:

$$S(\mathbf{f}) = \mathbf{f}_1 + \mathbf{f}_2 - \mathbf{m}_1 - \mathbf{m}_2 - \mathbf{f}_1 \log \left( \frac{\mathbf{f}_1}{\mathbf{m}_1} \right) - \mathbf{f}_2 \log \left( \frac{\mathbf{f}_2}{\mathbf{m}_2} \right) \quad (3.27)$$

$$\chi^2(\mathbf{f}) = (\mathbf{D} - \mathbf{Rf})^t [\sigma^{-2}] (\mathbf{D} - \mathbf{Rf}) \quad (3.28)$$

$$\chi^2(\mathbf{f}) = \frac{(0.56\mathbf{f}_1 + 0.83\mathbf{f}_2 - 5.32)^2}{0.48^2} + \frac{(0.83\mathbf{f}_1 - 0.56\mathbf{f}_2 - 4.24)^2}{2.00^2} \quad (3.29)$$

O  $\chi^2(\mathbf{f})$ , descreve uma elipse no espaço  $(\mathbf{f}_1, \mathbf{f}_2)$ . Um programa foi elaborado para calcular a trajetória das soluções de  $\hat{\mathbf{f}}$ , com  $\alpha$  assumindo o valor inicial, 2000, e sendo incrementado até a condição  $\chi^2 - 1 < \xi$ . O cálculo realizado utiliza o modelo  $\mathbf{m} = (1, 1)$ . A figura 3.6 apresenta a trajetória de  $\hat{\mathbf{f}}$  para o modelo. Os gráficos de superfícies da entropia  $S(\mathbf{f}, \mathbf{m})$  e de  $\chi^2(\mathbf{f})$  também são apresentados.

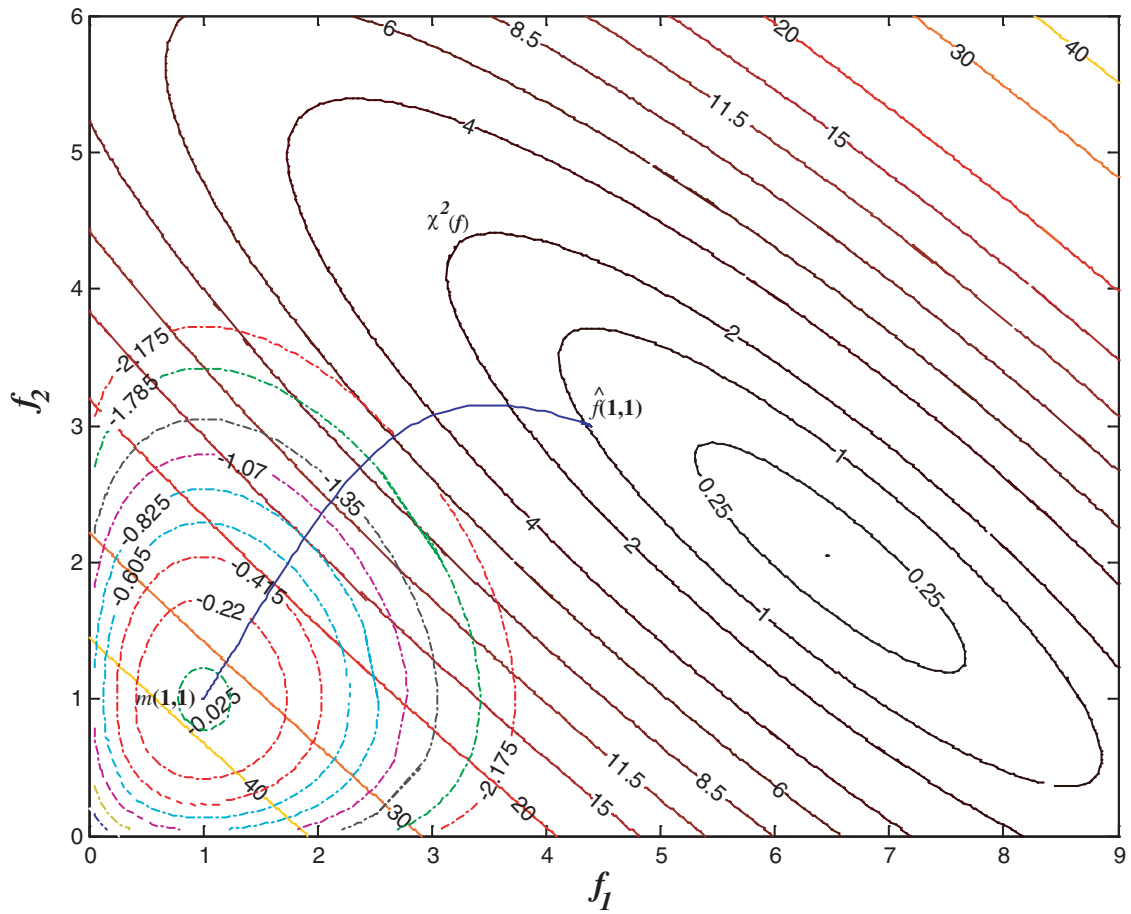


Figura 3.6: Exemplo da utilização do método do máximo de entropia por um sistema linear com duas incógnitas. Os cálculos são realizados a partir de um “modelo”:  $\mathbf{m} = (1, 1)$ , os gráficos de superfícies representa o  $\chi^2(\mathbf{f})$  e  $S(\mathbf{f}, \mathbf{m})$ . As curvas em traços fortes são as trajetórias das soluções aproximadas; a solução final a para a qual  $\chi^2 - 1 < \xi$ . Todos os valores contidos na elipse são soluções aceitáveis. A solução obtida é  $\hat{\mathbf{f}}_{1,1} = (4.34, 3.02)$ . A cada iteração, a nova solução aproximada é a que possui maior entropia

## Capítulo 4

# Aplicação do MME e Resultados

Este capítulo aborda a elaboração de um programa para a reconstrução de sinais por meio da inversão de Fourier utilizando o método do máximo de entropia (MME). Este programa utiliza o procedimento iterativo de Newton-Raphson a fim de ajustar um modelo que tenha o maior valor de entropia e que esteja em acordo com os dados experimentais.

O capítulo apresenta também uma análise do programa desenvolvido com dados simulados, discutindo os resultados encontrados, verificando sua eficiência e colocando em evidência as propriedades do MME.

### 4.1 Implementação Computacional do MME

A aplicação do MME na inversão de Fourier deve levar os dados do domínio freqüencial para o domínio do tempo e isto significa inverter o operador de densidade espectral  $\mathbf{R}$ . Este tipo de problema inverso, pode ser formulado por,

$$\mathbf{D}^{obs} = \mathbf{R}\mathbf{f} + \overrightarrow{erro} \quad (4.1)$$

onde  $\mathbf{D}^{obs}$  é um vetor de comprimento  $K$  (quantidade de dados medidos no instrumento),  $\mathbf{f}$  é a distribuição de interesse de comprimento  $N$ , e  $\mathbf{R}$  é uma matriz  $K \times N$ .

Para o cálculo da inversão de Fourier utilizando o método do máximo de entropia construímos uma sub-rotina em MATLAB (*MATLAB function*) denominada **idftme**. A figura 4.1 apresenta, em diagrama de blocos, quais os parâmetros de entrada e saída desta função.



Figura 4.1: Diagrama da função **idftme** construída no MATLAB, com os parâmetros de entrada [**DObs**, **SigDObs**, **modelo**, **R**, **LimConv**, **NIterMax**] e os respectivos parâmetros de saída [**S**,  $\chi^2$ , **f**, **FlagFim**].



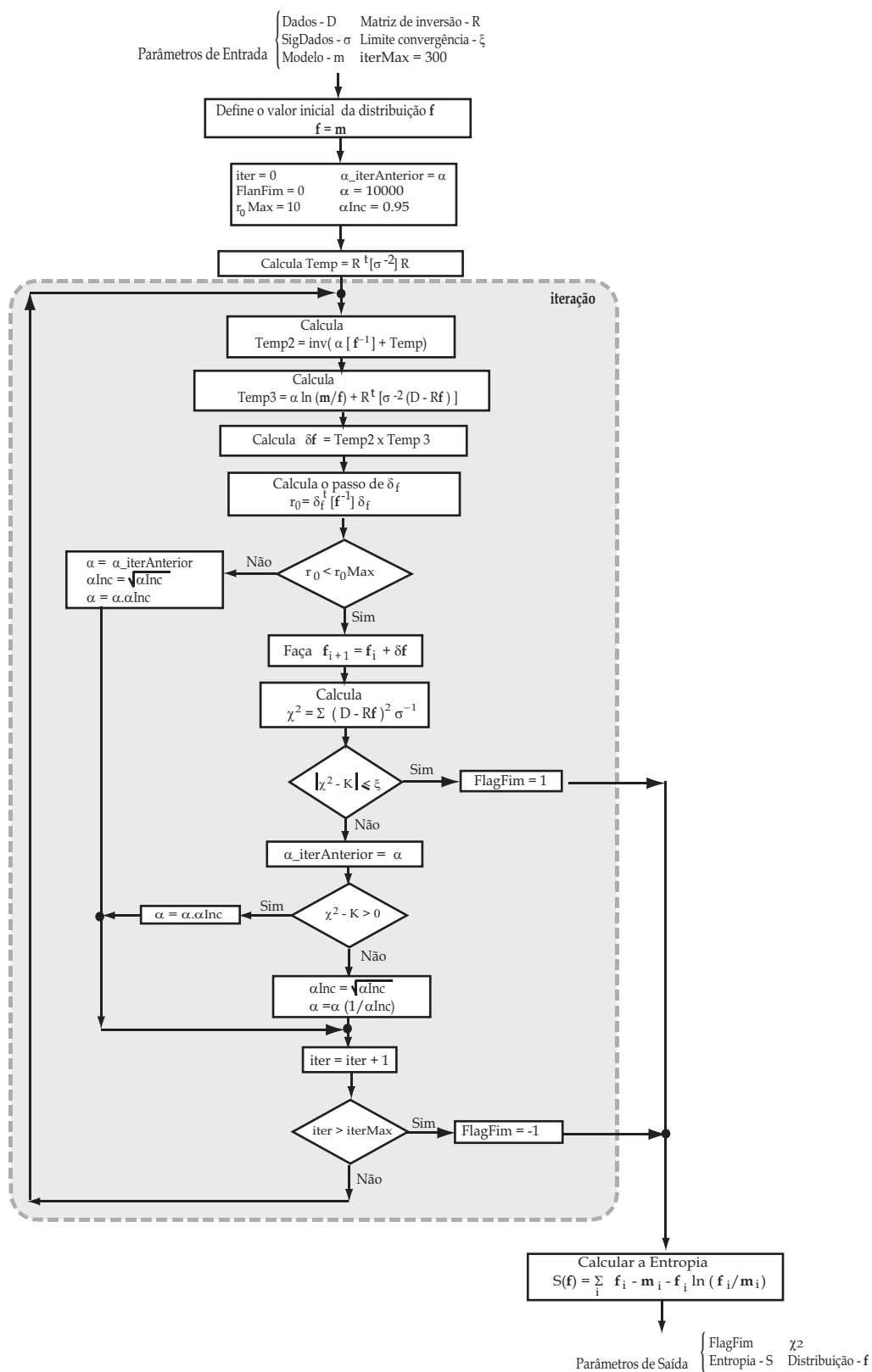


Figura 4.2: O fluxograma apresenta os passos computacionais para o cálculo da inversão de Fourier pelo método do máximo de entropia.

Para a utilização do **idftme** devemos colocar os valores das componentes de Fourier medidas no instrumento no vetor **DObs**, seus respectivos erros experimentais em **SigDObs**, o modelo que descreve o experimento em **modelo**, a matriz de inversão na variável **R**, o limite de convergência  $\xi$  na variável **LimConv** e o número máximo de iterações na variável **NIterMax**. A condição de parada do algoritmo é atingida quando se alcança a convergência de  $\chi^2 - K \leq \xi$  ou o número máximo de iterações. A variável *FlagFim* indica qual das duas condições foi alcançada. No caso de uma convergência de  $\chi^2$  (*FlagFim*=1) o programa devolve o resultado da inversão de Fourier mais provável no vetor **f**, o valor final calculado para  $\chi^2$  entre os dados observados experimentalmente e os dados estimados pelo método do máximo de entropia, e o valor final da entropia *S*.

A Figura 4.2 apresenta o fluxograma do algoritmo do MME utilizado no programa **idftme** foi dividido em três partes. A primeira parte define o valor inicial da distribuição procurada **f** e calcula as matrizes que são constantes durante a execução, i.e. independem de **f**. Em uma segunda parte, o programa calcula  $\delta\mathbf{f}$  que maximiza a entropia através da Equação 3.25, neste processo iterativo fazemos evoluir **f** até obtermos um valor de  $\chi^2$  próximo de 1, atingindo assim o limite de convergência  $\xi$  ou o número máximo de iterações. Em sua última parte o programa calcula o valor da entropia final pela Equação 3.8. Se o número máximo de iterações for atingido antes da convergência o programa pára e avisa ao usuário por meio de um *flag* que excedeu este valor (*FlagFim*=-1). Por fim a entropia é calculada (cabe lembrar que o cálculo da entropia ao final pode ser eliminado da sub-rotina quando o programa foi executado em *loop* a fim de diminuir o custo computacional). A sintaxe de utilização no MatLab é

```
[Entropia, c2, f] = idftme[Dobs, SigDobs, modelo, R, LimConv]
```

#### 4.1.1 Interface Gráfica

A fim de facilitar a utilização do **idftme** elaboramos uma interface gráfica (GUI - *Graphical User Interface*) para o Matlab, permitindo ao usuário uma melhor visualização dos resultados em um ambiente de trabalho mais agradável. Esta interface desenvolvida (veja Figura 4.3) permite ler os dados de um arquivo no formato ASCII dispostos sob a forma de duas colunas - dados experimentais, erros experimentais. O modelo a ser utilizado pode ser definido como sendo uniforme (valor constante) ou por meio de uma equação que descreve o resultado esperado. Em ambos os casos, é necessário definir o número de pontos utilizados para estimação de **f**.

Os resultados são apresentados na interface por meio de dois gráficos, sendo um no espaço de Fourier e o outro no espaço recíproco. Em ambos os gráficos os resultados são apresentados sobrepondo o modelo desejado ao sinal reconstruído calculado por **idftme** (i.e. que está em conformidade com os dados medidos). Os valores de  $\chi^2$  e da entropia (*S*) também são apresentados.

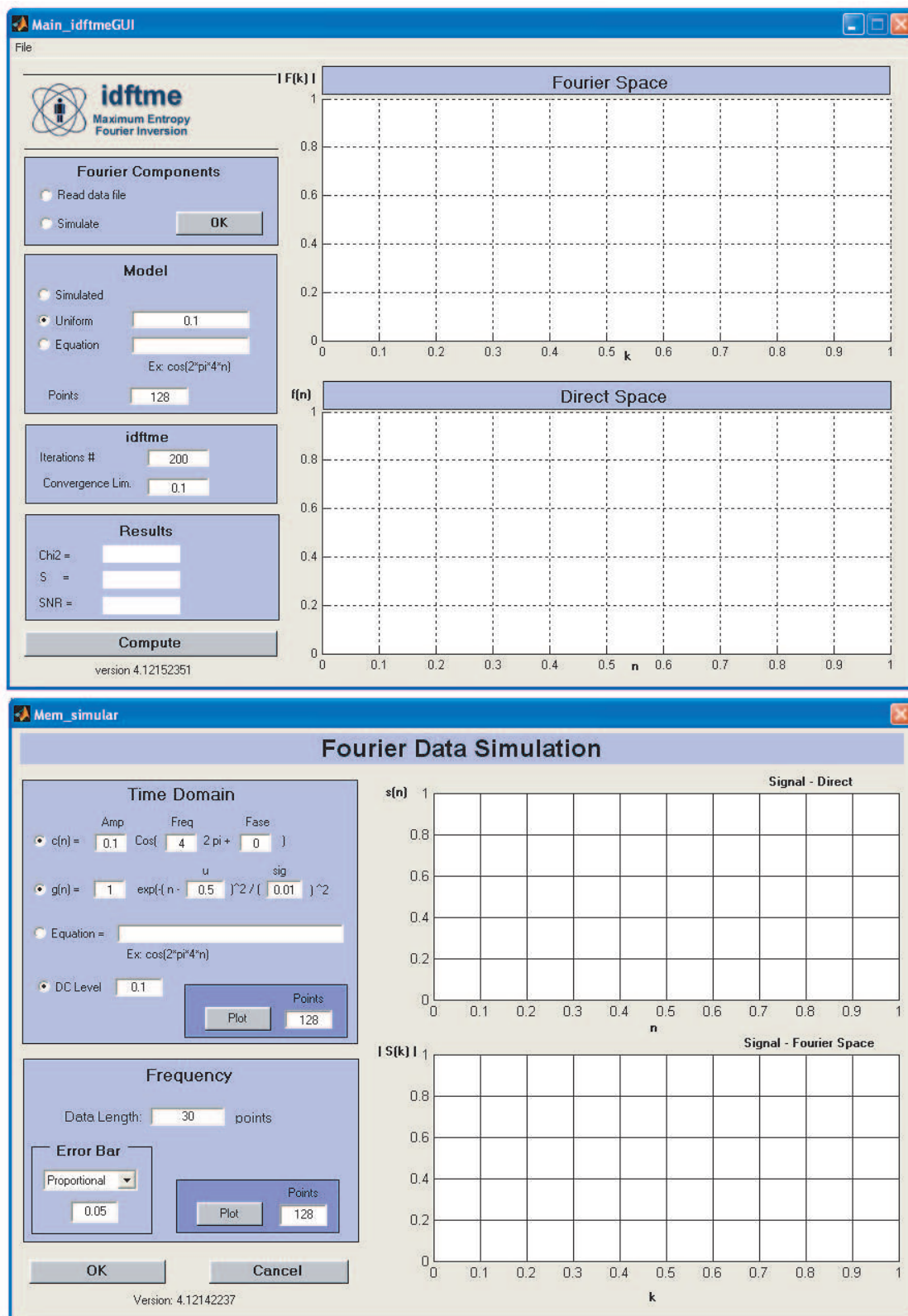


Figura 4.3: A interface gráfica foi elaborada para permitir ao usuário uma melhor visualização dos resultados alcançados pelo programa **idftme**.

## 4.2 Análise do método de Inversão de Fourier a Máximo de Entropia

A fim de analisar o programa **idftme** construímos uma simulação com objetivo de evidenciar as propriedades do MME, comprovando e validando sua funcionalidade.

A Figura 4.4 apresenta um diagrama de blocos de como foi realizada esta simulação. Inicialmente simulamos o sinal  $x_s(n)$  e calculamos a **dft** deste sinal a fim de obtermos as componentes de Fourier  $X_S(k)$ . Em seguida simulamos dados experimentais  $X_{S_{obs}}(k)$ , limitando as componentes de  $X_S(k)$  e adicionado um ruído gaussiano. Na terceira parte são aplicadas duas transformações no sinal  $X_{S_{obs}}(k)$ , *Inversão de Fourier clássica* e a *Inversão de Fourier a máximo de entropia*. Os resultados das transformações são comparados ao sinal original.

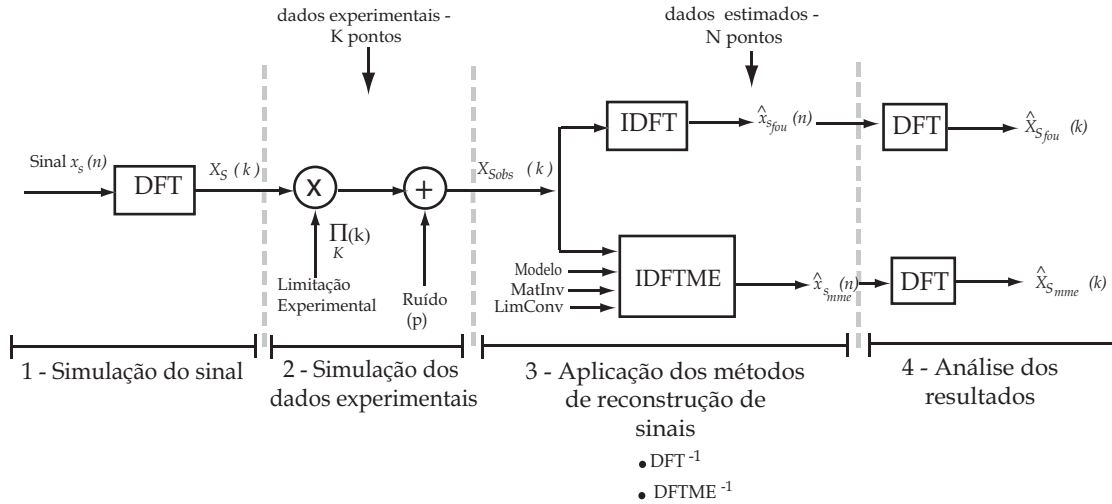


Figura 4.4: Diagrama de blocos da simulação para a avaliação do programa **idftme**.

A seguir descrevemos em detalhes as quatro etapas da simulação.

### 1. Simulação do sinal em frequência

Para teste do programa construímos um sinal no tempo  $x_s(n)$  composto de uma gaussiana e um cosseno

$$x_s(n) = \exp[-(n - \mu)^2/\sigma^2] + [0.1 \cos(2\pi f_0 n) + 1] \quad \forall \quad n \in [0, 1] \quad (4.2)$$

onde  $\mu$  é a posição da gaussiana,  $\sigma$  é a largura da gaussiana (desvio padrão),  $f_0$  é a frequência de oscilação do cosseno, e  $n$  o número de pontos. Para a simulação escolhemos  $\mu = 0,5$  e  $\sigma = 0.01$ .

Este sinal foi escolhido porque a gaussiana é um sinal de baixa frequência e sua representação espectral cai lentamente conforme aumentamos a frequência. O cosseno foi escolhido porque introduz oscilações no sinal que podem ser comparadas com as oscilações provenientes da reconstrução do máximo de entropia.

A Figura 4.5 apresenta o sinal  $x_s(n)$  e sua representação no espaço de Fourier, sinal  $X_S(k)$ <sup>1</sup>.

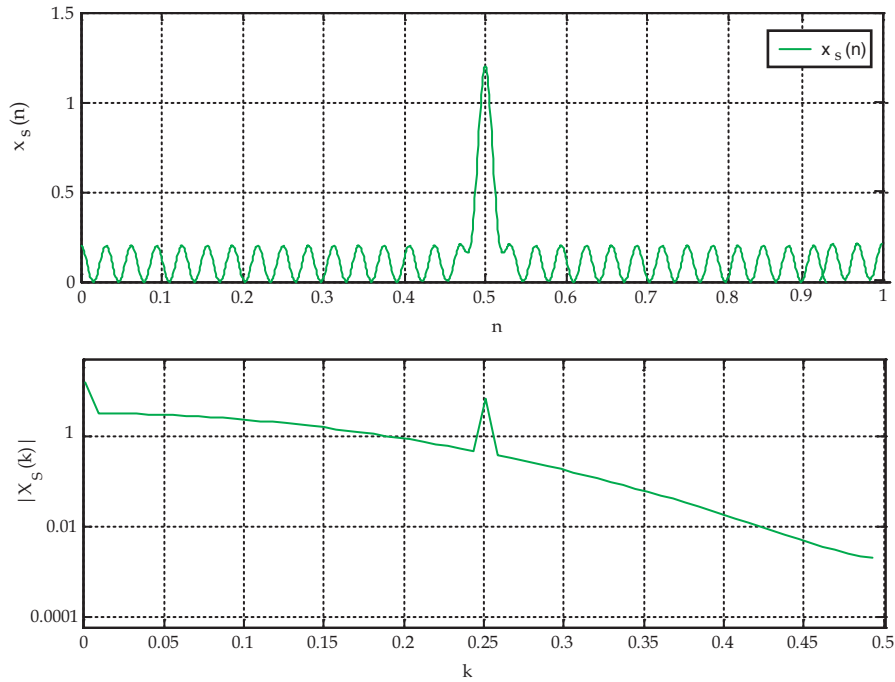


Figura 4.5: Representação do sinal  $x_s$  no tempo e em frequência normalizada.

## 2. Simulação dos dados experimentais

Esta parte da simulação limita em  $K$  componentes o sinal  $X_S(k)$ , adiciona um ruído gaussiano, e cria  $X_{S_{obs}}(k)$  que simula dados experimentais incompletos e ruidosos.

A limitação das componentes de frequência é realizada por meio da multiplicação do sinal  $X_S(k)$  por uma função porta,  $\prod_K(k)$ .

$$X'_{S_{obs}}(k) = X_S(k) \cdot \prod_K(k) \quad (4.3)$$

<sup>1</sup>Quando nos referimos a análise por **dft** o sinal é discreto e periódico, portanto sua representação é feita entre os valores de frequência normalizada  $[0, 0.5]$ .

Em seguida adicionamos ao sinal  $X'_{S_{obs}}(k)$  um ruído com desvio padrão  $\sigma_{X_{S_{obs}}(k)}$ . O ruído utilizado foi definido como uma distribuição gaussiana com média zero e desvio padrão proporcional a intensidade de  $X'_{S_{obs}}(k)$ . Nesta simulação o desvio padrão segue uma proporcionalidade constante tal que,

$$\sigma_{X_{S_{obs}}(k)} = X_{S_{obs}}(k) \cdot p \quad (4.4)$$

onde  $p$  é igual a 0.5% e 5%.

Ao final  $X_{S_{obs}}(k)$  é calculado por,

$$X_{S_{obs}}(k) = X'_{S_{obs}}(k) \pm N(0, \sigma_{X_{S_{obs}}(k)}) \quad \forall \quad k = 1, K \quad (4.5)$$

onde  $N(0, \sigma_{X_{S_{obs}}(k)})$  é um número aleatório com distribuição normal.

### 3. Aplicação dos métodos de reconstrução

O sinal  $X_{S_{obs}}(k)$ , é processado por meio de duas transformações, *Inversão de Fourier clássica- idft* e *Inversão de Fourier a Máximo de Entropia- idftme*.

A **idft** é calculada de acordo com,

$$x_{s_{fou}}(n) = \frac{1}{K} \sum_{i=0}^K X_{S_{obs}}(k) \exp(2\pi jnk/K) \quad (4.6)$$

A **idftme**, é calculada fornecendo alguns parâmetros ao algoritmo:

- **Modelo** - A escolha do modelo pode ser feita da seguinte forma:
  - Modelo Uniforme - atribui à variável *modelo* valores constantes isto é,

$$modelo(n) = 0,05 \quad \forall \quad n = [0, N - 1] \quad (4.7)$$

- Modelo Não-Uniforme - neste caso o modelo pode ser qualquer função. O modelo não uniforme utilizado pode ser, por exemplo,  $x_s(n)$  ou também somente a função gaussiana,  $g_s(n)$  descrita por:

$$g_s(n) = \exp[-(n - \mu)^2/\sigma^2] \quad (4.8)$$

- **Matriz de Inversão R**- A matriz de inversão utilizada é

$$R(k, n) = \exp(2\pi j(n - 1)(k - 1)/N) \quad \forall \quad n = 1, N; k = 1, K; \quad (4.9)$$

Nesta simulação utilizamos somente a parte real da matriz  $R$ , pois na maioria dos laboratórios experimentais os sinais medidos são números reais.

- Limite de convergência - é um valor que estabelece a condição de parada do algoritmo quando se alcança a condição  $\chi^2 - K \leq \xi$ . Nesta simulação foi utilizado  $\xi = 0.01$

#### 4. Análise dos resultados

Os resultados são analisados quantitativamente por meio da comparação entre os sinais:  $x_s(n)$  com  $\hat{x}_{sfou}(n)$  e  $\hat{x}_{smme}(n)$  e, qualitativamente entre os sinais  $X_S(k)$  com  $\hat{X}_{sfou}(k)$  e  $\hat{X}_{smme}(k)$  através da visualização em frequência da reconstrução das componentes retiradas.

A comparação é feita através da  $SNR$  (*signal-noise-ratio*) definida por:

$$SNR_{mem} = 10 \log \left[ \sum_{n=1}^N \left( \frac{x_s(n)^2}{(x_s(n) - \hat{x}_{smme}(n))^2} \right) \right] \quad (4.10)$$

para a transformação por **idftme**, e

$$SNR_{fou} = 10 \log \left[ \sum_{n=1}^N \left( \frac{x_s(n)^2}{(x_s(n) - \hat{x}_{sfou}(n))^2} \right) \right] \quad (4.11)$$

para a transformação por **idft**.

Outra análise que pode ser feita é uma observação da “distância” estatística (Capítulo 1 Seção 1.9 que existe entre o modelo inicial e resultado final i.e, o quanto nos afastamos do modelo para estarmos de acordo com os dados experimentais. Isto pode ser analisado pela entropia, se o modelo inicial for igual ao resultado final a entropia é zero. Entretanto, se a reconstrução se afasta do modelo, isto significa que a entropia é menor que zero, porém ela é máxima para o conjunto de dados.

##### 4.2.1 Simulações e Resultados

A partir do sinal  $X_{S_{obs}}(k)$ , quatro casos foram propostos a fim de reconstruir  $x_s$  por meio da **idftme**.

- **Caso 1:** Reconstrução das componentes retiradas e pressão em direção ao modelo

Este caso foi idealizado a fim de evidenciar as duas propriedades fundamentais do MME: i) reconstrução das informações que estão faltando; ii) e pressão para que o resultado final não se afaste do modelo. Para isto usamos os seguintes valores para os parâmetros: **modelo** = 0.05 (modelo uniforme), frequência  $f_0 = 6$ , e  $p = 5\%$  ou  $p = 0.5\%$ . A quantidade de dados  $X_{S_{obs}}(k)$  foi limitada em  $K = 25$  (componentes de Fourier) e a inversão de Fourier calculada com  $N = 128$ .  $S$  é o valor final da entropia após a convergência do algoritmo do MME.

Os dados simulados  $X_{S_{obs}}(k)$  ( $p = 0.5\%$ ) apresentados na Figura 4.8(b) foram utilizados para calcular por **idft** e **idftme** os  $\hat{x}_s$  da Figura 4.8(a). Nesta última Figura podemos observar que  $\hat{x}_{s_{fou}}$  contém oscilações devido ao uso da inversão de Fourier clássica. Por outro lado, o resultado da inversão utilizando o MEM,  $\hat{x}_{s_{mme}}$ , se aproxima melhor do sinal  $x_s$  original. A Figura 4.8(b) apresenta também a transformada de Fourier direta de  $\hat{x}_{s_{mme}}$ . Nesta Figura podemos observar que o MME tentou reconstruir as componentes de Fourier retiradas ( $k > 0.19$ ).  $\hat{x}_{s_{mme}}$  é dita como a solução final a máximo de entropia.

Na Figura 4.8(c)-(d), aumentamos a barra de erro dos dados fazendo  $p = 5\%$ . Isto deixa ao MME um maior grau de liberdade para que a solução final se aproxime do modelo. É importante notar que o valor final da entropia  $S$  é maior para a solução mais próxima do modelo ( $p = 5\%$ ). Nos dois casos, o valor final da entropia é o maior valor para o conjunto de dados fornecidos ( $X_{S_{obs}} \pm \sigma_{X_{S_{obs}}}$ ).

Observações relativas aos seguintes elementos:

**Co-seno:** A amplitude do co-seno diminui à medida que a barra de erro aumenta. Isto se deve a pressão exercida pela entropia em direção ao modelo inicial. Percebemos isto na representação no domínio da frequência, onde as componentes de  $\hat{X}_{S_{mme}}(k)$  que correspondem ao co-seno são menores que seu homólogo  $\hat{X}_S(k)$  e no domínio do tempo.

**Gaussiana:** as componentes de Fourier de alta frequência que foram reconstruídas, são exclusivas da gaussiana. Nos dois casos ( $p = 5\%$  e  $p = 0.5\%$ ), o resultado está muito próximo da realidade.

**Relação Sinal Ruído:** A estimação do  $\hat{x}_{s_{mme}}$  é melhor para uma barra de erro menor (Figura 4.8(a)-(b)). A informação contida no conjunto de dados é mais precisa e o algoritmo do MME pode exercer menos pressão em direção ao modelo uniforme. A reconstrução do MME eliminou praticamente todas as oscilações presentes em  $\hat{x}_{s_{fou}}$  devido a reconstrução das componentes de alta frequência. A Figura 4.6 apresenta graficamente que a relação sinal ruído para a reconstrução por MME é melhor.

**Entropia:** O valor da entropia é maior para o caso onde  $p = 5\%$ . Isto pode parecer paradoxal, pois a solução melhor (para  $p = 0,5\%$ ) tem uma entropia menor. Entretanto, devemos considerar que o modelo inicial (modelo uniforme) não é bom e a solução final está mais próxima do modelo quando a barra de erro é maior.



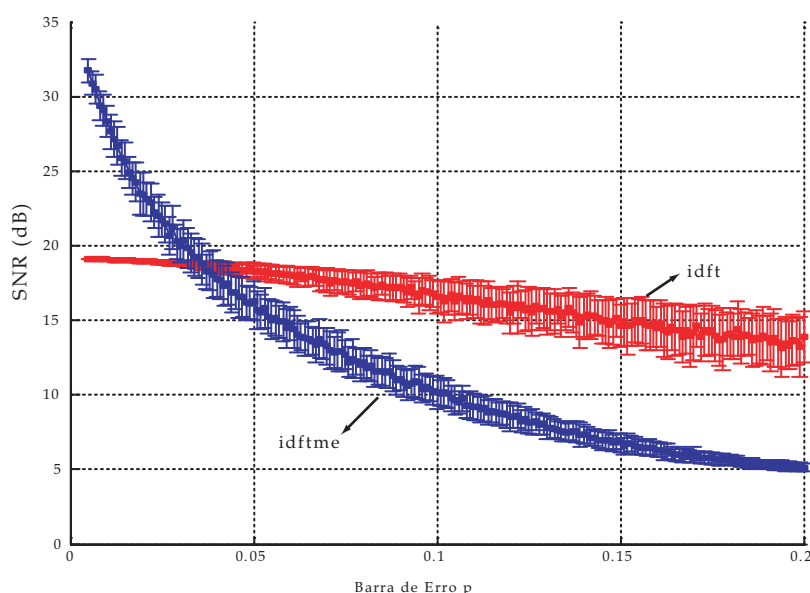


Figura 4.6: A relação sinal ruído em função da barra de erro  $p$  mostra que o resultado do MME é melhor .

- **Caso 2:** Reconstrução de acordo com o conjunto de dados

Este caso foi formulado para que o sinal  $X_{S_{obs}}$  possua apenas as componentes que correspondem a gaussiana, para isto os parâmetros que determinam  $X_{S_{obs}}$  foram escolhidos como: frequência  $f_0 = 34$ ,  $K = 25$ ,  $p$  igual a 5% e 0.5% e **modelo** = 0.05 (modelo uniforme). Podemos observar que em relação ao caso anterior, a frequência é o único parâmetro que foi modificado. O valor  $f_0 = 34$  faz com que o co-seno fique fora do conjunto de dados,  $X_{S_{obs}}$ .

Na Figura 4.9(a)-(c) observamos que  $\hat{x}_{s_{fou}}$  apresenta oscilações devido a limitação do número de componentes, enquanto que  $\hat{x}_{s_{mme}}$ , se aproximou somente da gaussiana. A Figura 4.9(b)-(d) mostra a reconstrução das componentes retiradas, ressaltando que o MME reconstruiu completamente as componentes de Fourier da gaussiana independente do valor de  $p$ . É importante notar que a reconstrução por MME não estima as oscilações do co-seno, isto se justifica pelo fato de não termos nem no conjunto de dados, nem no modelo nenhuma informação à respeito da existência de um co-seno.

Em ambas as situações de  $p = 5\%$  e  $p = 0.5\%$ , o valor final da entropia é específico do conjunto de dados utilizado e diferente de zero. Isto significa que as soluções encontradas pelo MME estão de acordo com os dados fornecidos e que os resultados se afastaram do modelo inicial.

- **Caso 3:** Modelo contido na reconstrução das componentes

Este caso trata da reconstrução do sinal  $X_{S_{obs}}$  feita a partir de um modelo não uniforme, ( $modelo = x_s$ ), os demais parâmetros têm a mesma configuração do caso 2.

A partir da Figura 4.10 podemos observar que a reconstrução do sinal utilizando o MME,  $\hat{x}_{smme}$  foi perfeita, e este foi capaz de reconstruir as componentes da gaussiana e do co-seno independente da amplitude do ruído (Veja a Figura 4.7). O valor da entropia nas duas situações é zero, sendo assim podemos afirmar que o modelo inicial está de acordo com o conjunto de dados fornecidos. Vale ressaltar que embora o co-seno não fizesse parte do conjunto de dados ele foi reconstruído, isto se deve ao fato do modelo utilizado possuir a informação referente ao co-seno, assim podemos dizer que o modelo está contido na reconstrução das componentes retiradas.

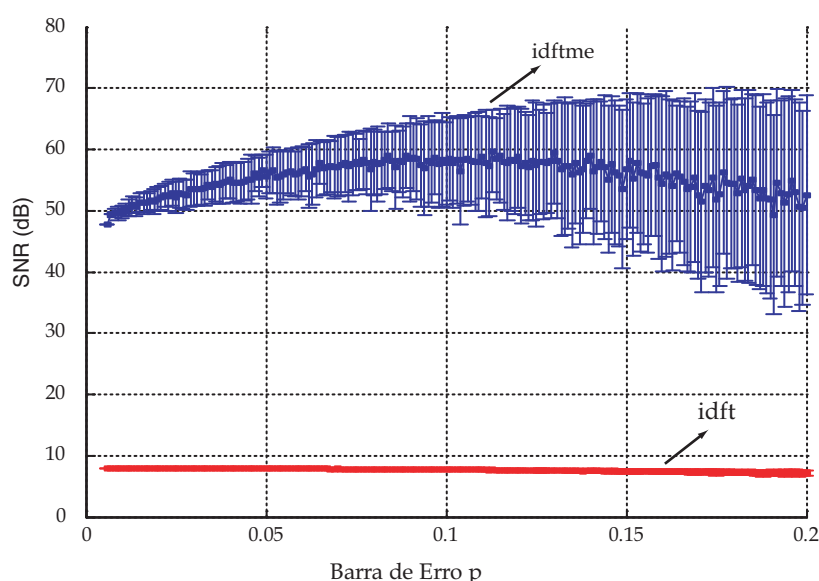


Figura 4.7: Relação sinal ruído em função da barra de erro  $p$ . A solução MME fornece sempre o melhor resultado.

- **Caso 4:** Validade do modelo

Este caso evidencia a importância dos problemas relacionados a escolha de um modelo. Para isso, utilizamos um modelo não uniforme que contém somente a gaussiana ( $modelo = g_s$ ). O número de componentes em  $X_{S_{obs}}$  assume dois valores,  $K = 25$  ou  $K = 36$  e o erro experimental é  $p = 5\%$ .

Na Figura 4.11 (a)-(b), o co-seno não faz parte do conjunto de dados  $K = 25$ . A Figura 4.11 (a) mostra que a inversão utilizando o máximo de entropia recuperou exatamente a gaussiana,  $\hat{x}_{smme}$ , e a Figura 4.11 (b) apresenta a reconstrução das componentes retiradas sendo que estas se referem apenas a gaussiana  $\hat{X}_{S_{smme}}$ . O valor final da entropia alcançou seu valor máximo,  $S = 0$  ou seja, a solução final é idêntica ao modelo.

A Figura 4.11 (c)-(d) apresenta a situação em que o co-seno faz parte do conjunto  $K = 36$  e o modelo inclui somente a gaussiana ( $modelo = g_s$ ). A Figura 4.11 (c) mostra que o sinal,  $\hat{x}_{s_{mme}}$  se aproximou melhor do sinal original  $x_s(n)$ . Na Figura 4.11 (d) podemos observar que ao tentar reconstruir as componentes retiradas o máximo de entropia superestimou estas componentes. A causa disto é a escolha inadequada do modelo, muito embora a entropia final comprove que a reconstrução não se afastou muito do modelo.

### 4.3 Conclusão

Este capítulo apresentou detalhadamente uma implementação computacional da inversão de Fourier a máximo de entropia.

O capítulo também apresentou por meio de simulações uma comparação da utilização da **idftme** e da **idft** na reconstrução de sinais. Mostramos que o processamento por meio da **idft** pode apresentar oscilações, enquanto que o resultado obtido pela **idftme** pode ser melhor em algumas circunstâncias.

Uma análise do programa desenvolvido foi proposta com quatro casos de estudos evidenciando as virtudes e fraquezas do MME. O caso 1 evidenciou as propriedades fundamentais do MME em reconstruir informações que estão faltando no conjunto de dados e a pressão para que o resultado final seja o mais próximo do modelo teórico. O caso 2 apresenta uma situação em que o modelo não é o correto, porém está de acordo com o conjunto de dados medidos no instrumento. O caso 3 mostra a utilização de um modelo correto, porém o resultado final não está totalmente de acordo com os dados medidos. Este caso evidencia que parte do sinal pode ser reconstruído baseado unicamente nas informações do modelo e que o modelo fornecido ao cálculo deve ser escolhido com muito cuidado. O caso 4 discute o problema da validade de um modelo, apresentando situações onde os dados experimentais forçam o resultado final mesmo que tenhamos utilizado um modelo ruim.

Os casos estudados demonstram também o papel do erro experimental na estimação do sinal. Medidas com barras de erro grandes permitem que o resultado final se aproxime do modelo teórico.

Enfim, os quatro casos evidenciaram que o valor final da entropia nos permite saber quanto sinal reconstruído se afasta do modelo, sendo a entropia uma informação útil para validá-lo.

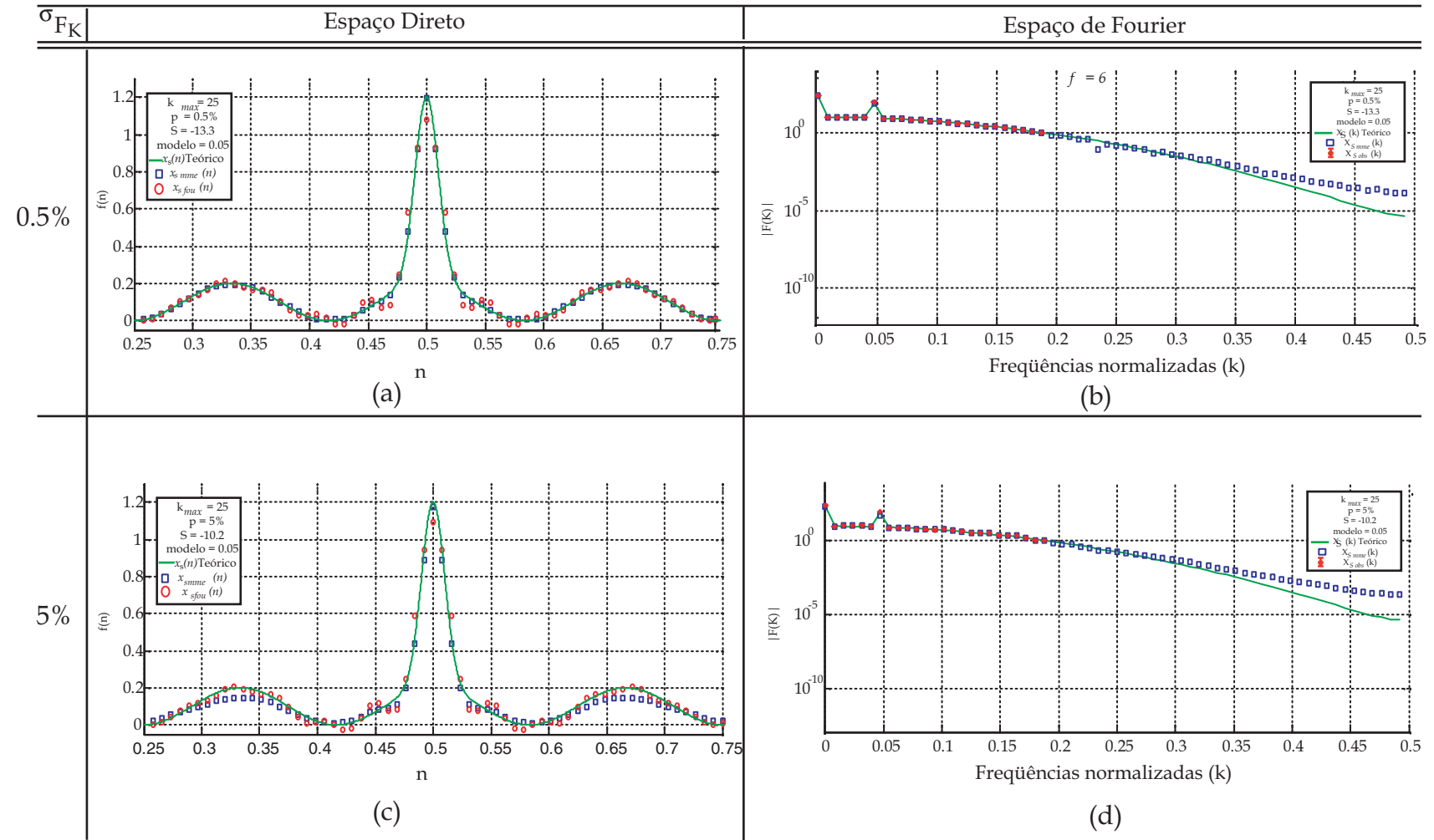


Figura 4.8: Inversão de Fourier com a utilização do máximo de entropia. As curvas (a) e (b) possuem barra de erro de 0.5%; as curvas (c) e (d) com barras de erro 5%. O cálculo é realizado conservando as 25 primeiras componentes de Fourier. O MME atribui valores às componentes que não foram fornecidas (altas frequências  $k > 0.19$ ). O erro da medida é um grau de liberdade utilizado por MME para aproximar a solução final do modelo inicial uniforme. (Frequência normalizada =  $k/128$ ).

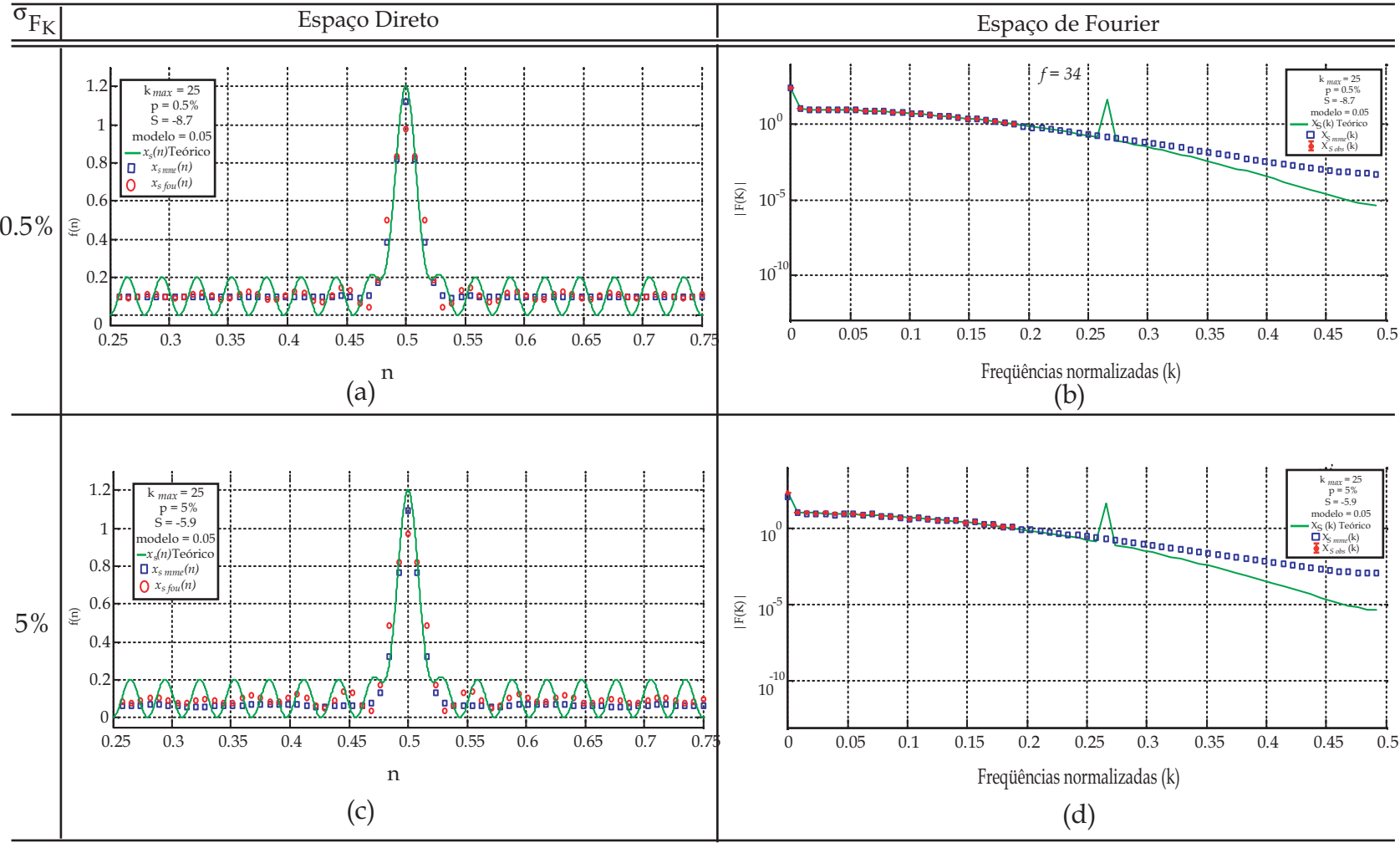


Figura 4.9: A figura apresenta as estimações de  $\hat{x}_s(n)$  se a componente de Fourier que descreve o cosseno encontra-se fora do jogo de dados. A solução do MME calculada com um modelo uniforme está de acordo com o jogo de dados. O MME estima apenas as componentes da gaussina.

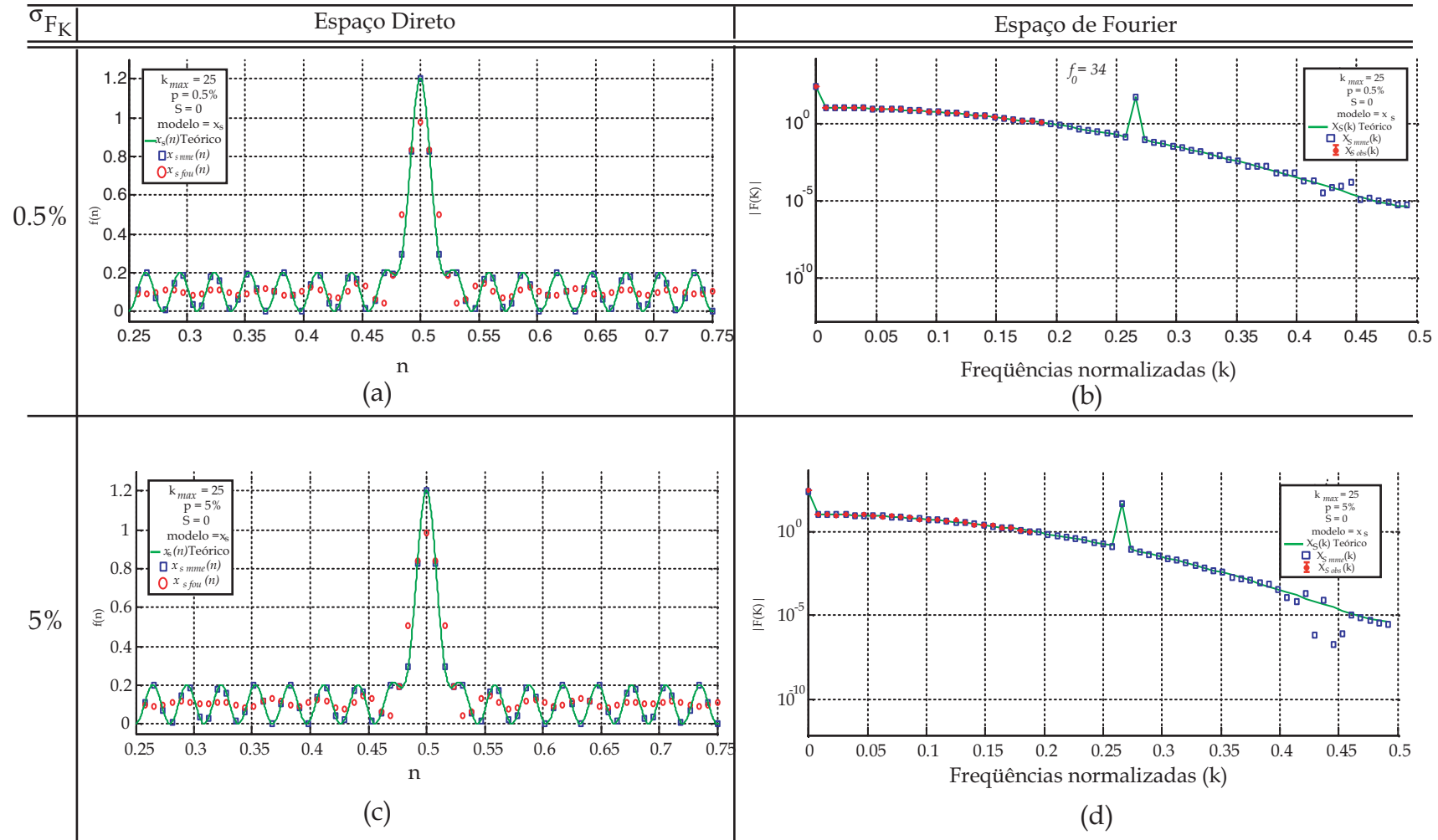


Figura 4.10: Reconstrução da curva utilizando a função  $x_s$  como modelo não uniforme. Qualquer que seja a amplitude da barra de erro, a solução do MME é perfeita e a entropia final é máxima. Evitamos fornecer as componentes que correspondem ao cosseno. Se o modelo contém informação que está em acordo com o conjunto de dados, o MME não tem razões para retirá-la.

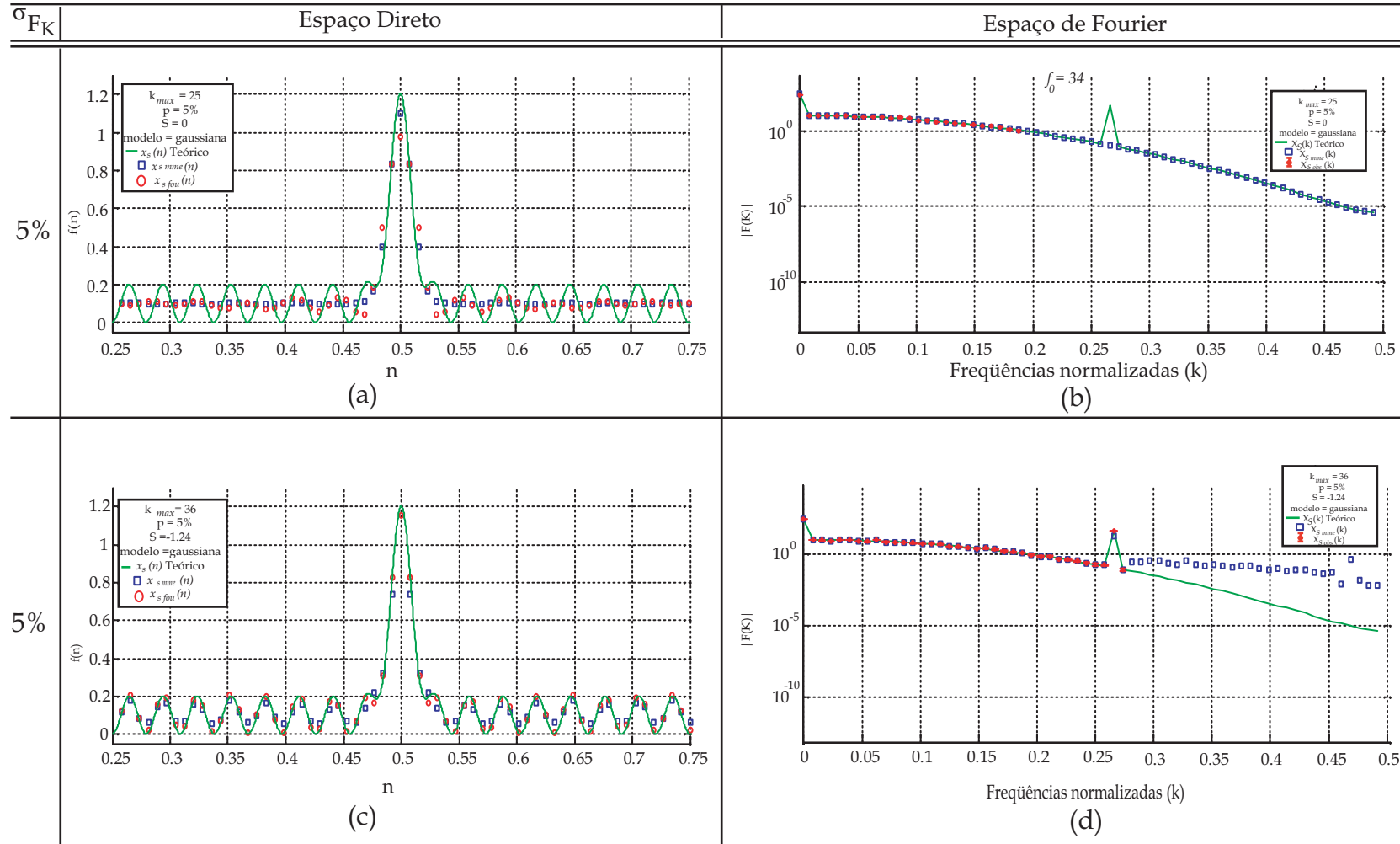


Figura 4.11: Reconstrução da curva utilizando a gaussiana como modelo não uniforme. Quando o conjunto de dados não possui a componente de Fourier do cosseno, o resultado é perfeito e a entropia é máxima  $S = 0$ . No caso contrário, o modelo não está completamente de acordo com o conjunto de dados  $S = -1, 24$ . A entropia permite verificar a imprecisão do modelo.

# Conclusão

Neste trabalho apresentamos uma ferramenta para o cálculo da transformada de Fourier inversa aplicado à reconstrução de sinais medidos no domínio da frequência. A tese abordou os problemas relacionados a realização de medidas neste domínio, como, por exemplo, a discretização do espectro e a utilização de uma janela de observação do mesmo. Estes efeitos comprometem a utilização da  $DFT^{-1}$ , resultando, em alguns casos, em oscilações no sinal estimado e dificultando a análise precisa do sinal. Este problema foi apresentado de forma analítica, demonstrando que as oscilações podem estar ligadas a limitações nos instrumentos de medidas. Além disso, o uso da  $DFT^{-1}$  despreza a informação do erro experimental.

Para tentar minimizar tais problemas estudamos a técnica de estimação de um sinal pela  $DFT^{-1}$  baseada no princípio do Máximo de Entropia (MME). Esta técnica nos permite incluir no cálculo da transformada o conhecimento prévio do resultado esperado para o experimento. Este método utiliza o Teorema de Bayes, onde as probabilidades condicionais são calculadas entre o conjunto de dados e o modelo matemático do resultado esperado. As probabilidades são descritas pelas funções da verossimilhança e da entropia do modelo matemático. A entropia é utilizada como informação a *priori* do experimento. Dois conceitos de entropia foram utilizados: a entropia clássica de Shannon e a entropia generalizada proposta por Shannon-Jaynes.

Uma ferramenta em MATLAB foi desenvolvida permitindo a utilização da inversão de Fourier a máximo de entropia (**idftme**).

Para avaliar o método desenvolvido, realizamos diversas simulações a fim de comparar a utilização da **idftme** e da **idft** na reconstrução de sinais. Quatro estudos de casos foram propostos. O primeiro caso mostrou que o MME reconstrói as informações que estão faltando no conjunto de dados. Esta reconstrução é obtida a partir do modelo teórico utilizado como informação a *priori*. Ainda no primeiro caso, evidenciamos que o erro experimental tem o papel de fazer com que o resultado final seja o mais próximo do modelo teórico, i.e. quanto maior a incerteza mais perto do modelo será o resultado final. Este caso teve como objetivo evidenciar as duas propriedades fundamentais do máximo de entropia: reconstrução de informações que estão faltando e pressão para que o resultado seja o mais próximo do modelo teórico.

O segundo caso utilizou um modelo incorreto, mas que estava em acordo com os dados. Neste caso a solução final é incorreta, entretanto a solução dada pelo MME é correta. O resultado final obtido pelo MME segue inexoravelmente os dados.



O terceiro caso apresentou a utilização de um modelo correto com parte do sinal sem comprovação experimental. Neste caso, o resultado final está de acordo com os dados, no entanto este tipo de reconstrução deve ser usado com muito cuidado, pois o resultado final não está totalmente evidenciado nos dados. Parte do sinal reconstruído está baseada unicamente nas informações do modelo.

O quarto caso abordou questões relativas a validade de um modelo. Um modelo errado foi apresentado, porém com parte dos dados sem comprovação no modelo. Mesmo que se utilize um modelo ruim o resultado final seguirá as informações contidas nos dados experimentais. A entropia tem um papel importante na validação do modelo, pois, por meio dela podemos saber o quanto o sinal reconstruído se diferencia do modelo.

As simulações procuraram ilustrar o interesse de se utilizar o MME, suas vantagens e possíveis armadilhas. Para a maximização da entropia e ajuste do modelo aos dados experimentais, utilizou-se o método de Newton-Raphson. O uso deste método para otimização mostrou instabilidades na convergência, o que já era esperado, pois quando os valores iniciais são escolhidos em um intervalo inadequado, o método não converge. O método de Newton nem sempre é o melhor método, pois é necessário que a estimativa inicial esteja dentro de um intervalo de convergência. Porém, a sua simplicidade e ordem de convergência quadrática fazem com que seja bastante utilizado.

Alguns testes poderiam ser realizados a fim de melhorar o problema de convergência. Uma proposta seria melhorar a parte do código que calcula  $\delta_f$ . As falhas no método de Newton-Raphson podem ter sido causadas por passos que resultaram em valores negativos na distribuição. O MME não é capaz de gerar um sinal com valores negativos. Dessa forma, em algumas situações os dados devem ser processados antes de serem utilizados. Para evitar os erros de convergência nesta condição, podemos verificar a existência de valores negativos, e se eles ocorrerem podemos diminuir  $\alpha$  mais lentamente e realizar uma nova tentativa.

Este trabalho pode ser estendido buscando investigar sinais reais, obtidos em instrumentos, a fim de testar a praticidade do uso da técnica e a convergência do método em situações reais.

Além disso, uma extensão natural é a reconstrução de sinais em 2D a fim de restaurar imagens onde a observação experimental é feita no domínio de Fourier, como, por exemplo, os tomógrafos de raios X<sup>2</sup>.

Outra sugestão para a evolução do trabalho é a substituição da entropia pela entropia relativa generalizada proposta por Tsallis. Esta entropia permitirá a introdução no processo de refinamento do parâmetro de não-extensividade  $q$ .

---

<sup>2</sup>Na tomografia computadorizada, um raio-X é concentrado em um feixe estreito que passa apenas por uma pequena parte do objeto examinado. A intensidade do raio-X chega em um detector e é convertida em um sinal digital. Movendo-se o emissor de raios-X e o detector obtém-se sinais de outros pontos do objeto, em ângulos variados. É o que se chama de “varredura” do feixe. Esse processo é repetido várias vezes para ângulos ligeiramente diferentes. Computadores armazenam os valores da intensidade dos raios-x. Matematicamente, esses valores constituem a “Transformada de Fourier” do objeto examinado. O computador processa essas transformadas e reconstrói uma imagem tri-dimensional do interior do objeto. Nos tomógrafos mais modernos apenas a fonte de raios-X é movida. A detecção é feita em um anel de detectores que envolve o objeto em estudo.

Em instrumentação científica, a teoria que descreve um experimento pode ser pensada como sendo um operador que liga o espaço de medida ao espaço das possíveis conclusões. Quando conhecemos o funcionamento por completo do aparato experimental; isto significa que temos uma teoria para cada parte do instrumento, nos relacionando os dados medidos até a grandeza física de interesse. Para ilustrar isso, vamos considerar, por exemplo, um experimento de difração. Pode ser que a fonte de excitação do sistema físico não seja completamente monocromática. Pode ser ainda que alguns dados não possam ser medidos devido a limitações no instrumento de medida; ou ainda que possam estar comprometidos devido a interferências diversas. Além disso, a resolução e a precisão da medida podem variar muito ao longo da medida. Ao final, podemos rejeitar as medidas muito ruins para termos um conjunto de dados medidos com boa qualidade. Entretanto, do ponto de vista convencional estes dados estão incompletos.

O operador que nos leva do espaço dos dados para o espaço das conclusões tem, normalmente, dimensões maiores para as conclusões que para os dados medidos. Esta diferença cria problemas de instabilidade e resultados distorcidos. A idéia deste trabalho foi de considerar todas as soluções no espaço de conclusões que podem gerar os dados. Existem varias soluções consistentes quando levamos em conta as incertezas experimentais. Entretanto, a escolha de uma das soluções deve seguir algum critério. Se todas as soluções são similares, então não há necessidade de um critério. Porém, podem existir várias soluções diferentes e precisamos de uma regra para escolhê-la. A escolha pelo método do máximo de entropia é uma boa opção e foi isto que buscamos mostrar nesta tese.

## Apêndice A

# Métodos Bayesianos e Máximo de Entropia: História e Filosofia

Este anexo tem por objetivo apresentar uma revisão da abordagem bayesiana e do princípio do máximo de entropia para a resolução dos problemas inversos.<sup>1</sup> As discussões apresentadas são mais filosóficas que técnicas. Discutimos a definição de probabilidade e suas interpretações, da lógica de tomada de decisão bayesiana, e das controvérsias sobre o princípio do máximo de entropia.

Jaynes é um dos principais autores modernos defensores dos métodos bayesianos e do princípio do máximo de entropia. É por isso que esta revisão histórica, é baseada em seus trabalhos e mais particularmente em seus artigos publicados durante a quarta oficina sobre o máximo de entropia e métodos bayesianos de estatística aplicada, editados por [Moh87].

### Lógica intuitiva de tomada de decisão

A necessidade de uma lógica de tomada de decisão em situações de falta de informação em ciências da engenharia e mesmo na vida cotidiana é evidente. Devemos incessantemente tomar decisões ainda que não estejamos certos das suas conseqüências. Por exemplo, uma pergunta simples em um dia nublado seria: “Devo levar um guarda-chuva hoje?”.

Antes mesmo de estudar matemática sabemos, por pura necessidade, como tratar estes problemas devido a um raciocínio lógico intuitivo. A nossa intuição organiza então o procedimento de tomada de decisão em quatro etapas:

1. Enumeração de todas as possibilidades que podem acontecer;
2. Definição de uma probabilidade para cada possibilidade baseado no que vemos (conjunto de dados) e a experiência do passado (*a priori*);
3. Predição das conseqüências prováveis de cada opinião;
4. Tomada de decisão.

---

<sup>1</sup>Este documento foi traduzido e adaptado da tese de doutorado de Mohammad-Djafari [Moh87] e trata de uma revisão histórica e de alguns elementos básicos dos métodos bayesianos e do máximo de entropia.

### Hérodote (500 antes de J.C.) traduz assim à lógica dos reis persas:

*“Uma decisão é sábia, mesmo que esta tenha uma conseqüência desastrosa, se as evidências disponíveis tivessem a indicado como a melhor. Ela não é sábia, mesmo que dê o resultado mais desejado, se as evidências indicassem que ela não era razoável e suas conseqüências improváveis”.*

Este raciocínio qualitativo pode, em certos casos, ser formulado de uma maneira quantitativa por um modelo matemático. Tal lógica é necessária nos diferentes domínios da ciência, onde os dados dos problemas a serem resolvidos não são, em geral, suficientes para nos fornecer soluções únicas a estes problemas.

Bernoulli (1713) chamou este raciocínio de “a arte da conjuntura” traduzindo um estado de conhecimento incompleto por um conjunto de  $N$  acontecimentos equiprováveis que se pode denotar por  $x_1, \dots, x_N$ . Jaynes chamou isto de espaço de hipóteses  $HO$ . O único conhecimento que se tem sobre estes acontecimentos é o número  $N$ .

Consideremos agora uma proposta  $A$ , definida como sendo verdadeira para o subconjunto  $H(A)$  de  $M$  elementos do espaço  $HO$ , sendo a proposta  $A$  falsa para qualquer outro conjunto.  $M$  é a multiplicidade de  $A$ , e a probabilidade de  $A$  então é definida por  $P(A) = \frac{M}{N}$ . Se definirmos as diferentes possibilidades por  $A, B, C$  etc. os seus complementos por  $\bar{A}, \bar{B}, \bar{C}$ , etc. Sendo  $AB$  a proposta “A e B” e  $A | B$  a proposta “A dado que já conhecemos B”; podemos definir, da mesma maneira, as probabilidades  $P(A), P(AB), P(A\bar{B}), P(A | B)$ , etc. Podemos também definir  $P(B | A)$ . Na verdade se sabemos que  $A$  é verdadeiro, o espaço das propostas reduz-se à  $H(A)$  e a probabilidade  $P(B | A)$  é então a proporção da multiplicidade de  $B$  neste subespaço.

Estas regras elementares podem ser utilizadas não somente para encontrar a melhor resposta num jogo de 20 perguntas, mas também em muitos problemas científicos que necessitam, às vezes, de cálculos combinatórios não triviais. Resta, contudo uma dificuldade: a boa determinação de  $HO$ . Se esta é simples para um jogo de dados, é menos evidente em outros problemas, como observou Bernoulli: “*qual mortal pode determinar o número de todas as doenças possíveis?*” Então como determinar a probabilidade de uma doença?

Jaynes baseou seu raciocínio nesta pergunta. Duas atitudes se apresentam. Seja abandonar esta teoria porque não é aplicável a todos os casos, quer dizer que não se trata de fazer onisciência, mas de uma tomada de decisão, “a melhor possível diante do que se conhece hoje”.  $HO$  não é a descrição da realidade final do universo, mas o que conhecemos hoje. Então escolhemos como  $HO$  “todas as doenças conhecidas” e seguimos.  $HO$  não é necessariamente  $HN$ , o espaço de hipóteses da natureza.

Suponhamos que a nossa previsão seja verdadeira, ou seja, que, entre as propostas ( $A_1 \dots, A_m$ ) a proposta  $A_k$  é escolhida como a mais provável se admitimos  $HO$ , sendo ela a mais confirmada pelas observações. Isto não quer dizer que  $HO = HN$ , mas que  $HO$  não é suficientemente diferente de  $HN$  para afetar a nossa previsão. Quanto mais as observações confirmam a nossa escolha, mais certeza nós temos de  $HO$ . Que  $HO$  seja igual à  $HN$  não é realmente importante. O que importa é que  $HO$  tenha um valor (poder) preditivo.

Suponha agora que a nossa previsão seja falsa e que a natureza persiste em tornar verdadeira a proposta  $A_k$ , pois ela é mais freqüente que uma outra proposta que foi escolhida como a melhor para  $HO$ . Neste caso, é necessário concluir que  $HO$  é muito diferente de  $HN$ . No entanto, mesmo neste caso não perdemos o nosso tempo porque os erros nos ajudam sobre a forma como é necessário alterar  $HO$ . Define-se então um melhor espaço das hipóteses  $H1$ , comparam-se as suas previsões com as observações, se necessário, podemos definir  $H2$ , e seguir a diante. Este se trata de um método científico.  $HO$  é um ponto de partida inevitável.

Bernoulli utilizou este raciocínio de maneira mais quantitativa, ele idealizou um método para testar  $HN$  nos casos onde é possível efetuar observações de um acontecimento que tenha um grande número de ocorrência de maneira independente. Jaynes concluiu sobre a sua abordagem: um acontecimento com uma probabilidade ( $M/N$ ) maior sobre  $HO$  deve ser previsto para acontecer mais frequentemente, enquanto um acontecimento com uma probabilidade maior sobre  $HN$  deve ser observado mais frequentemente. Bernoulli foi o primeiro a formular uma relação entre a probabilidade e a freqüência de realização de um acontecimento, hoje conhecida como a lei fraca dos grandes números: se realizamos  $n$  observações independentes e se observamos que o proposição  $A$  ocorre  $m$  vezes, a freqüência  $f(A) = m/n$  pode ser comparada com a probabilidade  $P(A) = M/N$ . Quando  $n \rightarrow \infty$  é quase certo que  $f(A)$  está bem próximo de  $P(A)$ .

### Estatística Bayesiana

Entre os diversos trabalhos de Bayes, o artigo publicado em 1763 apresenta uma solução completamente diferente ao problema de Bernoulli. Enquanto Bernoulli calculou a probabilidade de observar  $A$   $m$  vezes, dado o conjunto de dados  $N$ ,  $n$  e  $M$ , Bayes encontrou uma fórmula para calcular a probabilidade de modo que  $M$  tenha um valor definido, dado  $N$ ,  $n$  e  $m$ . A esta fórmula chamamos hoje de “Estatística Bayesiana”.

Alguns anos mais tarde, Laplace reinterpreta o princípio de Bayes em um de seus primeiros trabalhos publicados (1774) e explica-o com mais clareza e generalidade. Ele o utiliza durante quase 40 anos na resolução de problemas diversos: em astronomia, meteorologia, estatística das populações, e mesmo em direito (jurisprudência).

Os principais teoremas parecem simples e triviais porem são de uma importância real no método de inferência científica. Consideremos, por exemplo, três propostas  $A$ ,  $B$ , e  $C$ ; os seus complementos  $\bar{A}$ ,  $\bar{B}$ , e  $\bar{C}$ ; assim como as propostas  $AB$  (“A e B”) e a proposta  $A | B$  (“A dado B”); então as regras elementares de soma e produto de probabilidades são dadas por:

$$P(A | B) + P(\bar{A} | B) = 1 \quad (\text{A.1})$$

$$P(AB | C) = P(A | BC)P(B | C) \quad (\text{A.2})$$

As propostas  $AB$  e  $BA$  são idênticas, então a coerência lógica destas relações permite alterar a posição de  $A$  e de  $B$  na Eq. A.2. Além disso, se  $P(B | C) > 0$ , temos o que se chama hoje de regra de Bayes, embora Bayes nunca a tenha escrito:

$$P(A | BC) = \frac{P(A | C)P(B | AC)}{P(B | C)} \quad (\text{A.3})$$

Esta equação é na prática um processo de aprendizagem.  $P(A | C)$  é a probabilidade a *priori* quando se conhece apenas  $C$ .  $P(A | BC)$  é a probabilidade a *posteriori*, atualizada com a nova informação  $B$ . Tipicamente  $A$  representa uma hipótese (uma teoria) que queremos verificar,  $B$  representa o conjunto de dados, e  $C$  o conhecimento a *priori* que temos sobre  $A$  antes de observar  $B$ .

Um exemplo resolvido por Laplace é o seguinte:

A : A massa de Saturno  $m_s$  é compreendida entre  $m_1$  e  $m_2$

B : Os dados observados são as perturbações mútuas de Saturno e de Jupiter

C : O bom senso nos diz: a massa  $m_s$  não pode ser tão pequena pois assim Saturno perderia seus anéis e não pode ser tão grande pois Saturno deixaria o Sistema Solar

Laplace deu o seguinte resultado baseando-se nos dados disponíveis no final do século XVIII :  $m_s$  é igual à  $1/3512$  da massa solar, e a probabilidade que esta massa seja diferente em 1% deste valor é 0,99991. Hoje após 150 anos de dados acumulados, tem-se a mesma estimativa com a mesma probabilidade de 0,99991, mas com um erro relativo de 0,63% em vez de 1%.

A regra de Bayes corresponde a nossa intuição porem apresentando um sentido quantitativo. Uma probabilidade como  $P(A | C)$  está necessariamente presente em qualquer método de inferência. A determinação da probabilidade de  $A$  após conhecer  $B$  não pode ter uma resposta única (i.e. não é um problema “bem colocado”) se não conhecemos a probabilidade de  $A$  antes de dispor  $B$ .

A regra de Bayes pode ser aplicada de uma maneira repetitiva cada vez que se acrescenta um elemento de informação. Porém o resultado será idêntico se for aplicado uma só vez com todos os elementos de informação reunidos. Isto nos fornece (mais que a nossa intuição qualitativa) um julgamento sobre a coerência e a importância relativa destes diferentes elementos de informação.

No início do século, Jeffreys redescobriu o raciocínio de Laplace e o explicou melhor que o próprio Laplace. Mas, apesar destes trabalhos, a abordagem bayesiana permaneceu quase trinta anos sem ser realmente compreendida.

Baseado nos trabalhos de Bernoulli, Jaynes enunciou o seu primeiro princípio: “É um “direito constitucional” de utilizar  $HO$  mesmo se é possível que  $HO$  seja diferente de  $HN$ ”.

E, baseado nos trabalhos de Laplace e de Jeffreys, enunciou também seu segundo princípio: “Em dois problemas diferentes se dispomos do mesmo grau de conhecimento sobre uma determinada hipótese, deve-se atribuir a mesma probabilidade a esta hipótese.”

Jaynes apresenta um exemplo simples que ilustra este último princípio: se o espaço  $HO$  for escolhido como uma informação *a priori* que consiste exatamente nos acontecimentos possíveis  $x_i$ , então a única distribuição de probabilidade que se pode atribuir a estes acontecimentos é uma distribuição uniforme. Porque, atribuindo qualquer outra distribuição, a permutação de dois acontecimentos coloca um outro problema com o mesmo grau de conhecimento que no primeiro mas com uma distribuição de probabilidade diferente. Este é um exemplo do princípio geral de invariância por transformação de grupo, devido ao simples fato de atribuir uma distribuição de probabilidade *a priori* para representar a falta de conhecimento (“total ignorância”).

Embora Laplace tenha utilizado este princípio e justificado seu emprego por meio de exemplos concretos, ele não chegou a explicar que esta escolha não é arbitrária mas, resultado de uma lógica coerente para a representação de um grau de conhecimento.

É importante destacar que, em um método de inferência, a palavra probabilidade representa, de uma forma geral, um conhecimento incompleto e não tem necessariamente relação com variáveis aleatórias ou realizações de um dado acontecimento em uma experiência aleatória. Mesmo a noção de repetição no espaço das hipóteses  $HO$  não é necessária. No exemplo de Laplace as duas proposições: “ $m_s > 1/3512$  da massa solar” e “ $m_s < 1/3512$  da massa solar” têm as mesmas probabilidades, mas este caso não trata de um fenômeno aleatório.

Uma probabilidade  $P$  é um termo abstrato, um valor que atribuímos a uma proposta para representar o nosso grau de conhecimento, ou que calculamos a partir das três regras (A.1), (A.2) e (A.3).

Uma frequência  $f$ , numa situação onde falar de repetições tem um determinado sentido, é uma propriedade do mundo real que podemos medir ou estimar.

Assim, como Bernoulli, podemos calcular a probabilidade  $p(f)df$  em que uma frequência  $f$  se encontra em um intervalo  $df$ . Em alguns problemas, nossa informação, mesmo que incompleta, pode nos fornecer uma densidade de distribuição de probabilidade  $p(f)$  bastante concentrada ao redor de uma frequência  $f_0$ . Podemos então estimar esta frequência como um bom grau de confiança. Neste caso, podemos dizer que a teoria das probabilidades das variáveis aleatórias está de acordo com a teoria de Laplace-Jeffreys. Mas este último pode ser aplicável nos casos onde a primeira não tem sentido.

Esta é por exemplo a resolução dos problemas inversos. Porque em um problema inverso, a principal dificuldade é falta de informação. Não se trata de um fenômeno aleatório. Temos exatamente a necessidade de representar nosso conhecimento *a priori*, bem como a nossa incerteza sobre os dados e sobre a solução por probabilidades. Não é porque atribuímos uma probabilidade ou um parâmetro para representar o nosso grau de conhecimento de uma dada informação, ou à uma grandeza medida para representar a nossa incerteza sobre a sua precisão, que isso significa que estamos trabalhando com variáveis aleatórias.

Se as três regras do cálculo das probabilidades de Laplace-Jeffreys permitem combinar e estimar as frequências das variáveis aleatórias, isto não quer dizer que uma probabilidade será definida somente quando puder ser identificada à uma frequência. Esta tem sido uma

confusão de longa data, até que R.J. Cox demonstrou em 1946 que estas três regras são efetivamente as melhores regras de cálculo em um raciocínio indutivo.

Supondo que os graus de aceitabilidade sejam apresentados por números reais, ele encontrou as condições necessárias para o conjunto de regras de cálculo coerentes no sentido que, se existe dois métodos de cálculo diferentes para resolver um problema baseado nestas regras, os resultados devem ser idênticos.

Assim, graças a Cox, existe um teorema que diz que um conjunto de regras de condução de inferência, nas quais os graus de aceitabilidade são representados por números reais, são, necessariamente, equivalentes àquelas de Laplace-Jeffreys ou incoerentes.

Shannon (1948) utilizou o método de Cox para definir uma quantidade que ele posteriormente chamou de entropia, para medir a incerteza associada a um fenômeno aleatório. Mais uma vez as condições de coerência lhe permitiram definir esta medida. É a entropia associado à uma distribuição de probabilidade  $(p_1, \dots, p_N)$  de  $N$  acontecimentos, dado pela equação  $\sum_{i=1}^N p_i \log p_i$ . Esta definição de entropia corresponde a definição da entropia de Boltzmann em termodinâmica.

A palavra entropia tinha sido utilizada primeira vez por Rudolph Clausius em 1846 para descrever uma quantidade que acompanha a mudança da energia calórica em energia mecânica. Gibbs (1876) tinha dado uma interpretação variacional no qual a maximização da entropia de Clausius tivesse permitido estudar com sucesso a termodinâmica no equilíbrio.

Boltzmann (1877), Gibbs (1902) e Von Neumann (1928) apresentaram três métodos variacionais nos quais a maximização da entropia de Boltzmann lhes permitiu prever a entropia de Clausius, de uma maneira teórica, ao mesmo tempo nas teorias clássicas e quântica da termodinâmica, sem efetuar medidas.

Embora o formalismo matemático e elegante de Gibbs tivesse sucesso prático na teoria quântica, não havia ainda uma demonstração clara da sua justificativa teórica. Muitos debates sobre a necessidade da propriedade ergódica das equações de movimentos se sucederam, até o momento onde, finalmente, o princípio de coerência lógica de Shannon mostrou em qual sentido a maximização da entropia gera uma inferência lógica. Que o sistema de equações do movimento seja ergódico ou não, com a informação disponível, o formalismo do máximo de entropia dá as melhores previsões sobre o estado do sistema.

O espaço  $HO$  baseado na mecânica clássica de Gibbs era “ruim” para as atitudes convencionais, porque sua previsão para as equações de estado era correta mas a previsão de certas constantes de equilíbrio era falsa.

O novo raciocínio de Shannon ampliou o domínio de aplicação do princípio do máximo de entropia. Agora era possível utilizar os princípios variacionais de Gibbs e de Von Neumann, seja no estado de equilíbrio ou no estado de não-equilíbrio e mesmo em domínios diferentes da termodinâmica: análise espectral, reconstrução e restauração de imagens, economia, etc.



Consideremos o espaço  $HO$  que enumera as seguintes possibilidades  $x_1 \dots x_N$  e uma evidência  $E$  que não pode ser utilizada como os dados  $B$  na regra de Bayes, porque não se pode definir  $P(E | C)$ , mas que nos permite impor um limite sobre as probabilidades  $p_i = P(x_i)$  que atribuímos aos elementos  $HO$ . Este limite não é suficiente para definir a distribuição  $P_1 \dots P_N$  mas o obriga a não ser mais uniforme. Então, maximizando a entropia de Shannon sob esta condição, pode-se escolher uma entre estas distribuições que satisfazem este limite. Esta escolha é a mais objetiva descrição do nosso grau de conhecimento no sentido que este exprime somente a enumeração das possibilidades e a evidência  $E$ .

A partir de agora, se adquirimos uma informação suplementar  $B$ , podemos recalculamos o nosso grau de conhecimento pela regra de Bayes. Em outras palavras, o princípio do ME nos dá o meio para escapar ao domínio “equiprovável” de Bernoulli e Laplace, dando-nos o instrumento necessário para atribuir leis *a priori* não-uniformes a um parâmetro.

### **Estatísticas ortodoxas e estatísticas bayesianas:**

Entre a estatística ortodoxa e a estatística bayesiana existe uma longa história feita de confusão e de controvérsia que, em certos casos, causou uma paralisia de comunicação. As principais causas são:

- i) uma confusão entre a noção de probabilidade  $P$  e de frequência  $f$ ,
- ii) falta de distinção entre problemas diferentes que têm soluções matemáticas idênticas.

Os debates sobre este assunto datam de 1930 entre Jeffreys e Fisher. Jaynes resumiu assim estas controvérsias através das suas discussões com G. Bernard e apresentou seu ponto de vista bayesiano. Em uma abordagem bayesiana considera-se, a partir dos dados e da informação *a priori*, parâmetros constantes e desconhecidos (mas não aleatórios) no momento em que os dados estão disponíveis. Na estimativa dos parâmetros por meio de uma abordagem bayesiana, as distribuições de probabilidade *a priori* e *a posteriori* representam não uma função mensurável destes parâmetros, mas apenas o nosso grau de conhecimento. A largura da distribuição não indica o domínio de variação dos verdadeiros valores dos parâmetros mas o domínio dos valores coerentes com a nossa informação *a priori* e os nossos dados. Estes valores podem então ser admitidos objetivamente como possíveis valores.

G. Bernard (1984), que é um estatístico ortodoxo, crítica assim a abordagem bayesiana de estimativa dos parâmetros: “como pode-se pretender conhecer a distribuição de um parâmetro a partir dos dados disponíveis que podem nos informar apenas sobre um único valor do parâmetro?”

Jaynes comenta esta crítica afirmando que é exatamente a má compreensão da abordagem bayesiana pelos estatísticos ortodoxos, porque numa abordagem bayesiana não é a distribuição dos parâmetros que buscamos determinar, mas a distribuição das probabilidades. Não são os parâmetros que são distribuídos mas as probabilidades. É necessário distinguir entre dois problemas:

- i) considerar, a partir dos dados e das probabilidades *a priori*, um parâmetro desconhecido constante dando uma probabilidade *a posteriori* medindo o seu grau credibilidade.

- ii) deduzir, a partir do conhecimento *a priori* da distribuição de freqüências de um parâmetro aleatório (sobre uma larga classe  $C$  de repetições de uma experiência aleatória), a distribuição das freqüências deste parâmetro sobre uma classe  $C$  ( $D$ ) de experiências que produziram os dados  $D$ .

Estes dois problemas são tão diferentes que é evidente que não se pode resolvê-lo com o mesmo procedimento. Mas, pode-se, por meio de um raciocínio totalmente diferente encontrar um algoritmo matemático idêntico para resolvê-lo. É isto que aconteceu com estes dois problemas e onde começou a confusão. De qualquer forma, o que é evidente é que a abordagem bayesiana não pode e não deve ser utilizada para resolver o segundo problema. Além disso, embora se trate de um problema matemático interessante, a utilidade deste problema em uma situação real não é evidente por três razões:

- a) Em um problema real os parâmetros são em geral constantes desconhecidas e não variáveis aleatórias.
- b) Mesmo se os parâmetros forem variáveis aleatórias, o que nos interessa é conhecer os seus valores no momento em que a experiência é feita, não a sua distribuição de freqüência em uma classe de experiências imaginárias.
- c) Ainda que nos interessemos a esta distribuição de freqüências, nós não conhecemos a distribuição de freqüência *a priori* da qual se tem necessidade para resolver este problema.

A abordagem bayesiana e o princípio do máximo de entropia consistem em um método de aprendizagem e um instrumento de trabalho para um físico que tem a necessidade de um método que seja aberto ao mesmo tempo para “cima” e para “baixo”. Aberto para cima quer dizer que se possa deduzir qualquer outra proposta a partir de hipóteses e regras definidas. Aberto para baixo quer dizer que estas mesmas regras podem nos permitir testar uma hipótese ou medir a coerência de um conjunto de hipóteses.

A natureza não nos revela seus axiomas e para compreender, estudar e dominar um fenômeno físico, somos obrigados a começar por um ponto no meio. Começando por observações e estabelecendo diversas hipóteses, temos a necessidade de um método de inferência coerente que nos leve a outras conclusões para cima. Temos também a necessidade de agir da forma contrária, para baixo, ou seja a partir dos dados observados verificar a validade das hipóteses. É por isso que a abordagem bayesiana é um instrumento adequado para as ciências, para o engenheiro que se encontra em um ambiente diferente daquele de um matemático.

Se considerarmos agora as duas abordagens, da estatística ortodoxa e da estatística bayesiana, constatamos que a estatística bayesiana é mais adequada para resolver os problemas da física. Por exemplo, na teoria das probabilidades de Kolmogorov, dado as probabilidades dos acontecimentos simples, pode-se calcular as probabilidades dos acontecimentos cada vez mais complexos à ajuda das regras postuladas. Mas as probabilidades dos acontecimentos simples iniciais nos são dadas como grandezas determinadas noutro lugar.

Pelo contrário, na abordagem bayesiana, estas probabilidades necessariamente são condicionadas, desde o início, pelo nosso grau de conhecimento do problema. Elas não podem ser atribuídas arbitrariamente, o que permite, ao contrário da abordagem ortodoxa, uma análise para “baixo” do problema, porém mantendo uma coerência com a abordagem ortodoxa na análise para “cima”. Para um físico esta análise para “baixo” é somente a metade do problema e não pode ser ignorada.

Um exemplo real que pode ilustrar estes argumentos é o da mecânica estatística. Tomemos um pedaço de açúcar. As primeiras observações pelos nossos sentidos revelam-nos um objeto branco na forma de um cubo<sup>2</sup> com faces lisas. Não podemos a princípio, definir um espaço de hipóteses muito útil somente com estas percepções. No entanto, ao observarmos o mesmo por meio de uma lupa teremos um objeto constituído de cristais individuais e idênticos. Podemos por conseguinte definir  $H_0$  enumerando estes cristais e atribuindo-lhes uma distribuição de probabilidade. Podemos então estudar diversas propriedades do açúcar, como por exemplo sua porosidade.

Na mesma direção, podemos imaginar análises mais exaustivas, utilizando instrumentos mais adequados que nos revelem que cada cristal é constituído das várias moléculas arranjadas de uma maneira regular. Podemos por conseguinte definir um espaço  $H_1$  enumerando estas diferentes possibilidades de arranjo e atribuindo-lhes uma distribuição de probabilidade, uniforme, se não conhecemos sua estrutura nominal. Mas se conhecemos esta estrutura nominal (observando-a por exemplo através de uma imagem feita por Raios X) podemos então definir outro espaço  $H_2$  associando uma distribuição de probabilidade não-uniforme que daria uma probabilidade maior à estrutura nominal e estudar por exemplo as probabilidades de existência de redes não-uniformes, etc.

Podemos ir ainda mais adiante e definir  $H_3$ , constatando que cada molécula de açúcar é constituída de 12 átomos de carbono, 22 átomos de hidrogênio e 11 átomos de oxigênio, e enumerar assim a sua posição ou a sua velocidade (o espaço de fase Maxwell e de Gibbs).

Na presente fase, diversas propriedades do açúcar, como por exemplo a sua capacidade calórica, pode ser estudada. Mas um estudo quantitativo correto da capacidade calórica não pode ser concluído se não escolhermos uma distribuição de probabilidade correta para estas grandezas neste espaço.

Indo mais adiante e estudando os componentes (elétrons, prótons, e nêutrons) de cada átomo de  $C$ , de  $H$  e de  $O$ , pode-se definir  $H_4$  enumerando as posições e as velocidades destas componentes. Mas, como Sommerfeld demonstrou, se procedemos da mesma maneira para uma peça metálica,  $H_4$  pode nos dar resultados mais falsos que  $H_3$ , por exemplo, no valor dos coeficientes da capacidade calórica dos metais.

Mas se adicionamos o fato da posição e a velocidade das componentes dos átomos serem conhecidas de uma forma aproximada ou com uma dada precisão não-infinita, nos encontramos assim em um espaço  $H_5$  que consiste em enumerar os estados quânticos, o que fornece previsões mais coerentes com a realidade.

---

<sup>2</sup>em alguns países da Europa o açúcar é comercializado na forma de cubos

Em princípio, nesta etapa, se ignoramos as dificuldades de cálculo, somos capazes de estudar de uma maneira quantitativa todas as propriedades termodinâmicas e químicas do açúcar.

Este é um dos atuais desafios da mecânica estatística, ou seja à enumeração dos estados quânticos globais de um sistema macroscópico, e estes seriam suficientes para prever os parâmetros de equilíbrio das macroestruturas a partir dos dados microscópicos.

Esta simples enumeração dos estados quânticos é, em princípio, suficiente para prever todos os aspectos dos processos irreversíveis; mas o custo de cálculo computacional é tão grande que este assunto continua a ser ainda inexplorado.

Mas o fato de utilizar  $H5$  com sucesso nos mostra que mesmo que exista um outro espaço  $H6$  este não deve ter muita influência sobre as previsões que são feitas hoje com  $H5$ . Mesmo que pensemos em defini-lo os ganhos na prática serão poucos e perderemos muito em termos de custo computacional. Assim embora qualquer realização seja possível, na prática é necessário saber onde parar.

Em problemas mais recentes, como, por exemplo, na reconstrução de imagens, a análise espectral, a inversão de “sismogramas” em geofísica, etc, não estamos ainda nesta etapa. Não houve ainda uma atividade de pesquisa como as propostas por Liouville ou Gibbs para nos guiar nos espaços apropriados, como foi feito por Gibbs para a invariância do volume de fase para as equações do movimento e as transformações canônicas, mostrando que atribuir probabilidades uniformes aos volumes de fase tinha o mesmo significado para qualquer tempo e qualquer sistema de coordenadas. Ele aplicou realmente o princípio da invariância de grupo, proposto anteriormente por Jaynes.

Especificando o espaço das hipóteses, sobre o qual atribuímos um peso uniforme para todos os elementos, antes de levar em conta um pré-requisito a fim de obter uma distribuição de probabilidade (a *priori* não-uniforme que tenha a entropia máxima) é o ponto de partida da abordagem bayesiana. É o que é chamado o estado de “total ignorância” embora se conheça o número destes estados “equiprováveis”.

Laplace tinha notado este problema escrevendo que a apreciação exata destes estados “equiprováveis” é “um mais dos delicados pontos” da teoria das probabilidades.

Duzentos anos após, hoje, estamos ainda discutindo sobre esta “justa apreciação” deste estado equiprovável. Ele é de uma importância crucial para se aprender como traduzir uma informação a *priori* por uma distribuição de probabilidade a *priori*. Todos os problemas inversos têm necessidade de tal método e o avanço de muitos outros domínios é diretamente dependente.

O princípio do ME é explicado por várias abordagens:

- i) combinatórias (Boltzmann, Darwin, Fowler)
- ii) teoria da informação (Shannon, Jaynes)
- iii) coerência lógica (Shore, Johnson, Gull)

iv) pragmáticos e utilitários (Gibbs, Papanicolaou, Mead, Good, Skilling...)

v) codificação e comunicação (Shannon, Rissanen)

Muitos autores recentes concentraram-se preferencialmente na abordagem **(i)** que na **(ii)** pensando que uma vez o procedimento utilizado em situações onde há um significado evidente que torna a utilização do princípio mais confortável, estaremos mais dispostos a ir mais em frente e a utilizar o método em toda sua generalidade.

Contudo, é necessário ser vigilante. Embora todas as abordagens conduzam a um mesmo algoritmo matemático final, elas não descrevem exatamente os mesmos problemas e não é conveniente comparar os seus méritos relativos.

Por outro lado estas diferentes abordagens generalizam-se diferentemente, embora atualmente em algumas das suas aplicações há a impressão que elas se juntam. Mas em outras aplicações eles podem seguir seu próprio caminho. Assim é então efetivamente importante compreender as circunstâncias e os procedimentos do seu uso.

Na dúvida, podemos sempre tentar nos situar na abordagem **(iv)**, ou seja na forma pragmática. Mas no caso de sucesso ou insucesso, não se pode deixar de voltar aos princípios básicos.

No entanto, é necessário se prevenir para não se encontrar na situação onde estão, por exemplo, os procedimentos de análise espectral clássica (o pragmatismo de Tukey) que é bastante utilizado sem se compreender as justificativas teóricas.

O uso do princípio do máximo de entropia e o seu sucesso crescente na resolução de problemas da análise espectral é um outro exemplo da abertura desta abordagem para “cima” e para “baixo”. De fato, esta abordagem permitiu, não somente desenvolver novos algoritmos e obter melhores resultados, mas também dar uma nova interpretação aos algoritmos e técnicas mais antigos e nos ajudou a esclarecer sobre suas hipóteses e seus defeitos.

Hoje novos domínios de aplicação se abrem a esta abordagem. É, por exemplo, o problema da restauração e a reconstrução de imagens, sejam na óptica, astronomia ou tomografia. Estamos apenas no início de um processo de aprendizagem neste domínio. Não é um domínio que possa ser classificado como trivial embora os resultados pragmáticos já obtidos sejam bem animadores. A definição e escolha de um espaço de hipóteses  $H$ , sobre o qual devemos definir nossos “a priori” sobre uma imagem não são ainda bem definidos. Porque, em um problema novo, é normalmente inevitável que diferentes pessoas tenham diferentes “a priori”. As razões são várias:

- Diferentes conhecimentos a priori do fenômeno físico do problema
- Diferentes quantidades de experiências práticas.
- Diferente “treino mental” da experiência passada que estabeleceu uma espécie de canalização das idéias.

Então antes de se tornar um defensor de um ou de outro, é preferível, até agora, começar a relacionar a escolha às circunstâncias de uso. Temos necessidade ainda de várias dezenas de anos de trabalho neste domínio antes de poder afirmar a superioridade de uma escolha sobre a outra.

## Apêndice B

# Inversão de Fourier a Máximo de Entropia

```
% -----  
% Programa      : Inversao de Fourier a Maximo de Entropia  
% Escrito por   : E. MARQUES DA SILVA E M. PORTES DE ALBUQUERQUE  
% email        : emsilva@cbpf.br marcelo@cbpf.br  
% Versao       : 4.1202  
% Compilador    : MATLAB version 6.5.0 Realease 13  
% Projeto      : Tese de Mestrado em Instrumentação  
% -----  
  
function [F,Chi2, Entropia, FlagFim] = idftme(Dados, SigDados, Modelo, R, iterMax, FlagPrint);  
warning off  
TINY=1*10^(-20);  
FlagFim=0;  
  
% Declaração de Variaveis -----  
N=length(Modelo);  
K=length(Dados);  
Alpha=10000;  
AlphaInc = 0.95;  
AlphaAnterior=Alpha;  
rOMax=1;          % valor maximo do passo de DeltaF  
iter = 0;  
Chi2Ant=0;  
LimConv=0.5;  
% Define valor inicial de F=modelo -----  
Modelo=Modelo+TINY;  
F=Modelo;  
  
% Calcula SigDados.^2  
SigDados = SigDados+TINY;  
InvSigDados2 = 1./(SigDados.^2);  
  
% Calcula Temp=Rt [S.^2]R -----  
Temp= zeros(N,N);  
Rasc= zeros(K,N);  
for k=1:K  
    for n=1:N  
        Rasc(k,n)= InvSigDados2(k)*R(k,n);  
    end  
end  
  
for n1=1:N  
    for n2=1:N  
        for k=1:K  
            Temp(n1,n2) = Temp(n1,n2) + R(k,n1)*Rasc(k,n2);  
        end  
    end  
end  
  
if(FlagPrint==1) fprintf('\n'); end  
  
% Loop do Maximo de Entropia -----  
while(FlagFim==0)  
% Calcula Temp2 = inv(alpha 1/F + Temp) -----
```

```

Temp2= zeros(N,N);
for n1=1:N
for n2=1:N
if (n1==n2) Temp2(n1,n2) = Alpha / F(n1) + Temp(n1,n2);
else
Temp2(n1,n2) = Temp(n1,n2);
end
end
end
Temp2 = inv(Temp2);
% Temp3 = Alpha*Log(M/F)
Temp3= zeros(N);
for n=1:N
Temp3(n) = Alpha * log(Modelo(n)/F(n));
end
% Rasc = D - R *F
Rasc= Dados;
for k=1:K
for n=1:N
Rasc(k) = Rasc(k)- R(k,n)*F(n);
end
end
% Rasc = Sig2*(D - R*F)
for k=1:K
Rasc(k) = Rasc(k)*InvSigDados2(k);
end
% Temp3()= Alpha*Log(M/F) + Rt*(Sig2*(D - R*F))] -----
for n=1:N
for k=1:K
Temp3(n) = Temp3(n)+R(k,n)*Rasc(k);
end
end
% DeltaF = inv(alpha 1/F + Rt[S^-2]R)*Alpha*Log(M/F) + Rt*(Sig2*(D - R*F))]-----
for n1=1:N
DeltaF(n1)=0;
for n2=1:N
DeltaF(n1) = DeltaF(n1) + Temp2(n1,n2)*Temp3(n2);
end
end
% r0=DeltaFt [F^-1] DeltaFt -----
Temp4= DeltaF./F;
r0=sqrt(Temp4*F');
if(r0 < r0Max) % verifica se o passo de DeltaF e menor que seu valor maximo -----
% Calcula novo F(i+1)= F(i) + DeltaF-----
for n=1:N
F(n) = F(n) + DeltaF(n);
if(F(n) < TINY) F(n)=TINY;end
end
% Calcula o novo Chi2 -----
Chi2 =0;
for k=1:K
DadosCalculados=sum(R(k,:).*F);
Chi2 = Chi2+ (DadosCalculados-Dados(k))^2*InvSigDados2(k);
end
if(FlagPrint==1)
if(Chi2-Chi2Ant<0)
fprintf('v');
%fprintf('\n(%1.30f), (%1.30f)',Chi2-Chi2Ant, abs(Chi2 - K)); -----
else
fprintf('..');
end
end
Chi2Ant=Chi2;
% Convergencia foi alcancada? -----
if( abs(Chi2 - K) <= LimConv)
FlagFim=1;
else

```



```
AlphaAnterior=Alpha;
if( (Chi2 - K)>0)
    Alpha = Alpha*AlphaInc;
else
    AlphaInc = sqrt(AlphaInc);
    Alpha = Alpha*(1/AlphaInc);
end
end
else
    Alpha = AlphaAnterior;
    AlphaInc = sqrt(AlphaInc);
    Alpha = Alpha * AlphaInc;
end
% Mostra valores intermediarios
iter=iter+1;
if mod(iter,50)==0
    if(FlagPrint==1) fprintf('\nIter(%d)',iter); end
end

% Numero de iter max foi alcancado? -----
if(FlagFim==0 & iter > iterMax)
    FlagFim=-1;
end
end

% S= F(n) - Modelo(n) - F(n)*log(F(n)/Modelo(n))-----
if(FlagFim==-1)
    Entropia=-1; Chi2=-1; F=-1*ones(1,N);
else
    Entropia =0;
    for n=1:N
        Entropia = Entropia + F(n) - Modelo(n) - F(n)*log(F(n)/Modelo(n));
    end
end
end
```

# Referências Bibliográficas

- [A99] Oppenheim A. *Discrete-time signal processing*. Editora Prentice Hall, 2nd edition, (1999).
- [Alb99] M. P. Albuquerque. *Mesure Optimisée de Densités d'Aimantation*. PhD thesis, Institut National Polytechnique de Grenoble - INPG, (1999).
- [AS85] Livesey A.K. and J. Skilling. Maximum entropy theory. *Acta Cryst*, A(41):p.113–122, (1985).
- [BM91] B.Buck and V. A. Macaulay. *Maximum Entropy in Action*. Oxford University Press, Oxford, (1991).
- [Dan91] G. J Daniell. *Of maps and monkeys*. In *Maximum Entropy in Action*. Clarendon Oxford, Oxford, (1991).
- [DB04] Hanselmam Duane and L. Bruce. Matlab 6.5 - guia do usuário. <http://www.mathworks.com/matlabcentral>, junho (2004).
- [DC90] Alan S.S. David, L. D. Iain M.J. and Jeffrey C.H. Does the maximum entropy method improve sensitivity? *Proc. Natl. Acad. Sci. USA*, vol.87:p.5066–5068, Julho (1990).
- [Esq03] I. A. Esquef. *Técnicas de entropia em processamento de imagens*, (2003).
- [Fer73] J.Pedro. Fernandes. *Introdução à Teoria da Probabilidade*. Universidade de Brasília, Brasília, (1973).
- [Gul89] S. F. Gull. *Developments in maximum entropy data analysis*. In *Maximum Entropy and Bayesian Methods*, Cambridge, England . Kluwer Academic Publishers, Dordrecht, (1989).
- [H86] Justice J. H. *Maximum Entropy and bayesian methods in applied statistics*. Cambridge Univ. Press, Cambridge, (1986).
- [HS99] B. Van Veen Haykin S. *Sinais e Sistemas*. Editora John Wiley & Sons Ltd, New York, 1999.
- [Jay57] E. T. Jaynes. Information theory and statistical mechanics. *Physical Review*, vol.106(4):p.620–630, Maio (1957).
- [Jay78] E. T. Jaynes. *Where do we stand on maximum entropy?* In *Maximum Entropy Formalism*, volume 1. M.I.T Press, Cambridge, Dezembro (1978).

- [Jef61] H. Jeffreys. *Theory of probability*. Clarendon Press Oxford, Oxford, 3 edition, (1961).
- [JGP92] D. Manolakis John G. P. *Digital Signal Processing: Principles, Algorithms and Applications*. Editora MacMillan Coll Div, United States of America, 2 edition, (1992).
- [Joh80] J.E. Shore & R.W. Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. on information Theory*, 26(1):p.26–39, Janeiro (1980).
- [Lin95] W. Von. Linden. Maximum-entropy data analysis. *Applied Physics A*, 60(2):p.155–165, (1995).
- [Mat02] A. Veiga Mattos, R. S. Otimização de entropia: Implementação computacional dos princípios maxent e minxent. *Pesquisa Operacional*, vol.22(1):p.37–59, Janeiro a Junho (2002).
- [Mil03] I. Milho. Teoria de probabilidades. pages p.1–5, Dezembro (2003).
- [Moh87] A. A. D. Mohammad. *Synthese de Fourier Multivariables À maximum D'Entropie*. PhD thesis, Universite de Paris-Sud Centre D'Orsay, (1987).
- [Pré54] E. Carlos. Prélat. *Matemática de Las Probabilidades*. Nigar SRL, Buenos Aires, 2 edition, (1954).
- [S98] Mitra S. *Digital Signal Processing, a Computer Based Approach*. Editora Mc Graw Hill, (1998).
- [Sha48] C. E. Shannon. A mathematical theory of communication. *Bell Sus. Tech. J*, vol.27:p.379–423 e 623–656, Julho-Outubro (1948).
- [Ski84] S.F. Gull & Skilling. Maximum entropy image reconstruction. *IEEE Proc.*, vol.131(F):646, (1984).
- [Ski88] J. Skilling. *The axioms of maximum entropy*. In *Maximum Entropy and Bayesian Methods in Science and Engineering*, volume vol.1. Kluwer Academic Publishers, Dordrecht, (1988).
- [Ski89] J. Skilling. *Classic maximum entropy*. In *Maximum Entropy and Bayesian Methods*, Cambrigde, England, volume 36. Kluwer Academic Publishers, Dordrecht, (1989).
- [SS84a] S.F.Gull. and J. Skilling. *The maximum entropy method*. In *J.A. Roberts, editor Indirect Imaging*. Cambridge University Press, Cambridge, (1984).
- [SS84b] S.F.Gull. and J. Skilling. *MEMSYS 3 Users' manual*. *Maximum Entropy Data Consultants Ltd*. Meldreth SG8 6NR, England, 2 edition, (1984).
- [Sta03] G. A. Stark. Using maximum entropy methods to extract dynamics quantities from quantum monte carlo simulations, (2003).
- [Tri78] M. Tribus. *Thirty years of information theory*. In *The Maximum Entropy Formalism*, volume 1. M.I.T Press, Cambridge, Dezembro (1978).

- [TT91] Cover Thomas.M. and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, New York, (1991).
- [YTL84] N. Z Tishby Y. Tikochinsky and R. D. Levine. Consistent inference of propabilities for reproducible experiments. *Physics Review Letters*, vol.**52**:p.1357–1360, (1984).