A inversão harmônica do espectro de
ressonância magnética:
uma solução para o problema dos
autocampos.

Claudio José Magon

Tese apresentada para obtenção do título de Livre Docente

UNIVERSIDADE DE SÃO PAULO - USP
INSTITUTO DE FÍSICA DE SÃO CARLOS - IFSC
SÃO CARLOS, SP, BRASIL

2007

Resumo

Quando o Hamiltoniano de spin é uma função linear da intensidade do campo magnético, h , os campos de ressonância, h_r , podem ser determinados, em princípio, através de uma equação de autocampos. Neste trabalho mostraremos uma nova abordagem do problema da determinação dos campos de ressonância na qual a equação de autocampos se transforma em uma equação dinâmica, especificamente, em uma equação diferencial linear de primeira ordem em uma variável $L(x)$, onde x possui a unidade de campo magnético. Esta equação diferencial possui as propriedades: sua solução estacionária é a equação de autocampos e a informação espectral contida em $L(x)$ está diretamente relacionada com o espectro de ressonância e com os campos de ressonância, h_r . Este procedimento, conhecido pela denominação “Problema da Inversão Harmônica”, HIP, pode ser resolvido pelo “Método da Diagonalização Filtrada”, FDM, resultando em precisão e resolução suficientes para a análise espectral dos sinais dinâmicos. Os princípios básicos nos quais se baseia o FDM serão discutidos na primeira parte deste texto e, na segunda parte, a nova teoria para o problema dos campos de ressonância será apresentada. Serão mostrados alguns exemplos nos quais os campos de ressonância serão precisamente determinados em através de um único procedimento, sem a necessidade de resolver-se equações de autovalores.

Abstract

When the spin Hamiltonian is a linear function of the magnetic field intensity, h , the resonance fields, h_r , can be determined, in principle, by an eigenfield equation. In this work we show a new technical approach to the resonance field problem where the eigenfield equation leads to a dynamic equation or, more specifically, to a first order linear differential equation of a variable $L(x)$, where x has the unit of magnetic field. Such differential equation has the property that: its stationary solution is the eigenfield equation and the spectral information contained in $L(x)$ is directly related to the resonance spectrum and to the resonance fields, h_r . Such procedure, known as the “Harmonic Inversion Problem”, HIP, can be solved by the “Filter Diagonalization Method”, FDM, providing sufficient precision and resolution for the spectral analysis of the dynamic signals. The basic principles underlining the FDM will be discussed in details in the first part of this text and, in the second part, the new theory for the resonance field problem will be presented. Some examples are shown where the resonance fields are precisely determined in a single procedure, without the need to solve eigenvalue equations.

Este trabalho foi parcialmente publicado no

Journal of Magnetic Resonance (communication)

v. 184, p. 176-183, 2007

sob a autoria de:

Claudio José Magon

José Fernando Lima

Ronny Rocha Ribeiro

Mateus José Martins

Dedicamos este trabalho ao

Professor Horácio Carlos Panepucci

Desejando-lhe um futuro brilhante!
Tão brilhante quanto foi o seu passado.

Conteúdo

1	Introdução	1
2	O método da diagonalização filtrada	5
2.1	o HIP é solúvel pela algebra linear	8
2.2	A base de Krilov: KBDM	12
2.3	O problema dos autovalores generalizados	15
2.3.1	O método KBDM-QZ	16
2.3.2	O método KBDM-EVD	17
2.3.3	O método KBDM-SVD	18
2.4	As funções espectrais	22
2.5	Estimativa de erros	24
2.6	Exemplos ilustrativos do KBDM	26
2.6.1	Exemplo 1: Resolução	26
2.6.2	Exemplo 2: Correção de fase e tempo morto	30
2.6.3	Exemplo 3: Resolução artificial	34
2.6.4	Exemplo 4: KBDM vs. DFT	36
2.7	Análise espectral local: FDM	39
2.7.1	Uma breve revisão da DFT	41
2.7.2	A base tipo Fourier	44
2.7.3	Sumário do algoritmo do FDM	54

2.8	Exemplos ilustrativos do FDM	60
2.8.1	Exemplo 1: Resolução	60
2.8.2	Exemplo 2: Alta densidade de picos	64
2.8.3	Exemplo 3: FDM vs. DFT	66
2.8.4	Exemplo 4: Ajuste multi-exponencial	67
2.8.5	Exemplo 5: Picos não Lorentzianos	74
2.8.6	Exemplo 6: Ruído	75
2.9	Conclusões	81
3	O problema dos autocampos	83
3.1	A teoria básica dos autocampos	86
3.1.1	A equação mestra dos autocampos	86
3.1.2	Solução da equação mestra no espaço n^2	88
3.2	Uma nova proposta	92
3.2.1	Considerações gerais	92
3.2.2	A equação diagonal em G	94
3.2.3	A equação diagonal em F	96
3.2.4	Exemplo 1: o Hamiltoniano diagonal	99
3.3	Análise espectral da função de correlação	105
3.4	Exemplo 2: o Hamiltoniano não-diagonal	106
3.5	Exemplo 3: o caso singular	108
3.6	Exemplo 4: o caso não-definido-positivo	111
3.7	Conclusões	115
A	Álgebra linear	117
A.1	Definições	117
A.2	O produto interno	119
A.3	Mudança de base	120

A.3.1	Mudança de base no espaço de Hilbert	120
A.3.2	Mudança de base no espaço não-Hilbert	123
A.4	Autovalores e autovetores	124
A.5	Autovalores à esquerda e diagonalização de matrizes	127
A.6	Ortonormalização da base	129
A.7	Decomposição nos valores singulares	133

Capítulo 1

Introdução

Na maioria dos experimentos de ressonância magnética do tipo “onda contínua” (CW), também chamados “experimentos com varredura de campo”, a frequência da radiação é mantida fixa enquanto que a intensidade do campo magnético é lentamente variada. Este arranjo experimental implica que um parâmetro do Hamiltoniano de spin, o campo magnético neste caso, é variado durante o experimento e que, para alguns valores específicos deste parâmetro, denominados **autocampos**, uma (ou mais do que uma) transição pode ser observada.

O fato de a excitação possuir frequência fixa e o Hamiltoniano variar durante o experimento implica em uma dificuldade ao estabelecimento do procedimento teórico apropriado para determinar-se os valores corretos do campo magnético nos quais as transições podem ocorrer.

Atualmente, a metodologia CW, descrita acima, é pouco utilizada no cenário da Ressonância Magnética Nuclear, RMN. Todavia, esta metodologia CW é a mais utilizada no cenário da Ressonância Paramagnética Eletrônica, RPE, sendo que neste caso as técnicas pulsadas ainda são apenas complementares. Por este motivo, embora a teoria a ser proposta neste trabalho

possa ser aplicada em ambos os casos, utilizaremos exemplos da RPE ¹.

A partir dos experimentos de RPE, é possível extrair-se parâmetros magnéticos e deles deduzir informações dinâmicas e estruturais acerca das moléculas paramagnéticas ou centros paramagnéticos que estão sendo observados. No processo desta pesquisa, é de fundamental importância que o espectro de RPE possa ser simulado numericamente a partir de um modelo teórico e para este propósito os computadores tem sido utilizados durante décadas.

Existem diversos pacotes de programas computacionais disponíveis para executar esta tarefa. Alguns deles são produtos comerciais de alto custo, mas outros podem ser obtidos gratuitamente na comunidade da RPE. Particularmente, estamos familiarizados com dois destes programas gratuitos: com o conhecido QPOWA (dedicado a sistema de spins com $S = 1/2$) [3, 27] e com o recentemente desenvolvido pacote de programs executáveis no ambiente MatLab, denominado EasySpin [30, 31].

Para realizar-se a simulação de um espectro de RPE, a primeira dificuldade que surge é a determinação de todos os campos de ressonância associados a um dado Hamiltoniano de spin que sejam compatíveis com a energia do campo de radiação.

Apesar de existirem diversos algoritmos projetados para realizar esta tarefa, existe apenas um único procedimento conhecido na literatura que permite a determinação direta dos campos de ressonância através de primeiros princípios [2]. Através deste método, o problema de encontrar-se os campos de ressonância é formulado e resolvido de uma forma explícita, mas, como será explicado no Capítulo 3, o procedimento envolvido é computacional-

¹A frequência da radiação que induz transições entre estados de spin se situa na faixa de microondas (10^3 MHz) no caso da RPE, e na faixa das radio-freqüências (MHz) no caso da RMN.

mente inviável, mesmo atualmente, por causa que a ordem das matrizes envolvidas é n^2 , sendo n a dimensão do espaço do Hamiltoniano de spin. Por isso, este método é aplicável apenas em sistemas de baixa dimensão ou em computadores poderosos.

Sendo que não existe um método direto e prático para resolver-se o problema dos autocampos, soluções aproximadas, usualmente baseadas em teorias de perturbação (tanto através de fórmulas explícitas como por meio de tratamentos numéricos), são utilizadas de forma rotineira [29, 33]. Os métodos que utilizam procedimentos de extrapolação são baseados em *frequency-shift perturbation theory*, e são referenciados como *eigenfield perturbation theory* [14]. Algoritmos do tipo *root-finding* são baseados no método de Newton-Raphson e complementados com o método de Muller para realizar partição e interpolação quadrática. Tanto os métodos que utilizam-se de interpolação ou *root-finding* são somente aplicáveis em uma estreita faixa de campo magnético. Métodos de homotopia fazem uso do método dos quadrados mínimos [24] ou passos de Newton-Raphson combinados com *Rayleigh quotient iteration* [13]. O método utilizado no pacote computacional EasySpin implementa um modelamento dos níveis de energia do Hamiltoniano baseado em particionar a escala de campo magnético de uma forma iterativa e auto adaptável [31].

Recentemente, introduziu-se na literatura um procedimento numérico capaz de computar as autoenergias de um sistema quântico de **grande** dimensão, usando soluções de equações dependentes do tempo [22]. Este método foi denominado de “*Filter Diagonalization Method*”, ou FDM, e está sendo utilizado para a análise espectral de sinais dependentes do tempo [20].

Inspirados pelas semelhanças conceituais entre o problema dos autocampos e o enfoque do FDM, nós tivemos a idéia de re-visitar a teoria básica

mencionada acima, que foi proposta há mais de 30 anos atrás por Belford e colaboradores [2], teoria esta que embora seja sempre utilizada como referência ao problema, sempre foi rejeitada devido a dificuldades associadas à sua implementação computacional.

Neste trabalho, focalizaremos nossa atenção a este método de determinação dos autocampos proposto em 1973 [2] e proporemos uma nova formulação do problema dos campos de ressonância baseada na hipótese de que a equação de autocampos pode ser transformada em uma equação dinâmica, isto é, em uma equação diferencial linear de primeira ordem sobre uma variável associada ao campo magnético. A partir da solução desta equação diferencial os autocampos podem ser determinados exatamente sem a necessidade de resolver-se equações de autovalores.

Este enfoque é conhecido como “*Harmonic Inversion Problem*”, ou HIP, e pode ser resolvido através do FDM, alcançando-se suficiente precisão e resolução necessárias para a análise espectral dos sinais envolvidos.

Além do FDM, é possível que exista pelo menos um outro método capaz de realizar a análise espectral no contexto da nossa teoria. Trata-se de um novo método de análise espectral baseado nos polinômios ortogonais de Szegő [17, 5].

No Capítulo 2 introduziremos o FDM e no Capítulo 3 a nova teoria proposta. Acreditamos que, simplesmente por sugerir uma nova teoria, que resulta em um método prático, direto e aplicável à solução do problema dos campos de ressonância, estamos contribuindo para promover um maior entendimento de seus aspectos básicos e, possivelmente, direcionando nossos esforços para uma solução mais definitiva do problema.

Capítulo 2

O método da diagonalização filtrada

A denominação “método da diagonalização filtrada” é a nossa tradução direta a partir da língua inglesa do termo “*Filter Diagonalization Method*”, ou simplesmente FDM, que foi primeiramente introduzido por Neuhauser in 1990 [26] a fim de extrair-se autovalores e autovetores de um sistema quântico de muitos graus de liberdade e em uma dada faixa de energias na qual se tem interesse.

Em essência, aplica-se um filtro inspirado na técnica de Fourier para eliminar-se a correlação existente entre diferentes autovalores que diferem bastante entre si e, assim, através da diagonalização da matriz Hamiltoniana em um pequeno subspaço, os autoestados localizados em uma pequena faixa de energia podem ser calculados precisamente.

Existem duas vantagens da diagonalização filtrada: **primeiro**, através da filtragem a ordem da matriz Hamiltoniana é reduzida; **segundo**, através da diagonalização é possível alcançar-se uma precisão que vai além daquela prevista pelo princípio de incerteza de Heisenberg, $\delta E/\hbar\delta t$, onde δt é o intervalo

de propagação temporal.

Um marco inicial das aplicações do FDM talvez tenha sido o cálculo preciso das ressonâncias do espalhamento quântico do radical HO_2 [21], o que era muito difícil de ser reproduzido através dos outros métodos existentes na época.

Um outro trabalho seminal foi publicado em 1995 por Wall e Neuhauser [34], onde o cálculo da função de correlação temporal e a sua posterior decomposição espectral pôde ser dividida em duas etapas independentes. O fato importante foi a descoberta de que, desde que a função de correlação possa estar disponível, seja por cálculo direto ou por experimentação, a matriz Hamiltoniana pode ser diagonalizada e seus autoestados calculados precisamente.

Nesta nova formulação, o FDM passou a ser aplicável para a análise espectral de sinais temporais, simplesmente ignorando-se o primeiro passo, isto é, a geração destes sinais. Entretanto, esta nova linha de pesquisa permaneceu por mais algum tempo inexplorada até que a eficiência do método matemático fosse sensivelmente melhorada por Mandelshtam and Taylor [23, 22].

Esta nova versão do método foi então denominada de “*Filter Diagonalization Method*”, FDM, e usava um filtro retangular que era computacionalmente mais eficiente do que outros tipos usados nos outros métodos de ajuste de sinais, e foi formulada para tratar o problema da análise de sinais temporais discretos, definidos em uma grade equidistante.

Diante de diversos pontos de vista (convergência, eficiência computacional e estabilidade numérica) mostrou-se que o FDM era o método ideal para realizar o ajuste dos sinais por uma soma de linhas Lorentzianas. A partir de então, o FDM passou a ser bastante explorado e aplicado em muitos problemas, tanto teóricos como experimentais, nas áreas de física e química.

Os trabalhos mais importantes desta fase estão citados e discutidos em uma excelente revisão de Mandelshtam [20], o qual dá ênfase às aplicações do FDM na área da Ressonância Magnética Nuclear, RMN, a qual passou a ser uma das mais interessantes aplicações do método.

Do ponto de vista da análise espectral de sinais de RMN unidimensionais, RMN-1D, concluiu-se que o FDM não possuía limitações intrínsecas. Ele passou a ser uma alternativa viável ao método da Transformada de Fourier, FT, apresentando diversas vantagens (mas também desvantagens!) com relação à FT. A técnica permitia a análise de sinais extremamente longos que resultavam truncados a partir de um certo comprimento, e também de sinais truncados no seu início, isto é, afetados pela conhecida limitação experimental denominada “tempo morto”. Estas dificuldades não se constituíam em uma limitação desta nova técnica.

Por outro lado, os sinais de RMN multi-dimensionais são geralmente fortemente truncados na dimensão indireta devido a diversas limitações práticas, o que ocasiona distorções graves e perda de resolução dos espectros estimados pela FT. Aparentemente, uma das mais interessantes e desafiadoras aplicações do FDM está em suas extensões multi-dimensionais. O FDM é um método que permite processar simultaneamente todas as dimensões dos dados experimentais, de tal forma que a informação adicional codificada nas dimensões mais longas aumenta a resolução das dimensões mais curtas.

A implementação do FDM multi-dimensional na sua forma original requer a diagonalização de múltiplas matrizes de dados, o que é geralmente muito difícil, se não impossível, no caso de sinais temporais ruidosos com ressonâncias degeneradas. Diversas tentativas frustradas, assim como avanços reais tem sido realizados nesta área [20] e parece que este assunto ainda representa um desafio aos teóricos e experimentais da área.

Neste cenário, a análise de sinais temporais produzidos pelas técnicas da Ressonância Paramagnética Eletrônica de Onda Contínua, RPE, e Ressonância Paramagnética Eletrônica Pulsada, RPEP, representa um caso a parte.

Devido ao fato de que todos os tempos envolvidos em RPEP serem cerca de três ordens de grandeza mais curtos do aqueles da RMN, as limitações experimentais impostas à RPEP são muito mais severas. Particularmente, devido ao tempo morto do espectrômetro, os sinais de RPEP são geralmente medidos durante uma fração do seu decaimento total o que ocasiona graves alargamentos de linha e distorções no espectro FT.

As aplicações da FDM nesta nova área estão ainda em uma fase de incubação, sendo que nota-se claramente uma certa empatia com relação aos avanços desta área por uma grande parte da comunidade científica envolvida. Poucas publicações acerca das aplicações do FDM para RPE e RPEP estão disponíveis. Citamos como exemplos [16, 19].

Neste capítulo faremos uma exposição detalhada dos principais conceitos básicos envolvidos na formulação do FDM, deixando para o Apêndice A uma revisão de alguns tópicos da Álgebra Linear que são importantes para a formulação matemática do método.

No próximo capítulo apresentaremos uma aplicação recente do FDM que nós desenvolvemos, e que consiste na determinação computacional dos campos ressonantes da RPE de um sistema descrito por um Hamiltoniano de spin efetivo, linear com o campo magnético.

2.1 o HIP é solúvel pela algebra linear

O objeto de estudo do FDM-1D é ajustar um dado sinal temporal complexo, $c_n = c(t_n)$, definido em uma grade temporal eqüidistante, $t_n = n\tau$,

$n = 0, 1, 2 \dots N - 1$, a uma soma de funções harmônicas exponencialmente moduladas:

$$c_n = \sum_{k=0}^{K-1} |d_k| e^{i\theta_k} e^{-i2\pi n\tau f_k} e^{-n\tau\gamma_k} = \sum_{k=0}^{K-1} d_k e^{-in\tau\omega_k}. \quad (2.1)$$

Esta definição introduz um total de $4K$ incógnitas reais: partes real e imaginária de K amplitudes complexas, $d_k = |d_k| e^{i\theta}$, e partes real e imaginária de K frequências complexas, $\omega_k = 2\pi f_k - i\gamma_k$. A parte real da frequência complexa fornece a posição da linha, f_k , e a parte imaginária é a largura da linha, γ_k . O valor absoluto da amplitude, $|d_k|$, corresponde à integral da linha e o ângulo de fase, θ_k , define a fase da linha.

Embora não seja uma condição necessária, iremos assumir, na maioria dos casos, que γ_k é positivo (isto é, linhas Lorentzianas).

A lista $\{d_k, \omega_k\}$ corresponde à representação paramétrica do sinal temporal, c_n , na qual cada par (d_k, ω_k) , para um dado k , define uma linha (algumas vezes utilizaremos sinônimos para linha: pico ou polo serão também empregados).

Embora o problema de ajuste formulado pela Equação 2.1 seja não-linear, sua solução pode ser obtida através da Algebra Linear.

Para chegar-se a uma proposição do formalismo do FDM assume-se que o sinal complex, c_n , está associado a uma função de correlação temporal de um certo sistema quântico dissipativo, que é descrito por um operador linear, denominado Hamiltoniano, $\hat{\Omega}$, e algum estado inicial Φ_0 :

$$c_n = (\Phi_0|\Phi_n) = (\Phi_0|U^n|\Phi_0) = (\Phi_0| \left[e^{-i\tau\hat{\Omega}} \right]^n |\Phi_0) = (\Phi_0|e^{-in\tau\hat{\Omega}}|\Phi_0), \quad (2.2)$$

onde U é o operador evolução associado (também linear), $\hat{U} = e^{-i\tau\hat{\Omega}}$.

O formalismo que se segue é baseado na hipótese de que os operadores

atuam em um espaço em que o produto interno é complexo-simétrico, isto é, $c_n = (\Phi_0|\Phi_n) = (\Phi_n|\Phi_0)$, no qual não se conjuga o segundo termo e utiliza-se parênteses para distingui-lo do produto interno Hermitiano, $\langle a|b \rangle = \langle b|a \rangle^*$. Além disso, considera-se que $\hat{\Omega}$ é um operador complexo-simétrico (portanto, não-Hermitiano):

$$(\Psi|\hat{\Omega}|\Phi) = \{(\Psi|\}\{\hat{\Omega}|\Phi)\} = \{(\Psi|\hat{\Omega})\}\{|\Phi)\} = \{(\Phi|\}\{\hat{\Omega}|\Psi)\} = (\Phi|\hat{\Omega}|\Psi). \quad (2.3)$$

Assumindo que $\hat{\Omega}$ seja diagonalizável, com autovalores ω_k e autovetores $|\Upsilon_k\rangle$,

$$\hat{\Omega}|\Upsilon_k\rangle = \omega_k|\Upsilon_k\rangle, \quad (2.4)$$

ele pode ser escrito na sua representação espectral,

$$\hat{\Omega} = \sum_k \omega_k |\Upsilon_k\rangle \langle \Upsilon_k|. \quad (2.5)$$

Além disso, assume-se que para qualquer função de $\hat{\Omega}$ vale a suposição:

$$f(\hat{\Omega}) = \sum_k f(\omega_k) |\Upsilon_k\rangle \langle \Upsilon_k|. \quad (2.6)$$

Os autovetores são ortonormais com relação ao produto interno complexo-simétrico, isto é,

$$(\Upsilon_k|\Upsilon_{k'}) = \delta_{kk'}. \quad (2.7)$$

Inserindo 2.6 em 2.2, pode-se recuperar a Equação 2.1:

$$c_n = (\Phi_0|e^{-in\tau\hat{\Omega}}|\Phi_0) = \sum_k e^{-in\tau\omega_k} (\Phi_0|\Upsilon_k) (\Upsilon_k|\Phi_0) = \sum_k d_k e^{-in\tau\omega_k}, \quad (2.8)$$

com

$$d_k^{1/2} = (\Upsilon_k | \Phi_0). \quad (2.9)$$

A partir deste resultado é possível entender porque consideramos inicialmente que o produto escalar é complexo-simétrico e que o operador Hamiltoniano $\hat{\Omega}$ é complexo-simétrico. Em primeiro lugar, se $\hat{\Omega}$ fosse Hermitiano as frequências ω_k resultariam reais e a Equação 2.8 apenas representaria uma soma de componentes harmônicas de amplitudes constantes: o fato de o operador Hamiltoniano não ser Hermitiano produz ω_k complexo, resultando componentes harmônicas moduladas exponencialmente, conforme foi proposto como definição na Equação 2.1. Em segundo lugar, se tivéssemos assumido o produto interno Hermitiano obteríamos $d_k = \langle \Upsilon_k | \Phi_0 \rangle \langle \Phi_0 | \Upsilon_k \rangle = \langle \Upsilon_k | \Phi_0 \rangle \langle \Upsilon_k | \Phi_0 \rangle^* = |\langle \Upsilon_k | \Phi_0 \rangle|^2$, o que representaria apenas componentes harmônicas de fase nula (amplitudes reais).

O problema de ajuste associado à Equação 2.1 se torna então equivalente ao de diagonalizar-se o Hamiltoniano $\hat{\Omega}$ e por isso ele é popularmente denominado de “*Harmonic Inversion Problem*”, ou **HIP**. Equivalentemente, este problema é o mesmo que diagonalizar-se o operador evolução associado, $\hat{U} = e^{-i\tau\hat{\Omega}}$, o qual possui os autovetores comuns com $\hat{\Omega}$:

$$\hat{U}|\Upsilon_k\rangle = e^{-i\tau\hat{\Omega}}|\Upsilon_k\rangle = e^{-i\tau\omega_k}|\Upsilon_k\rangle = \mu_k|\Upsilon_k\rangle. \quad (2.10)$$

A partir dos autovalores de $\hat{U}$ determina-se as posições e larguras das linhas e a partir dos seus autovetores determina-se as amplitudes e fases.

Embora nem $\hat{U}$ nem $\hat{\Omega}$ sejam conhecidos explicitamente, **suas representações matriciais numa base apropriada podem ser calculadas simplesmente através do sinal c_n** . Isto é o que mostraremos a seguir.

2.2 A base de Krilov: KBDM

A base mais simples que pode ser utilizada corresponde a um conjunto de vetores de Krilov $|\Phi_n\rangle$, $n = 0, 1, 2 \dots M - 1$, gerados a partir da propagação do estado inicial $|\Phi_0\rangle$, utilizando-se para isso o operador evolução $\hat{U}$, tal que:

$$|\Phi_n\rangle = \hat{U}^n |\Phi_0\rangle = e^{-in\tau\hat{\Omega}} |\Phi_0\rangle \quad \text{e} \quad c_n = (\Phi_0 | \Phi_n). \quad (2.11)$$

Os elementos de matriz de $\hat{U}^p$ nesta base podem ser trivialmente obtidos:

$$\begin{aligned} U_{nm}^p &= (\Phi_n | \hat{U}^p | \Phi_m) = (\Phi_0 | \hat{U}^n \hat{U}^p \hat{U}^m | \Phi_0) \\ &= (\Phi_0 | \hat{U}^{n+m+p} | \Phi_0) = (\Phi_0 | \Phi_{n+m+p}) = c_{n+m+p}. \end{aligned} \quad (2.12)$$

Sendo que a base de Krilov não é ortonormal, a matriz U_{nm}^0 também precisa ser calculada:

$$U_{nm}^0 = (\Phi_n | \hat{U}^0 | \Phi_m) = (\Phi_n | \Phi_m) = (\Phi_0 | \hat{U}^{n+m} | \Phi_0) = c_{n+m}. \quad (2.13)$$

Note que as matrizes U_{nm}^p são simétricas, o que é uma consequência direta do fato do produto interno ser complexo-simétrico.

Cada função $|\Phi_n\rangle$ é certamente uma combinação linear dos autovetores $|\Upsilon_k\rangle$ de $\hat{U}$. Entretanto, não se pode assumir a priori que o espaço formado pelos vetores de Krilov, $|\Phi_n\rangle$, seja completo, mas, normalmente estes vetores são linearmente independentes se M for da ordem da dimensão K do Hamiltoniano. Então, se M for suficientemente grande pode-se assumir que o conjunto $\{|\Phi_n\rangle\}$ forme uma base válida para se descrever o sistema quântico idealizado. Deste ponto de vista, a base $\{|\Upsilon_k\rangle\}$ poderia ser expandida na base $\{|\Phi_n\rangle\}$ e a seguinte expressão passaria a ser válida:

$$|\Upsilon_k\rangle = \sum_{n=0}^{M-1} [B_k]_n |\Phi_n\rangle. \quad (2.14)$$

Define-se B_k como um vetor coluna que decompõe $|\Upsilon_k\rangle$ na base $\{|\Phi_n\rangle\}$. Inserindo esta equação na 2.10 e multiplicando ambos os lados por $|\Phi_m\rangle$ à esquerda, obtemos uma equação de autovalores generalizada para os autovalores μ_k e autovetores B_k ,

$$\sum_{n=0}^{M-1} U_{mn}^p [B_k]_n = \mu_k^p \sum_{n=0}^{M-1} U_{mn}^0 [B_k]_n, \quad (2.15)$$

ou seja,

$$U^p B_k = \mu_k^p U^0 B_k. \quad (2.16)$$

Em particular, se $p = 1$,

$$U^1 B_k = \mu_k U^0 B_k. \quad (2.17)$$

Analogamente, para $p = 2$ obtemos $U^2 B_k = \mu_k^2 U^0 B_k = \mu_k U^1 B_k$. Generalizando-se este procedimento, obtém-se:

$$U^p B_k = \mu_k U^{p-1} B_k. \quad (2.18)$$

Uma maneira mais formal de escrever esta equação, para todos os valores de k , é:

$$U^p B = U^{p-1} B \mu \quad \text{ou, para } p = 1 : \quad U^1 B = U^0 B \mu, \quad (2.19)$$

onde μ é uma matriz diagonal na qual os autovalores μ_k estão dispostos na diagonal e B é uma matriz na qual cada coluna é o autovetor correspondente ao autovalor na mesma coluna de μ .

Substituindo-se a Equação 2.14 na Equação 2.7 conclui-se que os autovetores B_k devem ser normalizados com relação a U^0 :

$$B_k^T U^0 B_{k'} = \delta_{kk'}. \quad (2.20)$$

Os autovalores, μ_k , da Equação 2.17 fornecem as frequências, ω_k , e a partir dos seus autovetores normalizados, B_k , calcula-se as amplitudes, d_k , através das Equações 2.9 e 2.14:

$$\sqrt{d_k} = (\Phi_0 | \Upsilon_k) = \sum_{n=0}^{M-1} B_{kn} (\Phi_0 | \Phi_n) = \sum_{n=0}^{M-1} B_{kn} C_n = [B^T]_k C, \quad (2.21)$$

onde, $[B^T]_k$ é um vetor linha $(1, M)$, cujos elementos são os da k -ésima coluna da matriz B e $C = (c_0, c_1, \dots, c_{M-1})^T$ é um vetor coluna $(M, 1)$.

Note também que, de acordo com a Equação 2.12, o cálculo de U_{nm}^p para $n, m = 0, 1, 2, \dots, M-1$ requer o conhecimento de c_n para $n = p, p+1, p+2, \dots, p+2M-2$, o que significa que para montar-se a Equação 2.17 é necessário conhecer c_n para $n = 0, 1, 2, \dots, 2M-1$. Logo, conclui-se que $N \geq 2M$ é uma condição necessária para a solução da Equação 2.17.

Esta conclusão é compatível com o princípio informacional, que prevê que o número de informações contidas no sinal, $2N$, deve ser maior ou igual ao número de incógnitas a serem determinadas, $4K$, o que implicaria $N \geq 2K$. Em princípio, a situação $N = 2K$ seria suficiente para levar o problema a uma única solução numérica, pois poderíamos definir a base de Krilov com $M = N/2 = K$ vetores e, então, cada uma das soluções da Equação 2.17 corresponderia a um pico válido.

Entretanto, este número K não é conhecido a priori! Não obstante, este fato não representa um problema desde que o comprimento do sinal, N , seja suficientemente longo ou que o número de componentes, K , não seja

excessivo, isto é, desde que $N/2 \geq M \geq K$.

Embora K não seja conhecido, o valor correto de M (isto é, M que satisfaz a condição $M = K$) poderia, em princípio, ser determinado através da comparação entre os resultados de experimentos numéricos, nos quais o número de vetores da base, M , é sistematicamente variado. Para uma certa região em que $M \geq K$ (mas não $M \gg K$) espera-se que a solução da Equação 2.17 seja numericamente estável. É claro que, quando $M > K$ a base de Krilov fica sobre-completa, implicando em instabilidades numéricas que certamente ocorrerão caso não se tome os certos cuidados especiais que serão levantados e discutidos adiante. Na situação extrema em que $M \gg K$, a base estaria super-completa, o que certamente levaria a problemas numéricos sérios, tais como erros de *roundoff* e *overflow*.

O formalismo matemático delineado nesta seção não possui um nome específico popularizado na literatura. Entretanto, no trabalho de revisão de Mandelshtam [20] ele o denomina de “*Krilov Basis Diagonalization Method*”, ou **KBDM**, a fim distingui-lo da análise espectral local, denominada de FDM, a qual será discutida adiante.

2.3 O problema dos autovalores generalizados

Conforme ficou esclarecido na seção anterior, o KBDM pode ser utilizado para solucionar o HIP e para isso é necessário apenas montar-se e resolver-se a equação de autovalores generalizada, Equação 2.17. Entretanto, como já antecipamos, a tarefa de resolver-se esta equação não pode ser trivialmente (ou, ingenuamente) executada. Apesar disso, os problemas que surgem ao executar-se esta tarefa são puramente matemáticos e podem ser resolvidos a partir de técnicas especiais.

A dimensão das matrizes U^0 e U^1 é M e, se M for maior que o número de componentes espectrais presentes no sinal em análise, isto é, $M > K$, estas matrizes se tornam singulares, possuindo exatamente $M - K$ autovalores nulos e, o sub-espço não-nulo (denominado “*range space of rank K*”) definirá inequivocamente as K frequências e amplitudes do sinal. Por outro lado, se uma pequena quantidade de ruído for acrescentada ao sinal ela será suficiente para tornar as matrizes não-singulares ou, como é normalmente e corretamente especificado, quase-singulares.

Em qualquer uma destas duas situações diz-se que o problema de autovalores está mal-condicionado e por isso ele só pode ser resolvido através de métodos especiais que, felizmente, já são conhecidos há várias décadas. Os dois métodos mais populares, discutidos nos textos de Algebra Linear, são o método QZ [25] e o método GUPTRI [9, 10]. Programas computacionais que resolvem o problema generalizado de autovalores de matrizes singulares pelo método QZ podem ser encontrados facilmente em pacotes disponíveis gratuitamente pela internet, tal como os conhecidos LAPACK e IMSL.

Após realizarmos experimentos numéricos com os diversos programas e técnicas disponíveis na literatura, conseguimos obter resultados satisfatórios com os três diferentes métodos que serão apresentados a seguir.

2.3.1 O método KBDM-QZ

Este método resolve a Equação 2.17 através do conhecido método QZ [25]. Para implementá-lo utilizamos a subrotina DGVCCG do pacote matemático IMSL-FORTRAN.

2.3.2 O método KBDM-EVD

Este é o método proposto por Wall e Neuhauser [34]. Estes autores utilizaram as técnicas conhecidas, “*Eigen Value Decomposition*” e “*canonical orthogonalization*”, aqui designadas por simplesmente **EVD**. Este assunto será sucintamente descrito a seguir mas, para obter maiores detalhes consulte o Apêndice 1.

Primeiro, a matriz U^0 é diagonalizada:

$$U^0 = QSQ^T \quad \text{com} \quad QQ^T = Q^TQ = I, \quad (2.22)$$

onde Q é uma matriz com os autovetores dispostos nas colunas, S é uma matriz diagonal, com os autovalores associados dispostos ao longo da diagonal principal e I é a matriz identidade. Todas as matrizes desta última equação possuem dimensão M . Note que a normalização dos autovetores não é feita de acordo com a norma Euclidiana usual $QQ^H = I$.

O segundo passo consiste em contrair-se as matrizes S e Q , eliminando-se os autovalores que possuem módulo menor que um certo limite mínimo pré-estabelecido, $s_{\min}$, e eliminando-se também as colunas da matriz Q que correspondem a estes autovalores. Se existem L autovalores com módulo maior que $s_{\min}$, as matrizes contraídas $\bar{S}$ e $\bar{Q}$ passam a serem da ordem (L, L) e (M, L) , respectivamente.

O próximo passo é a formação da matriz contraída $\bar{U}^1$ de ordem (L, L) :

$$\bar{U}^1 = \bar{S}^{-1/2} \bar{Q}^T U^1 \bar{Q} \bar{S}^{-1/2}. \quad (2.23)$$

Esta matriz complexa-simétrica, $\bar{U}^1$, é então diagonalizada:

$$\bar{U}^1 = P \mu P^T, \quad (2.24)$$

onde μ é uma matriz diagonal com os autovalores $\mu_k = e^{-i\tau\omega_k}$, $k = 0, 1, 2, \dots, L-1$, e P é uma matriz complexa com os autovetores de $\bar{U}^1$ dispostos nas colunas. Note que estes autovalores são aqueles que buscamos e que correspondem ao conjunto de autovalores válidos da Equação 2.17. Os autovetores correspondentes são obtidos pela equação:

$$B_{jk} = \sum_{n=0}^{L-1} \bar{Q}_{jn} \bar{S}_n^{-1/2} P_{nk}, \quad (2.25)$$

para $j = 0, 1, 2, \dots, M-1$ e $k = 0, 1, 2, \dots, L-1$.

Os autovalores e correspondentes autovetores calculados a partir de 2.24 e 2.25 são de dois tipos: aqueles que resultam nas verdadeiras (precisas) frequências e amplitudes dos picos e, espúrios, associados a artefatos provocados por singularidades remanescentes.

2.3.3 O método KBDM-SVD

O terceiro e último método que utilizamos foi a decomposição da matriz U^0 nos seus valores singulares, método este que é conhecido por “*Singular Value Decomposition*”, ou simplesmente **SVD**. Para maiores detalhes a respeito do SVD consulte o Apêndice 1.

Este método é bastante similar ao anterior, porem difere no fato de que o anterior decompõe a matriz U^0 nos seus autovalores e este a decompõe nos seus valores singulares. A diferença é que, no anterior, os autovalores eram números complexos e neste, os valores singulares são reais e positivos. Como veremos adiante, isto facilita a implementação de critérios para contrair-se a matriz U^0 de uma maneira mais eficiente.

Primeiro, decompomos a matriz U^0 :

$$U^0 = L D R^T, \quad \text{com} \quad LL^T = L^T L = I \quad \text{e} \quad RR^T = R^T R = I. \quad (2.26)$$

A matriz D é uma matriz diagonal na qual os elementos da diagonal principal são números reais positivos, isto é, $D = \text{diag}(d_i)$ com $d_i \geq 0$. Os elementos da diagonal de D são chamados de “**valores singulares**” da matriz U^0 . Normalmente os valores singulares são ordenados em ordem decrescente, $d_1 > d_2 \dots > d_n$, de tal forma que o maior deles ocupa o canto superior esquerdo da matriz D .

O segundo passo consiste em contrair-se as matrizes L , R e D , eliminando-se os valores singulares que são menores que um certo limite mínimo pré-estabelecido, $d_{\min}$, e eliminando-se também as colunas da matriz L e linhas de R^T que correspondem a estes valores singulares eliminados. Se o número de valores singulares retidos é L , as matrizes contraídas, $\bar{U}^0$, $\bar{L}$, $\bar{D}$ e $\bar{R}^T$, passam a ser da ordem (M, M) , (M, L) , (L, L) e (L, M) , respectivamente.

O próximo passo é a formação da matriz contraída $\bar{U}^1$ de ordem (L, L) :

$$\bar{U}^1 = \bar{D}^{-1/2} \bar{L}^T U^1 \bar{R} \bar{D}^{-1/2}. \quad (2.27)$$

Esta matriz, $\bar{U}^1$, é então diagonalizada:

$$\bar{U}^1 = P \mu P^T, \quad (2.28)$$

onde μ é uma matriz diagonal com os autovalores $\mu_k = e^{-i\tau\omega_k}$, $k = 0, 1, 2, \dots, L-1$, e P é uma matriz complexa com os autovetores de $\bar{U}^1$ dispostos nas colunas. Note que estes autovalores são aqueles que buscamos e que correspondem ao conjunto de autovalores válidos da Equação 2.17. Os correspondentes autovetores da Equação 2.17 são obtidos pela equação:

$$B_{jk} = \sum_{n=0}^{L-1} \bar{R}_{jn} \bar{D}_n^{-1/2} P_{nk}, \quad (2.29)$$

para $j = 0, 1, 2, \dots, M-1$ e $k = 0, 1, 2, \dots, L-1$.

Os autovalores e correspondentes autovetores calculados a partir de 2.28 e 2.29 são de dois tipos: aqueles que resultam nas verdadeiras (precisas) frequências e amplitudes dos picos, e espúrios associados a artefatos provocados por singularidades remanescentes.

Do ponto de vista numérico, a técnica SVD é mais eficiente que a EVD por um motivo bastante sutil. Para entender isso, lembremo-nos que foi mencionado na seção anterior que as matrizes U^0 e U^1 são, em grande parte das situações, patológicas (tecnicamente, diz-se que as matrizes são *ill-conditioned*). Em outras palavras, estas matrizes são singulares, ou quase-singulares, sendo que este segundo caso representa uma dificuldade numérica ainda mais grave que a do primeiro.

Quando a matriz U^0 é singular, ou quase-singular, pode ser possível, por inspeção, decidir-se quantos L autovalores (e valores singulares) são muito pequenos a ponto de gerarem instabilidades numéricas e, a partir daí, realizar-se o truncamento adequado das matrizes de transformação, Q , ou, L e R .

A nossa experiência adquirida ao tratar muitos exemplos práticos nos diz que esta “inspeção” é mais simples e mais precisa quando se trata de selecionar valores singulares nulos ou quase-nulos, do que selecionar-se autovalores (complexos) de módulo nulo ou quase-nulo. Particularmente, quando os elementos da matriz estão afetados por ruído, a matriz deixa de ser (exatamente) singular para se tornar quase-singular.

Em muitos casos práticos, o efeito deste processo faz com que os autovalores nulos deixem de ser nulos para se tornar quase indistinguíveis daqueles não-nulos da matriz original sem ruído. Por outro lado, os valores singu-

Figura 2.1: Distribuição típica dos autovalores, $|s|$, em (a) e (c), e valores singulares, d , em (b) e (d), da matriz U^0 . Para este exemplo utilizamos dados simulados, com $N = 1024$, $K = 6$ e $M = 50$. Nos gráficos (c) e (d) uma pequena quantidade de ruído Gaussiano, de valor rms igual a 5% da amplitude do sinal, foi adicionada. A linha tracejada assinala uma possível posição para o truncamento das matrizes de transformação. Em (c) este posicionamento não é óbvio.

lares são mais sensíveis a esta “imperfeição” da matriz, o que propicia um truncamento mais correto das matrizes de transformação.

Para exemplificar o que dissemos, a Figura 2.1 compara, para uma situação típica, os autovalores e os valores singulares da matriz U^0 . Nota-se claramente que, no caso do método KBDM-SVD, o nível adequado para o

truncamento pode ser mais facilmente escolhido.

Infelizmente, nem o primeiro método, nem o segundo, podem dar a palavra final sobre este assunto. Particularmente, quando o sinal está sobreposto a um ruído excessivo, nem o KBDM-EVD, assim como o KBDM-SVD facilitam a escolha de um truncamento adequado. Nestes casos, escolher-se um nível de truncamento pode vir a ser um procedimento de risco, visto que picos genuínos podem ser removidos.

2.4 As funções espectrais

Após a determinação da lista $\{d_k, \omega_k\}$ costuma-se buscar pelo correspondente espectro $S(\varphi)$, que é função de um parâmetro real φ identificado como frequência. Entretanto, esta função, $S(\varphi)$, não é única e aqui discutiremos duas maneiras diferentes de expressá-la [20].

A forma mais simples de se obter um espectro a partir da lista $\{d_k, \omega_k\}$, aqui denominado de “**espectro FT**”, é baseada na Transformada de Fourier da função contínua $c(t)$:

$$S(\varphi) = \int_0^\infty c(t) e^{it\varphi} dt = \sum_k d_k \int_0^\infty e^{it(\varphi - \omega_k)} dt = \sum_k \frac{d_k}{i(\varphi - \omega_k)}. \quad (2.30)$$

O segundo método corresponde à Transformada de Fourier Discreta (ou DFT) da função discreta $c(n\tau)$ [20]:

$$\begin{aligned} S^\tau(\varphi) &= \tau \sum_{n=0}^{\infty} \left(1 - \frac{\delta_{n0}}{2}\right) c_n e^{in\tau\varphi} = \tau \sum_k d_k \sum_{n=0}^{\infty} \left(1 - \frac{\delta_{n0}}{2}\right) e^{in\tau(\varphi - \omega_k)} \\ &= \sum_k \tau d_k \left(\frac{1}{1 - e^{i\tau(\varphi - \omega_k)}} - \frac{1}{2} \right). \end{aligned} \quad (2.31)$$

Na primeira igualdade, o termo $(1 - \frac{\delta_{n0}}{2})$ multiplica c_0 por $1/2$ a fim de corrigir o erro que é introduzido ao aproximarmos a transformada de Fourier contínua por uma soma discreta.

Denominamos este espectro de “**espectro DFT**”. Quando $\gamma_k \tau \ll 1$ teremos $S(\varphi) \approx S^\tau(\varphi)$ mas, em geral, os dois espectros diferem entre si. De fato, o espectro DFT é periódico em φ , com um período igual ao intervalo de Nyquist $2\pi/\tau$.

A Figura 2.2 mostra, através de um exemplo, que os dois espectros calculados destas duas maneiras diferentes são, de fato, diferentes entre si. Observe que, enquanto que o espectro FT tem a forma correta de uma linha Lorentziana (a absorção é simétrica e a dispersão antisimétrica com relação ao centro, e ambas tendendo a zero nos extremos), o espectro DFT é periódico (o que significa que a DFT requer continuidade entre o primeiro e o último ponto), o que causa uma deformação na forma da linha.

Para a maioria das aplicações o espectro denominado de “**absorção**” é o que contem as informações relevantes e, nos dois casos, pode ser expresso como a parte real dos espectros FT ou DFT:

$$A(\varphi) = \text{Re}\{S(\varphi)\} \quad \text{ou} \quad A(\varphi) = \text{Re}\{S^\tau(\varphi)\}, \quad (2.32)$$

desde que ambos espectros estejam com a fase correta, isto é, que d_k seja real. Ainda neste caso, o espectro denominado “**dispersão**” corresponde à parte imaginária:

$$D(\varphi) = \text{Im}\{S(\varphi)\} \quad \text{ou} \quad D(\varphi) = \text{Im}\{S^\tau(\varphi)\}, \quad (2.33)$$

Figura 2.2: Espectros de absorção calculados a partir das Equações 2.30 (FT) e 2.31 (DFT) para um sinal temporal de uma única componente espectral. Usando a notação da Equação 2.1, os parâmetros do sinal de entrada são: $K = 1$, $\omega = 10$, $|d| = 10.0$, $\theta = 0.0$, $1/\gamma = 0.02$, $N = 50$ e $\tau = 0.02$.

2.5 Estimativa de erros

A solução do KBDM produz em geral uma lista $\{d_k, \omega_k\}$ que representa não somente os picos genuínos como também os picos espúrios. Com a finalidade de remover da lista os picos espúrios foram criados alguns métodos especiais [22, 20]. Nesta seção apresentaremos apenas o método mais simples, que é baseado na fórmula de recorrência ditada pela Equação 2.16.

A lista poderá conter picos espúrios mesmo quando o ruído presente no sinal em análise é desprezível. Em um primeiro caso, isto pode ocorrer quando a base $\{|\Phi_n\rangle\}$, $n = 0, 1, 2, \dots, M-1$ está super-dimensionada, isto é, $M > K$, assim como quando ela não é completa, $M < K$. Nestes casos, alguns dos

autovalores, μ_k , e autovetores, B_k , da Equação 2.17, não convergem aos valores corretos que se esperaria para a lista $\{d_k, \omega_k\}$. Quando o sinal é excessivamente ruidoso e o sinal é amostrado em um número muito reduzido de pontos N , (e, lembrando que a dimensão da base é sempre limitada a $M = N/2$), a base pode ser insuficiente para o ajuste simultâneo do sinal de interesse e do ruído presente.

A identificação dos autovalores e autovetores espúrios pode ser realizada de várias maneiras e uma delas consiste em resolver-se a Equação 2.16 para dois valores diferentes de p (por exemplo, $p = 1$ e $p = 2$) e então, por inspeção, comparar-se os resultados obtidos. Embora esta técnica pareça ser pouco automática, felizmente, a estrutura do problema permite que se resulte em um método prático. Considerando que os autovetores são normalizados por $B^T U^0 B = I$, obtemos a partir da Equação 2.16:

$$\begin{aligned} U^2 B_k &= \mu_k^2 U^0 B_k \\ B_k^T U^2 B_k &= \mu_k^2 B_k^T U^0 B_k = \mu_k^2. \end{aligned} \quad (2.34)$$

Aceitando-se somente os autovalores que satisfazem $B_k^T U^2 B_k - \mu_k^2 < \varepsilon_k^2$, onde ε_k é suficientemente pequeno, os autovalores espúrios podem ser selecionados. Um outro procedimento consiste na inspeção a posteriori dos “erros” calculados a partir de:

$$\varepsilon_k = \sqrt{B_k^T U^2 B_k - \mu_k^2}, \quad (2.35)$$

com a ressalva de que estes “erros” são apenas números indicativos (quanto maior pior!), isto é, eles não representam erros estatísticos que permitiriam escrever-se $\mu_k = \mu_k^m \pm \varepsilon_k$, com μ_k^m um certo “valor médio”.

Tendo em vista que se tem o interesse em calcular-se também os erros

associados à determinação dos autovalores, μ_k , necessita-se também determinar a matriz U^2 , além das U^0 e U^1 necessárias para a execução do KBDM descrito pela equação de autovalores 2.17. Com isto, este novo procedimento vai requerer o conhecimento de c_n para $n = 0, 1, \dots, 2M$, ou seja, o conhecimento de $N = 2M + 1$ pontos experimentais, onde M é a dimensão da base.

2.6 Exemplos ilustrativos do KBDM

Nesta seção mostraremos alguns exemplos ilustrativos, para os quais utilizaremos dados de entrada, c_n , simulados numericamente. Além disso, aproveitaremos a oportunidade para esclarecer alguns detalhes da técnica e também introduzir alguns novos conceitos.

2.6.1 Exemplo 1: Resolução

Iniciaremos nossa pequena série de exemplos com um caso simples, porém não-trivial, que pode ser resolvido analiticamente.

Este exemplo, adaptado de [20], corresponde a um sinal de entrada que possui apenas duas componentes harmônicas de frequências (reais), Ω_1 e Ω_2 , e amplitudes (reais), a_1 e a_2 , respectivamente; isto é, consideraremos o sinal temporal:

$$c_n = a_1 e^{-in\tau\Omega_1} + a_2 e^{-in\tau\Omega_2}. \quad (2.36)$$

O objetivo é determinar a_1 , a_2 , Ω_1 e Ω_2 , **todos reais**, a partir da série $\{c_n\}$, $n = 0, 1, \dots, N - 1$.

Sendo que $K = 2$ e, de acordo com o que concluímos anteriormente, o problema tem solução única se $M = K = 2$ e $N = 2M = 4$. Portanto, vamos

assumir que temos os quatro valores de c_n :

$$\begin{aligned}
 c_0 &= a_1 e^0 + a_2 e^0 = a_1 + a_2 \\
 c_1 &= a_1 e^{i\tau\Omega_1} + a_2 e^{i\tau\Omega_2} = a_1 \eta_1 + a_2 \eta_2 \\
 c_2 &= a_1 e^{i2\tau\Omega_1} + a_2 e^{i2\tau\Omega_2} = a_1 \eta_1^2 + a_2 \eta_2^2 \\
 c_3 &= a_1 e^{i3\tau\Omega_1} + a_2 e^{i3\tau\Omega_2} = a_1 \eta_1^3 + a_2 \eta_2^3,
 \end{aligned} \tag{2.37}$$

com $\eta_1^n = e^{in\tau\Omega_1}$ e $\eta_2^n = e^{in\tau\Omega_2}$.

O problema agora é resolver a seguinte equação de autovalores generalizada de dimensão $(2, 2)$:

$$\begin{pmatrix} c_1 & c_2 \\ c_2 & c_3 \end{pmatrix} \begin{pmatrix} [B_k]_1 \\ [B_k]_2 \end{pmatrix} = \mu_k \begin{pmatrix} c_0 & c_1 \\ c_1 & c_2 \end{pmatrix} \begin{pmatrix} [B_k]_1 \\ [B_k]_2 \end{pmatrix}. \tag{2.38}$$

Os autovalores, μ_k , com $k = 1, 2$, podem ser encontrados a partir das raízes da equação secular:

$$\det \begin{pmatrix} c_1 - \mu_k c_0 & c_2 - \mu_k c_1 \\ c_2 - \mu_k c_1 & c_3 - \mu_k c_2 \end{pmatrix} = 0, \tag{2.39}$$

de onde se obtém:

$$(c_0 c_2 - c_1^2) \mu_k^2 - (c_0 c_3 - c_1 c_2) \mu_k + (c_1 c_3 - c_2^2) = 0. \tag{2.40}$$

Substituindo os valores de c_n dados na Equação 2.37, encontramos que os autovalores são as soluções da seguinte equação do segundo grau:

$$(\eta_1 - \eta_2) \mu_k^2 - (\eta_1^2 - \eta_2^2) \mu_k + \eta_1 \eta_2 (\eta_1 - \eta_2) = 0. \tag{2.41}$$

Observe que as amplitudes, a_1 e a_2 , foram eliminadas. Após resolver-se esta equação encontramos os dois autovalores μ_k , com $k = 1, 2$: $\mu_1 = e^{-i\tau\omega_1} = \eta_1 = e^{-i\tau\Omega_1}$ e $\mu_2 = e^{-i\tau\omega_2} = \eta_2 = e^{-i\tau\Omega_2}$, assim como era esperado, pois isto implica que as duas frequências obtidas pelo ajuste são: $\omega_1 = \Omega_1$ e $\omega_2 = \Omega_2$.

O próximo passo consiste em encontrar-se os autovetores da Equação 2.38 e normalizá-los de acordo com a Equação 2.20:

$$\begin{pmatrix} [B_k]_1 & [B_k]_2 \end{pmatrix} \begin{pmatrix} c_0 & c_1 \\ c_1 & c_2 \end{pmatrix} \begin{pmatrix} [B_k]_1 \\ [B_k]_2 \end{pmatrix} = 1. \quad (2.42)$$

Finalmente, usando a Equação 2.21:

$$\sqrt{d_k} = \begin{pmatrix} [B_k]_1 & [B_k]_2 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \end{pmatrix}, \quad (2.43)$$

encontramos as duas amplitudes, $d_1 = a_1$ e $d_2 = a_2$, como também era esperado.

Este exemplo serve para mostrar que, dado um sinal de entrada sem ruído e composto por duas componentes harmônicas, os parâmetros espectrais podem ser determinados, dentro da precisão numérica, a partir de apenas quatro pontos do sinal de entrada. Além disso, a escolha dos quatro pontos não depende de quão próximas são as duas frequências.

Este resultado é importante e é bastante diferente do que se obtém com a técnica da transformada de Fourier discreta, DFT, onde se requer uma grade muito fina, com espaçamento menor que $|\Omega_1 - \Omega_2|$ e, portanto, requer um número muito grande de dados de entrada, N , para resolver-se duas linhas muito próximas, não obstante a alta relação sinal/ruído do sinal.

Isto significa que a DFT não consegue levar vantagem do fato do sinal ter uma alta relação sinal/ruído, enquanto que no caso do KBDM a alta relação sinal/ruído é automaticamente convertida em aumento da resolução.

Embora tenhamos tratado aqui, por simplicidade, de um sinal com apenas duas componentes, estes resultados podem ser generalizados para sinais com número arbitrário de componentes, para os quais a resolução é determinada apenas pela precisão do cálculo numérico.

Para ilustrar ainda mais o KBDM pretendemos resolver novamente este mesmo exemplo, considerando que apenas são conhecidos $N = 2$ pontos experimentais. Embora esta não seja a situação ideal, podemos através dela conhecer um pouco a respeito das propriedades do método.

Se $N = 2$ então a única possibilidade é tomarmos $M = 1$. Neste caso o problema de autovalores generalizado é de dimensão $(1, 1)$:

$$c_1 B_1 = \mu_1 c_0 B_1. \quad (2.44)$$

Portanto, existe somente um autovalor:

$$\mu_1 = e^{-i\tau\omega_1} = \frac{c_1}{c_0} = \frac{a_1 \eta_1 + a_2 \eta_2}{a_1 + a_2}. \quad (2.45)$$

O autovetor, que é apenas um número, é normalizado de acordo com $B_1 c_0 B_1 = 1$, e resulta:

$$B_1 = \frac{1}{\sqrt{c_0}}. \quad (2.46)$$

Portanto, a amplitude do único pico que pode ser determinado é dada por:

$$d_1 = B_1 c_0 = a_1 + a_2. \quad (2.47)$$

Estes resultados mostram que uma única componente harmônica, de amplitude $a_1 + a_2$ e de frequência que corresponde a uma certa média ponderada entre as duas frequências verdadeiras, pode ajustar o sinal. A Equação 2.45 diz que, embora as duas frequências verdadeiras, Ω_1 e Ω_2 , sejam números reais, a solução dada por ω_1 é complexa. Disto conclui-se que o único pico ajustado aqui está em uma posição intermediária entre Ω_1 e Ω_2 e que ele possui uma largura, que tenta levar em consideração a existência de dois picos separados por $|\Omega_1 - \Omega_2|$.

Em outras palavras, o espectro obtido, dado por um único pico Lorentziano, é de baixa resolução mas, apesar disso, ele “tenta” ser fiel às características espectrais do sinal original.

Portanto, esta segunda parte do Exemplo 1 demonstra que, mesmo quando a base está sub-completa ($M < K$), o KBDM se mantém “bem-comportado”. Note também, que no limite em que $|\Omega_1 - \Omega_2| \rightarrow 0$, as Equações 2.45 e 2.47 dão o resultado correto que corresponderia a um sinal com uma única componente.

2.6.2 Exemplo 2: Correção de fase e tempo morto

O segundo exemplo é simples e consiste de seis linhas Lorentzianas igualmente espaçadas no intervalo de frequência (10,20), conforme mostra a Tabela 2.1. Os outros parâmetros foram escolhidos arbitrariamente, com o objetivo de facilitar-se a observação do espectro. Observe que enquanto todas as componentes possuem a mesma amplitude, as correspondentes fases e constantes de decaimento diferem entre si.

Tabela 2.1: Parâmetros do sinal temporal mostrado na Figura 2.3.

k	f_k	$1/\gamma_k$	$ d_k $	$\theta_k(\text{rad})$
1	10	1.0	10	$\pi/3$
2	12	2.0	10	$\pi/4$
3	14	2.0	10	$\pi/4$
4	16	3.0	10	$\pi/5$
5	18	3.0	10	$\pi/5$
6	20	4.0	10	$\pi/6$

Para sintetizar-se o sinal é também necessário, além dos parâmetros citados, especificar-se o **intervalo de amostragem**, τ , (ou, *dwel time* em in-

glês) e o **tempo morto**, t_d , (ou *dead time*). O primeiro já foi definido anteriormente e, por “tempo morto”, queremos dizer que o primeiro dado adquirido, c_0 , arbitrariamente atribuído na Equação 2.1 a $n = 0$, ou $t = 0$, não corresponde realmente ao instante $t = 0$, mas, que ele possui um tempo de atraso, t_d , com relação ao instante inicial real.

Em outras palavras, os dados de entrada estão truncados, ou seja, lhes falta uma parte inicial de duração t_d . Sendo que os parâmetros da Tabela 2.1, especificamente as amplitudes e fases, referem-se ao instante inicial real, o espectro calculado pelo KBDM, a partir dos dados de entrada truncados, não poderá fornecer os parâmetros $|d_k|$ e θ_k corretos que estão mostrados na Tabela 2.1. Mais adiante explicaremos também como se pode corrigir apropriadamente as amplitudes e fases do espectro neste caso.

A Figura 2.3 mostra em (a) o sinal temporal simulado com um total de 2048 pontos, tomando-se $\tau = 0.02$ e $t_d = 1.0$. Em (b) é mostrada uma ampliação do sinal em torno dos $N = 13$ pontos (marcados em vermelho) que serão utilizados. No caso em que $N = 13$ a teoria prevê que uma base com dimensão $M = (N - 1)/2 = 6$ vetores será suficiente para determinar-se as $K = 6$ componentes espectrais e também os erros associados aos autovalores.

De fato, conforme mostra a Figura 2.3(c) o cálculo é bem sucedido em determinar o espectro com precisão. Por simplicidade, somente a parte real (absorção) do espectro está sendo mostrada. A Figura 2.3(d) mostra as amplitudes, $|d_k|$, de cada componente harmônica, que correspondem também às áreas de cada um dos picos mostrados em (c).

Com relação aos gráficos das Figuras 2.3(c,d) duas observações importantes devem ser consideradas. Primeiro, as amplitudes mostradas em (d) não correspondem às amplitudes reais definidas na Tabela 2.1. De fato, nesta tabela todas as amplitudes são iguais a 10 enquanto que as mostradas diferem

Figura 2.3: Resultados do exemplo 2, baseado nos parâmetros da Tabela 2.1, com $N = 13$, $\tau = 0.02$, $t_d = 1.0$ e $M = 6$. Em (c,d) espectros sem correção do tempo morto e (e,f) com correção do tempo morto.

de 10. Na verdade, as amplitudes calculadas em (d) correspondem àquelas no instante $t = t_d$ e não ao que se refere a tabela, isto é, $t = 0$. Em segundo lugar, observa-se que a parte real do espectro calculado possui a forma correta de um espectro de absorção, não obstante o fato de as fases terem sido tomadas arbitrárias a priori. Isto ocorre porque o gráfico mostrado em (c) corresponde ao espectro calculado e posteriormente manipulado para resultar no espectro de absorção correto. Esta manipulação posterior recebe o nome de “**ajuste de fase de ordem zero**” e é realizada simplesmente através da substituição de todas as amplitudes complexas calculadas pelo seu módulo

correspondente, isto é:

$$\text{Ajuste de fase de ordem zero: } d_k \rightarrow |d_k|. \quad (2.48)$$

Entretanto, considerando que existe um tempo morto não nulo, este ajuste de fase não é suficiente para resultar nas amplitudes corretas. Nos gráficos (e) e (f) estão mostrados o espectro e as amplitudes, ambas manipulados posteriormente através de dois processos. Primeiro, é realizada uma correção devido à existência do tempo morto e em seguida a correção de fase de ordem zero.

As frequências dos picos e suas larguras não são afetadas pela presença do tempo morto: somente as amplitudes o são. Por isso:

$$\begin{aligned} c_n &= \sum_k d_k e^{-in\tau\omega_k} = \sum_k [d_k e^{-it_d\omega_k}] e^{-i(n\tau-t_d)\omega_k} \\ &= \sum_k D_k e^{-i(n\tau-t_d)\omega_k} \quad \text{para } n = 0, 1, \dots, N-1. \end{aligned} \quad (2.49)$$

Portanto, $D_k = d_k e^{-it_d\omega_k}$ é a amplitude da componente k no instante de tempo $t = -t_d$ (o mesmo que $n = 0$), ou seja, D_k é a amplitude inicial que seria obtida caso o sinal não houvesse sido truncado. Assim, para cancelar-se o efeito do tempo morto, substitui-se as amplitudes calculadas, d_k , por D_k :

$$\text{Correção devido ao tempo morto: } d_k \rightarrow d_k e^{-it_d\omega_k}. \quad (2.50)$$

Aplicando-se esta correção obtém-se as amplitudes corretas mostradas Figura 2.3(f) e, se em seguida é feito o ajuste de fase de ordem zero, obtém-se também o espectro correto da Figura 2.3(e).

2.6.3 Exemplo 3: Resolução artificial

Para este exemplo o nosso objetivo é introduzir o conceito de “aumento artificial da resolução” e, para isso, utilizaremos os parâmetros da Tabela 2.2. O sinal simulado de acordo com estes parâmetros consiste de dois quintopletos centrados nas frequências -4.4 e 2.8, para os quais as separações entre as linhas valem 0.2 e 0.1, respectivamente.

Os resultados obtidos estão mostrados na Figura 2.4, para os quais foi utilizada uma base de dimensão $M = 10$. Em (a) está o sinal temporal (2048 pontos), tomando-se $\tau = 0.085$, ligeiramente abaixo do limite de Nyquist, $\tau < 1/(2f_k^{max}) \approx 0.1$, e $t_d = 0.0$. Em (b) é mostrada uma ampliação do sinal em torno dos $N = 21$ pontos (marcados em vermelho) que serão utilizados. No caso em que $N = 21$ a teoria prevê que uma base com dimensão $M = (N - 1)/2 = 10$ vetores será suficiente para determinar-se as $K = 10$ componentes espectrais e também os erros associados aos autovalores.

De fato, conforme mostra a Figura 2.4(c) o cálculo é bem sucedido em determinar o espectro com precisão. Por simplicidade, somente a parte real (absorção) do espectro, após correção de fase de ordem zero, está sendo mostrada. A Figura 2.4(d) mostra as amplitudes, $|d_k|$, de cada componente harmônica, que correspondem também às áreas de cada um dos picos mostrados em (c), as quais estão de acordo com os valores dados na Tabela 2.2.

As Figuras 2.4(c,e,f) mostram uma propriedade interessante do KBDM: a lista $\{d_k, \omega_k\}$ contém mais informação do que o espectro convencional de DFT.

Por exemplo, algumas linhas que parecem singletos em um espectro DFT, não são necessariamente singletos! Elas parecem singletos porque os múltiplos picos que as formam são largos e se sobrepõem. No cenário do DFT, estes múltiplos picos “podem” ser resolvidos através do emprego de filtros

Tabela 2.2: Parâmetros do sinal temporal mostrado na Figura 2.4.

k	f_k	$1/\gamma_k$	$ d_k $	$\theta_k(\text{rad})$
1	-4.0	2	10	$\pi/1$
2	-4.2	3	10	$\pi/2$
3	-4.4	3	10	$\pi/3$
4	-4.6	3	10	$\pi/4$
5	-4.8	2	10	$\pi/5$
6	2.6	2	10	$\pi/1$
7	2.7	3	10	$\pi/2$
8	2.8	3	10	$\pi/3$
9	2.9	3	10	$\pi/4$
10	3.0	2	10	$\pi/5$

apropriados. Estes filtros, algumas vezes muito agressivos, melhoram a resolução às custas de uma severa distorção da sua forma e do aumento dos artefatos produzidos pelo ruído.

Se as formas de linha são realmente Lorentzianas, é possível construir uma outra representação (espectro) da lista, na qual as larguras das linhas são reduzidas por um fator $\alpha > 1$, tal que:

$$\text{resolução artificial:} \quad \omega_k \equiv 2\pi f_k - i\gamma_k \quad \rightarrow \quad \omega_k \equiv 2\pi f_k - i\gamma_k/\alpha. \quad (2.51)$$

Aqui, denominamos o processo de “resolução artificial”, conforme diz o título da seção, ao processo denominado “*linewidth cheating*”, introduzido por Mandelshtam [20]. O espectro de absorção do exemplo em estudo está mostrado na Figura 2.4(c), para $\alpha = 1.0$. Em (e) e (f) utilizou-se $\alpha = 3.0$ e $\alpha = 6.0$, respectivamente, onde se observa o aumento da resolução com o

Figura 2.4: Resultados do exemplo 3, baseado nos parâmetros da Tabela 2.2, com $N = 21$, $\tau = 0.085$, $t_d = 0.0$ e $M = 10$. Em (c,e,f) espectros para diferentes valores do parâmetro de resolução artificial α .

aumento de α .

2.6.4 Exemplo 4: KBDM vs. DFT

Com este exemplo queremos fazer uma comparação pontual entre as técnicas KBDM e DFT. Embora estas duas técnicas sejam aparentemente direcionadas a solucionar o mesmo problema, elas são diferentes, tanto na sua concepção quanto nos objetivos.

Em particular, a transformada de Fourier é uma técnica extremamente geral, válida para qualquer função. Por outro lado, o KBDM é um procedimento de ajuste que se aplica somente para sinais harmônicos modulados exponencialmente (linhas Lorentzianas no caso de sinais amortecidos).

É importante que se diga:

(a) O KBDM e o FDM **não são transformações**: o KBDM e o FDM são **procedimentos de ajuste** de sinais (*fitting procedures*) baseados no molde da Equação 2.1.

(b) Portanto, O KBDM e o FDM **não são e nem tentam ser** equivalentes à transformada de Fourier, FT, ou à transformada de Fourier discreta, DFT, as quais **são transformações**.

Sendo que, sem dúvida, a técnica da DFT é mais prática, mais robusta, de maior rapidez e, além disso funciona como uma “*black box*”, ela deve sempre ser prioritariamente utilizada **nos casos em que ela funciona bem!**

Um dos casos em que a DFT não funciona bem será discutido neste exemplo, e é uma adaptação de um caso mostrado em [11]. Nesta referência os autores introduziram um novo método denominado “*cross-term averaging*” para resolver o problema e, aqui, utilizaremos o KBDM para resolver este mesmo problema.

O exemplo consiste na decomposição espectral de um sinal cujos parâmetros estão mostrados na Tabela 2.3. São apenas cinco componentes espectrais, com decaimentos relativamente curtos: o mais curto (0.013) é da mesma ordem que o período da maior frequência (76).

A Figura 2.5(a) mostra o sinal amostrado com $N = 2048$, $\tau = 0.004$ e $t_d = 0.0$ (sinal em cor preta). Em (f) está mostrada a parte real da DFT deste sinal, que corresponde ao espectro de absorção pois, de acordo com a tabela, as fases são nulas.

Tabela 2.3: Parâmetros do sinal temporal mostrado na Figura 2.5.

k	f_k	$1/\gamma_k$	$ d_k $	$\theta_k(\text{rad})$
1	34	0.034	2.6	0
2	40	0.050	1.0	0
3	50	0.040	3.0	0
4	56	0.032	2.3	0
5	76	0.013	1.7	0

A dificuldade associada a este exemplo é que o sinal verdadeiro não corresponde a este que mencionamos, mas sim a um outro truncado, afetado por um tempo morto $t_d = 0.062$. Levando em consideração este tempo morto, a parte real da DFT não mais irá coincidir com o espectro de absorção; de fato, o espectro (módulo da DFT) obtido para $N = 2048$, $\tau = 0.004$ e $t_d = 0.062$ está mostrado em (e) e é bem distinto do espectro de absorção mostrado anteriormente.

No cenário da DFT, nem sempre é possível ajustar-se a fase de um espectro e, por isso, analisa-se o módulo da transformada que, por ser sempre positivo (e apesar de ser mais alargado), fornece um resultado parecido com o espectro de absorção. Porém, nem sempre este é o caso; e, de fato, este não é o caso presente. O artigo citado discute este problema com muitos detalhes e também justifica que o ajuste de fase apropriado deste espectro não pode ser obtido dentro do cenário da DFT.

Todavia, o espectro correto do sinal truncado pode ser obtido com precisão através do KBDM e as Figuras 2.5(c,d) mostram os resultados obtidos. Nota-se que o espectro (c), obtido pelo KBDM, com os mesmos parâmetros adotados para o espectro DFT (e), é idêntico ao (f), que foi obtido pela DFT com $t_d = 0$. Nota-se também que as saliências do espectro (c) ocorrem nas

Figura 2.5: Resultados do exemplo 4, baseado nos parâmetros da Tabela 2.3, com $N = 11$, $\tau = 0.0045$ e $M = 5$. Para referência, em (f) está mostrada a DFT do sinal calculado com $t_d = 0.0$. Em (c) está o resultado do FDM e em (e) o resultado do DFT, ambos para $t_d = 0.062$.

freqüências que foram estipuladas na tabela e, que o mesmo não ocorre com o espectro DFT (e). Para se ter um perfeito posicionamento das linhas de absorção, poder-se-ia aplicar-se a técnica de aumento artificial da resolução que foi discutida no exemplo anterior.

2.7 Análise espectral local: FDM

Pudemos concluir, a partir da teoria apresentada e dos exemplos, que o KBDM, como solução do HIP, é um método robusto e fácil de ser implementado. Entretanto, e infelizmente, este problema está longe de ser facilmente

resolvido nas situações práticas mais comuns. Existe mais do que um problema envolvido que merece ser discutido aqui.

Em primeiro lugar, e como já constatamos, a dimensão das matrizes envolvidas na Equação 2.17 é determinada pelo número de vetores da base de Krilov. Para sinais muito longos, que representam um número grande de componentes espectrais, a dimensão destas matrizes cresce e a solução da equação de autovalores generalizada se torna computacionalmente impraticável, pois o tempo de execução das rotinas conhecidas para executar esta tarefa cresce com o cubo da dimensão das matrizes e, a quantidade de memória necessária para armazenar cada matriz cresce com o quadrado de sua dimensão. Além disso, o processamento de matrizes grandes levam a problemas numéricos (*roundoff* e *overflow*) difíceis de serem evitados.

Em segundo lugar, existem problemas ainda mais sérios que estes já mencionados! Para entendê-los é preciso notar que, na vida real, não existem sinais sem ruído. Até mesmo sinais calculados através de simulações numéricas estão sujeitos à presença de ruídos provocados pelo fato de que o número de algarismos significativos dos computadores reais é finito. Com relação a resultados experimentais, a presença do ruído é uma característica física natural. Resumidamente, podemos adiantar que a presença do ruído requer $M > K$, porque o ruído também possui sua composição espectral e esta será tratada no mesmo pé de igualdade com o “sinal sem ruído” de interesse. Em princípio, K vetores da base serão necessários para ajustar o sinal de interesse e $M - K$ vetores serão necessários para descrever o ruído. Obviamente, sendo que o ruído possui uma densa composição espectral, sinais ruidosos vão requerer um número muito maior de amostragens.

Concluindo, o formalismo simples baseado na base de Krilov que descrevemos não pode ser aplicado, com os computadores atuais, para sinais

reais (ruidosos) que possuem mais do que cerca de alguns milhares de pontos experimentais. Isto porque, no caso de sinais experimentais reais (com ruído), a base primitiva de Krylov produz um sistema linear mal-condicionado e de grande dimensão (as matrizes U^0 e U^1 se tornam enormes e densas), o que faz do KBDM uma escolha impraticável.

Felizmente, existem maneiras de se evitar a diagonalização de grandes matrizes de dados em um único passo e uma delas consiste na implementação de um filtro de Fourier sobre a base de Krilov. Aplicando-se uma transformação unitária apropriada sobre as matrizes U^0 e U^1 , a diagonalização pode ser efetuada em múltiplos passos. Mais especificamente, em vez de analisar-se todo o espectro de uma única vez, é possível dividi-lo em pequenas janelas espectrais e analisá-las separadamente. Esta possibilidade de análise espectral local é uma propriedade importante do FDM.

2.7.1 Uma breve revisão da DFT

Antes de nos aprofundarmos nos detalhes do FDM, vamos aqui fazer uma breve revisão de alguns conceitos importantes da conhecida Transformada de Fourier Discreta, DFT.

A transformada de Fourier, FT, de um sinal contínuo e complexo, $c(t)$, definido no intervalo de tempo $(-\infty, +\infty)$ pode ser estimada através do mesmo sinal amostrado em uma grade temporal finita e equidistante, $c_n = c(t_n) = c(n\tau)$, com $n = 0, 1, \dots, N-1$ e τ o intervalo de amostragem (*dwell time*), da seguinte maneira:

$$F_j = F(\varphi_j) = \int_{-\infty}^{+\infty} c(t) e^{i\varphi_j t} dt \approx \tau \sum_{n=0}^{N-1} c_n e^{in\tau\varphi_j}. \quad (2.52)$$

O conjunto $\{\varphi_j\}$, denominado de “grade de frequências da DFT”, é formado pelos números reais:

$$\varphi_j = -\frac{\pi}{\tau} + \frac{2\pi}{N\tau}j \quad \text{com} \quad j = 0, 1, \dots, N-1. \quad (2.53)$$

A soma da Equação 2.52 é chamada de “Transformada de Fourier Discreta” (DFT) dos N pontos c_n , a qual é uma transformação linear e unitária e que portanto possui inversa. Usualmente, escolhe-se $N = 2^k$, com $k = 0, 1, 2, \dots$, porque existe um algoritmo de execução rápida, denominado “*Fast Fourier Transform*”, ou FFT, que é específico para calcular a soma da Equação 2.52 nesta situação particular.

A Figura 2.6 mostra um diagrama com as escalas de tempo e frequências, $\varphi_j/(2\pi)$, associadas aos índices correspondentes.

Figura 2.6: A estrutura da DFT para $N = 16$.

Conforme assinalado na Figura 2.6, um parâmetro importante da DFT é a sua “**resolução**”. Resolução é a capacidade de distinguir-se componentes do sinal que estão próximas em frequência e a resolução da DFT é limitada pelo chamado “princípio de incerteza tempo-frequência”, que é definido por:

$$\frac{1}{\delta\varphi} = \frac{N\tau}{2\pi}, \quad (2.54)$$

cujos inverso corresponde ao intervalo entre duas frequências consecutivas da grade.

Isto significa que, quanto maior a resolução, menor será o espaçamento, $\delta\varphi$, entre duas componentes observáveis e, a resolução cresce linearmente com o intervalo de tempo no qual o sinal c_n foi definido: $N\tau$. Portanto, para aumentar-se a resolução é necessário aumentar-se o número de pontos, N , ou o intervalo de amostragem, τ .

Todavia, se o objetivo for o de aumentar-se a resolução, o intervalo de amostragem não pode ser aumentado arbitrariamente, pois, para qualquer intervalo de amostragem, τ , existe uma frequência especial, φ_c , chamada de “frequência crítica de Nyquist”:

$$\varphi_c = \frac{2\pi}{2\tau} = \frac{\pi}{\tau}. \quad (2.55)$$

A frequência crítica de Nyquist é importante por causa do conhecido “teorema da amostragem”:

“Se uma função contínua, $c(t)$, é amostrada em intervalos de tempo τ , a largura de banda fica limitada a frequências menores que $\varphi_c = \pi/\tau$ ”,

ou

“Se $F(\varphi) = 0$ para $|\varphi| > \varphi_c$ a função $c(t)$ está completamente determinada pelo sinal $\{c_n\}$ ”.

Em outras palavras, embora a informação contida no sinal discreto é muito menor do que aquela contida no correspondente sinal contínuo, toda a informação que o sinal contínuo contém dentro de uma dada banda, φ_c , pode ser obtida através do sinal discreto, desde que este seja amostrado em intervalos de tempo $\tau < 2\pi/\varphi_c$.

Se esta condição não for obedecida, qualquer frequência fora do intervalo, $(-\varphi_c, \varphi_c)$, será artificialmente (ou falsamente) translada para dentro do intervalo, provocando artefatos no espectro que dificultarão a identificação das

freqüências genuínas. Este fenômeno é denominado (em inglês) por *aliasing* e está discutido na maioria dos livros textos.

2.7.2 A base tipo Fourier

Qualquer combinação linear entre as funções da base de Krilov pode formar uma outra base. Dentre as possíveis bases que poderiam ser encontradas, uma escolha útil é a seguinte:

$$|\Psi_j\rangle = \sum_{n=0}^{M-1} e^{in\tau\varphi_j} |\Phi_n\rangle = \sum_{n=0}^{M-1} e^{in\tau\varphi_j} \{e^{-in\tau\hat{\Omega}} |\Phi_0\rangle\} = \sum_{n=0}^{M-1} e^{in\tau(\varphi_j - \hat{\Omega})} |\Phi_0\rangle, \quad (2.56)$$

onde, daqui em diante, M representa a máxima dimensão possível para a base de Krilov e $\{\varphi_j\}$, $j = 0, 1, 2, \dots, J-1$, é um conjunto de valores (freqüências), equidistantes ou não, tomados dentro da faixa de freqüências de interesse. Em princípio, J pode ter qualquer valor inteiro positivo. A máxima dimensão da base de Krilov depende do número de pontos do sinal de entrada, mas também do cálculo que será efetuado. Vimos anteriormente, que para a determinação da lista $\{d_k, \omega_k\}$ a dimensão máxima da base é $M = N/2$ mas, vimos também que para o cálculo dos erros associados aos autovalores se requer $M = (N-1)/2$.

Existem várias possibilidades para a escolha do conjunto de freqüências $\{\varphi_j\}$. Primeiro, e lembrando que não tem sentido escolher freqüências fora do intervalo de Nyquist, escolhe-se a grade equidistante esquematizada na Figura 2.7(a) e composta pelos $J = M$ valores:

$$\varphi_j = -\frac{\pi}{\tau} + \frac{2\pi}{M\tau}j \quad j = 0, 1, \dots, M-1. \quad (2.57)$$

Com esta escolha, a transformação da base de Krilov, $\{|\Phi_n\rangle\}$, $n = 0, 1$,

Figura 2.7: Três possibilidades para a escolha da grade $\{\varphi_j\}$, $j = 0, 1, \dots, J-1$, utilizada no FDM, para $N = 15$. (a) $J = M = N/2 = 7$, corresponde a uma transformação unitária. (b) Grade equidistante com $J = 24$. (c) Grade não-equidistante com $J = 24$.

$\dots M-1$, para a base de Fourier, $\{|\Psi_j\rangle\}$, $j = 0, 1, \dots, M-1$, é unitária.

É interessante observar também que a transformação 2.56 pode ser interpretada através da projeção de $|\Psi_j\rangle$ no estado inicial $|\Phi_0\rangle$:

$$\tilde{C}_j = (\Phi_0|\Psi_j) = \sum_{n=0}^{M-1} e^{in\tau\varphi_j} (\Phi_0|\Phi_n) = \sum_{n=0}^{M-1} c_n e^{in\tau\varphi_j}. \quad (2.58)$$

Comparando-se esta equação com a definição da DFT em 2.52 conclui-se que, no caso em que a escala dos valores φ_j é a definida em 2.57, o vetor $\tilde{C}$ é a transformada de Fourier discreta dos M pontos iniciais do sinal $\{c_n\}$.

Daqui por diante, o símbolo $\sim$ sobre vetores ou matrizes será utilizado para identificar o uso da base de Fourier.

A definição 2.57 não é a única escolha possível para o conjunto $\{\varphi_j\}$ e, de fato, esta pode não ser a escolha mais apropriada. As Figuras 2.7(b,c)

mostram outras duas possibilidades, com $J = 24$ escolhido arbitrariamente. Em (b) os valores φ_j são equidistantes e em (c) estão distribuídos aleatoriamente no intervalo de Nyquist. Qualquer uma destas duas escolhas, assim como uma infinidade de outras mais, poderiam ser usadas no método que iremos descrever adiante. Por exemplo, uma quarta escolha com $J = 5$ também poderia ter sido sugerida.

Um parâmetro utilizado para caracterizar-se uma determinada grade de freqüências é a sua densidade, que expressa o número de pontos da grade por unidade de freqüência. Assim, se uma certa grade possui L pontos no intervalo de freqüência (φ_a, φ_b) , com $\varphi_b > \varphi_a$, então a sua densidade seria $L/(\varphi_b - \varphi_a)$. Utilizando-se desta definição, a grade de Fourier (Equação 2.53) teria a densidade $N/(2\pi/\tau) = N\tau/(2\pi)$. Assim, se expressarmos por ρ a densidade de uma grade qualquer, relativamente à da base de Fourier, obtemos:

$$\rho = \frac{2\pi}{N\tau} \frac{L}{\varphi_b - \varphi_a} = \frac{2\pi L}{N\tau(\varphi_b - \varphi_a)} \quad (2.59)$$

Em particular, se $\varphi_b - \varphi_a = 2\pi/\tau$ e $L = J = M = N/2$, como no caso (a) da Figura 2.7, então a densidade vale 0.5, o que significa que esta escala possui metade da densidade da escala da DFT. Analogamente, se escolhermos $L = N\tau(\varphi_b - \varphi_a)/(4\pi)$, obteremos $\rho = 0.5$. Para o caso (b) da Figura 2.7, a densidade vale $\rho = (2\pi)(24)/[(15\tau(2\pi/\tau))] = 1.6$. Por outro lado, no caso (c) a grade não é equidistante e a densidade global seria ainda a mesma ($\rho = 1.6$), mas a densidade local, calculada em intervalos $(\varphi_b - \varphi_a) < 2\pi/\tau$, pode depender de qual intervalo está sendo considerado.

Vamos agora analisar algumas conseqüências destas definições.

Primeiro: cada função da base tipo Fourier é localizada no domínio da freqüência, isto é, cada nova função, $|\Psi_j\rangle$, é uma combinação linear de somente

aquelas autofunções $|\Upsilon_k\rangle$ de $\hat{\Omega}$ para as quais $\omega_k \approx \varphi_j$.

Para demonstrar esta afirmação, realiza-se a soma da Equação 2.56 analiticamente e usa-se a representação espectral de $\hat{\Omega}$:

$$\begin{aligned}
 |\Psi_j\rangle &= \sum_{n=0}^{M-1} e^{in\tau(\varphi_j - \hat{\Omega})} |\Phi_0\rangle = \left\{ \sum_{n=0}^{M-1} [e^{i\tau(\varphi_j - \hat{\Omega})}]^n \right\} |\Phi_0\rangle \\
 &= \frac{1 - e^{iM\tau(\varphi_j - \hat{\Omega})}}{1 - e^{i\tau(\varphi_j - \hat{\Omega})}} |\Phi_0\rangle \\
 &= \sum_k \frac{1 - e^{iM\tau(\varphi_j - \omega_k)}}{1 - e^{i\tau(\varphi_j - \omega_k)}} (\Upsilon_k | \Phi_0) |\Upsilon_k\rangle \\
 &= \sum_k \frac{1 - e^{iM\tau(\varphi_j - \omega_k)}}{1 - e^{i\tau(\varphi_j - \omega_k)}} d_k^{1/2} |\Upsilon_k\rangle \\
 &= \sum_k f(\omega_k - \varphi_j) d_k^{1/2} |\Upsilon_k\rangle,
 \end{aligned} \tag{2.60}$$

com

$$f(\omega_k - \varphi_j) = \frac{1 - e^{-iM\tau(\omega_k - \varphi_j)}}{1 - e^{-i\tau(\omega_k - \varphi_j)}}. \tag{2.61}$$

Logo, $|\Psi_j\rangle$ é uma certa combinação linear das funções da base do Hamiltoniano, $|\Upsilon_k\rangle$. Isto leva a crer que, se J é suficientemente grande, esta transformação possui inversa:

$$|\Upsilon_k\rangle = \sum_{j=0}^{J-1} [\tilde{B}_k]_j |\Psi_j\rangle, \quad k = 0, 1, 2 \dots M-1. \tag{2.62}$$

A função $f(\omega_k - \varphi_j)$ possui uma estrutura do tipo “*sinc*” centrada em $\omega_k - \varphi_j = 0$, cuja amplitude decai conforme $\omega_k - \varphi_j$ se afasta de zero. A largura desta distribuição é da ordem do intervalo médio entre as frequências φ_j , isto é, $\approx 2\pi/(J\tau)$, e diminui com o aumento de J . Portanto, somente aquelas frequências ω_k que satisfazem $\omega_k \approx \varphi_j$ contribuem para a soma:

$$|\Psi_j\rangle \approx \sum_{\omega_k \approx \varphi_j} f(\omega_k - \varphi_j) d_k^{1/2} |\Upsilon_k\rangle. \quad (2.63)$$

A relação inversa fica:

$$|\Upsilon_k\rangle \approx \sum_{\varphi_j \approx \omega_k} [\tilde{B}_k]_j |\Psi_j\rangle. \quad (2.64)$$

Isto demonstra a afirmação anterior. Agora, utilizando-se a Equação 2.56, vamos considerar os elementos de matriz, $[\tilde{U}^p]_{jk}$, com $j, k = 0, 1, 2, \dots, J-1$, do operador $\hat{U}^p$ na base $\{|\Psi_j\rangle\}$:

$$\begin{aligned} [\tilde{U}^p]_{jk} &= \tilde{U}_{jk}^p = (\Psi_j | \hat{U}^p | \Psi_k) = \sum_{n=0}^{M-1} \sum_{m=0}^{M-1} e^{in\tau\varphi_j} e^{im\tau\varphi_k} (\Phi_n | \hat{U}^p | \Phi_m) \\ &= \sum_{n=0}^{M-1} \sum_{m=0}^{M-1} e^{in\tau\varphi_j} e^{im\tau\varphi_k} c_{n+m+p}. \end{aligned} \quad (2.65)$$

Considere-se $p = 0$ e os elementos da diagonal secundária de $\tilde{U}^0$, dados por $\tilde{u}_j^0 \equiv \tilde{U}_{j, J-j-1}^0$, com $j = 0, 1, \dots, J-1$, tal que $\tilde{u}_0^0$ é o elemento do canto superior direito de $\tilde{U}^0$ e $\tilde{u}_{J-1}^0$ é o elemento do canto inferior esquerdo de $\tilde{U}^0$. Então, a partir de 2.65 podemos escrever para $p = 0$:

$$\begin{aligned} \tilde{u}_j^0 &= (\Phi_0 | \sum_{n=0}^{M-1} \sum_{m=0}^{M-1} e^{in\tau\varphi_j} e^{im\tau\varphi_{J-j-1}} \hat{U}^{n+m} | \Phi_0) \\ &= (\Phi_0 | \sum_{n=0}^{M-1} \sum_{m=0}^{M-1} e^{in\tau\varphi_j} e^{im\tau\varphi_{J-j-1}} e^{-i(n+m)\tau\hat{\Omega}} | \Phi_0) \\ &= (\Phi_0 | \left\{ \sum_{n=0}^{M-1} \left[e^{-i\tau(\hat{\Omega}-\varphi_j)} \right]^n \right\} \left\{ \sum_{m=0}^{M-1} \left[e^{-i\tau(\hat{\Omega}-\varphi_{J-j-1})} \right]^m \right\} | \Phi_0) \\ &= (\Phi_0 | \left[\frac{1 - e^{-iM\tau(\hat{\Omega}-\varphi_j)}}{1 - e^{-i\tau(\hat{\Omega}-\varphi_j)}} \right] \left[\frac{1 - e^{-iM\tau(\hat{\Omega}-\varphi_{J-j-1})}}{1 - e^{-i\tau(\hat{\Omega}-\varphi_{J-j-1})}} \right] | \Phi_0) \\ &= (\Phi_0 | \sum_i f(\omega_k - \varphi_j) |\Upsilon_i\rangle \langle \Upsilon_i| \sum_k f(\omega_k - \varphi_{J-j-1}) |\Upsilon_k\rangle \langle \Upsilon_k| | \Phi_0) \end{aligned}$$

$$\begin{aligned}
&= \sum_k f(\omega_k - \varphi_j) f(\omega_k - \varphi_{J-j-1}) (\Upsilon_k | \Phi_0)^2 \\
&= \sum_k f(\omega_k - \varphi_j) f(\omega_k - \varphi_{J-j-1}) d_k.
\end{aligned} \tag{2.66}$$

Assim como ressaltamos anteriormente, $f(\omega_k - \varphi_j)$ é do tipo *sinc*, centrada em $\omega_k - \varphi_j = 0$, enquanto que $f(\omega_k - \varphi_{J-j-1})$ é também do tipo *sinc*, centrada em $\omega_k - \varphi_{J-j-1} = 0$. Portanto, para distribuições localizadas, $|\varphi_j - \varphi_{J-j-1}| \gg 2\pi/(J\tau)$, o produto entre as duas funções será sempre aproximadamente nulo. Conclui-se então que $\tilde{u}_j^0$ poderá ser diferente de zero somente quando $\omega_k \approx \varphi_j \approx \varphi_{J-j-1}$, ou seja quando $j \approx (J-1)/2$.

Resultados análogos podem ser obtidos para as outras diagonais secundárias de $\tilde{U}^0$, assim como para as de $\tilde{U}^p$, para qualquer p . Isto significa que somente os elementos $\tilde{U}_{jk}^p$ localizados próximo da diagonal principal ($j \approx k$) podem ser significativamente diferentes de zero. Logo, **as matrizes $\tilde{U}^p$ são quase-diagonais**.

Antes de tirar proveito desta propriedade das matrizes $\tilde{U}^p$, vamos delinear a seguir o procedimento que se utiliza para resolver o HIP na base tipo Fourier.

Substituindo-se 2.62 em 2.10 e multiplicando ambos os lados por $|\Psi_m\rangle$ à esquerda, obtemos uma equação de autovalores generalizada para os autovalores μ_k e autovetores $\tilde{B}_k$:

$$\begin{aligned}
\sum_{n=0}^{J-1} (\Psi_m | \hat{U}^p | \Psi_n) [\tilde{B}_k]_n &= \mu_k^p \sum_{n=0}^{J-1} (\Psi_m | \Psi_n) [\tilde{B}_k]_n \\
\sum_{n=0}^{J-1} \tilde{U}_{mn}^p [\tilde{B}_k]_n &= \mu_k^p \sum_{n=0}^{J-1} \tilde{U}_{mn}^0 [\tilde{B}_k]_n,
\end{aligned} \tag{2.67}$$

ou seja,

$$\tilde{U}^p \tilde{B}_k = \mu_k^p \tilde{U}^0 \tilde{B}_k, \tag{2.68}$$

com as equações decorrentes:

$$\begin{aligned}\tilde{U}^1 \tilde{B}_k &= \mu_k \tilde{U}^0 \tilde{B}_k \\ \tilde{U}^2 \tilde{B}_k &= \mu_k^2 \tilde{U}^0 \tilde{B}_k \\ \tilde{U}^p \tilde{B}_k &= \mu_k \tilde{U}^{p-1} \tilde{B}_k.\end{aligned}\tag{2.69}$$

Substituindo-se a Equação 2.62 na Equação 2.7 conclui-se que os autovetores $\tilde{B}_k$ devem ser normalizados com relação a $\tilde{U}^0$:

$$\tilde{B}_k^T \tilde{U}^0 \tilde{B}_{k'} = \delta_{kk'}.\tag{2.70}$$

As frequências, ω_k , são calculadas a partir dos autovalores, μ_k , e, a partir dos autovetores normalizados, $\tilde{B}_k$, calcula-se as amplitudes, d_k , através das Equações 2.9 e 2.62:

$$\sqrt{d_k} = (\Phi_0 | \Upsilon_k) = \sum_{n=0}^{J-1} [\tilde{B}_k]_n (\Phi_0 | \Psi_n) = \sum_{n=0}^{J-1} [\tilde{B}_k]_n \tilde{C}_n = \tilde{B}_k^T \tilde{C},\tag{2.71}$$

onde, $\tilde{B}_k^T$ é um vetor linha $(1, J)$, cujos elementos são os da k -ésima coluna da matriz $\tilde{B}$ e $\tilde{C} = (c_0, c_1, \dots, c_{J-1})^T$ é um vetor coluna $(J, 1)$, já definido em 2.58.

Agora, vamos tirar proveito do fato de que as matrizes $\tilde{U}^p$, de dimensão (J, J) , são quase-diagonais. O resultado fundamental desta propriedade é que as matrizes $\tilde{U}^p$ podem ser diagonalizadas em blocos de dimensões $L < J$, simplesmente ignorando-se as correlações entre pontos φ_j e φ_k que estão muito separados entre si, conforme esquematizado na Figura 2.8.

Em outras palavras, qualquer pequeno bloco diagonal da matriz $\tilde{U}^p$, formado pelas frequências $\varphi_j, \varphi_{j+1}, \dots, \varphi_{j+L-1}$, com $1 \ll L < J$, pode ser diagonalizado isoladamente e, as frequências ω_k , com $k = 0, 1, \dots, L-1$, das

Figura 2.8: Divisão das matrizes $\tilde{U}^p$ de dimensão J em pequenos blocos de dimensão L . Elementos muito fora da diagonal principal são quase-nulos. Há uma superposição de 50% entre cada dois blocos adjacentes.

componentes espectrais calculadas para este bloco a partir dos autovalores da equação generalizada (Equação 2.68), serão as corretas frequências do HIP para $\omega_k \in (\varphi_j, \varphi_{j+L-1})$.

Denominaremos, daqui em diante, o intervalo $(\varphi_j, \varphi_{j+L-1})$ por “**janela espectral j de dimensão L** ”. A Figura 2.9 ilustra o fato de que, não obstante a complexidade do espectro associado ao sinal de interesse, a análise espectral de apenas um segmento de todo o espectro pode ser processada isoladamente.

A Figura 2.10 mostra um exemplo do procedimento utilizado para combinar-se os resultados de diversas janelas sobrepostas em 50%. Para cada janela, denotada pelo índice w , somente as frequências calculadas ω_k^w (com correspondentes amplitudes d_k^w) no intervalo $(\varphi_a^w, \varphi_b^w)$ são mantidas e uti-

Figura 2.9: Um exemplo de uma janela espectral utilizada nos cálculos do FDM.

lizadas para construir o espectro $S_w(\varphi)$ de acordo com a Equação 2.30 (ou 2.31). Somando-se os espectros correspondentes a cada janela, obtém-se o espectro total:

$$S(\varphi) = \sum_w g_w(\varphi) S_w(\varphi) \quad \text{com} \quad \sum_w g_w(\varphi) = 1, \quad (2.72)$$

onde a função peso, $g_w(\varphi)$, é diferente de zero somente dentro de cada janela e é utilizada para rejeitar-se os picos com frequência próxima da borda de cada janela, onde o cálculo é menos preciso.

Normalmente utiliza-se uma mesma função peso para todas as janelas escolhidas em um mesmo procedimento, todavia, dois tipos de funções peso tem sido mais comumente utilizadas. Primeiro, a mais simples consiste em fazer $g_w(\varphi) = 0$ para valores φ localizados nos extremos de cada janela e $g_w(\varphi) = 1$ para a porção central (50% do intervalo total). Em outras palavras, simplesmente despreza-se os autovalores correspondentes a frequên-

Figura 2.10: Um exemplo de implementação do FDM multi-janelas.

cias próximas da borda de cada janela. Uma outra maneira, esquematizada na Figura 2.10, consiste em considerar:

$$g_w(\varphi) = \frac{1}{2} \left[1 - \cos \left(2\pi \frac{\varphi - \varphi_a^w}{\varphi_b^w - \varphi_a^w} \right) \right]. \quad (2.73)$$

Agora, vamos discutir com um pouco mais de detalhes a implementação prática do FDM.

O cálculo das matrizes $\tilde{U}^p$ através da soma dupla da Equação 2.65 requer muito esforço computacional e deve ser reduzida a somas simples através de

mudanças de variáveis [22, 20, 8]. O resultado final se obtém por meio do conjunto de fórmulas escrito abaixo:

$$\begin{aligned}
\tilde{U}_{jk}^p &= \frac{1}{z_j - z_k} \left\{ z_j \left[G_k^p - z_j^{-M} G_k^{M+p} \right] - z_k \left[G_j^p - z_k^{-M} G_j^{M+p} \right] \right\} \\
\tilde{U}_j^p &= \sum_{n=0}^{M-1} \left[(n+1) z_j^{-n} c_{n+p} + (M-n-1) z_j^{-M} z_j^{-n} c_{n+M+p} \right] \\
G_j^\alpha &= \sum_{n=0}^{M-1} z_j^{-n} c_{n+\alpha} \\
z_j &= e^{-i\tau\varphi_j}.
\end{aligned} \tag{2.74}$$

$\tilde{U}_j^p \equiv \tilde{U}_{jj}^p$ são os elementos da diagonal de $\tilde{U}^p$, e $\tilde{U}_{jk}^p$ são os elementos fora da diagonal de $\tilde{U}^p$.

Também muito útil são as relações de recorrência:

$$\begin{aligned}
\tilde{U}_{jk}^{p+1} &= \frac{z_j z_k}{z_j - z_k} \left[(G_k^p - z_j^{-M} G_k^{M+p}) - (G_j^p - z_k^{-M} G_j^{M+p}) \right] \\
\tilde{U}_j^{p+1} &= z_j \left[\tilde{U}_j^p - (G_j^p - z_j^{-M} G_j^{M+p}) \right] \\
G_j^{\alpha+1} &= z_j \left(G_j^\alpha + z_j^{-M} c_{M+\alpha} - c_\alpha \right).
\end{aligned} \tag{2.75}$$

2.7.3 Sumário do algoritmo do FDM

A partir do conjunto de equações elaborado na seção anterior, pode-se estabelecer uma receita de cálculo baseada na seqüência de passos descrita a seguir.

Passo 1

O primeiro passo consiste na configuração inicial do cálculo. Primeiro, o sinal de entrada estabelece os valores de N e de τ . Segundo, assumiremos

Figura 2.11: Distribuição das janelas com 50% de superposição para $W = 7$. Os retângulos negros representam as regiões ativas das janelas.

que todas as janelas são definidas com os mesmos parâmetros, L , ρ e $\Delta\varphi$ (embora o início de cada janela, φ_a^w , e o final da mesma, φ_b^w , dependam de w , a diferença $\Delta\varphi \equiv \varphi_b^w - \varphi_a^w$ não depende de w porque, embora não seja uma condição necessária, estamos admitindo que todas as janelas possuem a mesma largura). Estes três parâmetros que caracterizam as janelas não são independentes entre si, de fato, pela Equação 2.59 eles devem satisfazer: $2\pi L/(\rho\Delta\varphi) = N\tau$.

A primeira decisão a ser tomada é a escolha do intervalo de frequência (φ_a, φ_b) , com $\varphi_b > \varphi_a$ e $-\pi/\tau \leq \varphi_a, \varphi_b \leq \pi/\tau$, no qual se deseja realizar a análise espectral do sinal de entrada.

A segunda decisão a ser tomada é a escolha da dimensão da grade, $L \gg 1$, associada a cada janela $w = 0, 1, 2, \dots, W-1$ no intervalo $(\varphi_a^w, \varphi_b^w)$, com $\varphi_b^w > \varphi_a^w$. Uma escolha apropriada seria no intervalo $L \approx 10 - 100$. A especificação destes três parâmetros define: $\rho\Delta\varphi = 2\pi L/(N\tau)$.

Em seguida, escolhe-se um valor para a densidade de pontos da grade, ρ . Um valor apropriado é $\rho \approx 0.5 - 0.6$ mas, em geral, existe um valor mínimo para ρ que é estabelecido pelo fato de que a largura da janela tem que, necessariamente, ser menor ou igual a $\varphi_b - \varphi_a$. Portanto, a seguinte condição tem que ser satisfeita: $\rho \geq 2\pi L/[N\tau(\varphi_b - \varphi_a)]$. Num caso hipotético em que $N = 125$, $L = 100$ e $(\varphi_b - \varphi_a) = 2\pi/\tau$ esta condição implicaria $\rho \geq 0.8$. Feito isso, estabelece-se a largura das janelas: $\Delta\varphi = 2\pi L/(N\tau\rho)$.

Não é aconselhável aumentar-se muito a densidade, pois a diminuição do intervalo entre os valores da grade pode provocar erros numéricos, por isso, é mais conveniente diminuir-se L quando N for pequeno. Todavia, se L ficar próximo da unidade, é mais conveniente realizar-se o cálculo do FDM em uma única janela ($W = 1$), escolhendo-se $\Delta\varphi = \varphi_b - \varphi_a$ e $\rho = 2\pi L/[N\tau(\varphi_b - \varphi_a)]$ com $L \gg 1$.

A fim de que a análise espectral possa ser realizada em todo o intervalo (φ_a, φ_b) , através do conjunto de W janelas nos intervalos $(\varphi_a^w, \varphi_b^w)$ é necessário um esquema prático para o posicionamento de tais janelas. Um esquema prático está mostrado na Figura 2.11 e, este esquema em particular, requer que o intervalo total de frequências, $\varphi_b - \varphi_a$, seja múltiplo inteiro de $\Delta\varphi/2$. Para isso é necessário reajustar-se o $\Delta\varphi$ calculado para o valor mais próximo da série $\Delta\varphi_n = 2(\varphi_b - \varphi_a)/n$, que resulta em uma das possíveis densidades $\rho_n = \pi n L/[N\tau(\varphi_b - \varphi_a)]$, com $n = 2, 3, \dots$. Por último, calcula-se os limites de cada janela, $\varphi_a^w = \varphi_a + w\Delta\varphi/2$ e $\varphi_b^w = \varphi_a^w + \Delta\varphi$, com $w = 0, 1, 2 \dots W - 1$; e o número total de janelas para as quais o FDM será aplicado, $W = n - 1$.

Passo 2

Toma-se $w = 0$ e constrói-se uma grade de valores φ_j , $j = 0, 1, 2, \dots, L - 1$ (igualmente espaçada é a sugestão inicial) tal que $\varphi_a^w \leq \varphi_j < \varphi_b^w$:

$$\varphi_j = \varphi_a^w + \frac{\varphi_b^w - \varphi_a^w}{L} j \quad j = 0, 1, \dots, L-1. \quad (2.76)$$

Calcula-se os vetores:

$$z_j = e^{-i\tau\varphi_j} \quad \text{e} \quad z_j^{-M}. \quad (2.77)$$

para $j = 0, 1, 2, \dots, L-1$, onde, conforme já definido no início desta seção, $M = N/2$ (ou, $M = (N-1)/2$ se o cálculo de erros estiver habilitado).

Passo 3

Calcula-se as somas:

$$\begin{aligned} G_j^0 &= \sum_{n=0}^{M-1} z_j^{-n} c_n \\ G_j^M &= \sum_{n=0}^{M-1} z_j^{-n} c_{n+M} \\ \tilde{U}_j^0 &= \sum_{n=0}^{M-1} [(n+1)z_j^{-n} c_n + (M-n-1)z_j^{-M} z_j^{-n} c_{n+M}]. \end{aligned} \quad (2.78)$$

Passo 4

Calcula-se o vetor G^1 :

$$G_j^1 = z_j (G_j^0 + z_j^{-M} c_M - c_0). \quad (2.79)$$

Passo 5

Se o cálculo de erros está habilitado, calcula-se também o vetor G^{M+1} :

$$G_j^{M+1} = z_j (G_j^M + z_j^{-M} c_{2M} - c_M). \quad (2.80)$$

Passo 6

Calcula-se os elementos fora da diagonal de $\tilde{U}^0$ e a matriz $\tilde{U}^1$:

$$\begin{aligned}\tilde{U}_{jk}^0 &= \frac{1}{z_j - z_k} \{ z_j [G_k^0 - z_j^{-M} G_k^M] - z_k [G_j^0 - z_k^{-M} G_j^M] \} \\ \tilde{U}_j^1 &= z_j [\tilde{U}_j^0 - (G_j^0 - z_j^{-M} G_j^M)] \\ \tilde{U}_{jk}^1 &= \frac{z_j z_k}{z_j - z_k} [(G_k^0 - z_j^{-M} G_k^M) - (G_j^0 - z_k^{-M} G_j^M)].\end{aligned}\quad (2.81)$$

Passo 7

Se o cálculo de erros está habilitado, calcula-se a matriz $\tilde{U}^2$:

$$\begin{aligned}\tilde{U}_j^2 &= z_j [\tilde{U}_j^1 - (G_j^1 - z_j^{-M} G_j^{M+1})] \\ \tilde{U}_{jk}^2 &= \frac{z_j z_k}{z_j - z_k} [(G_k^1 - z_j^{-M} G_k^{M+1}) - (G_j^1 - z_k^{-M} G_j^{M+1})].\end{aligned}\quad (2.82)$$

Passo 8

Resolve-se a equação de autovalores generalizada:

$$\tilde{U}^1 \tilde{B}_k = \mu_k \tilde{U}^0 \tilde{B}_k, \quad (2.83)$$

calculando-se os autovalores, μ_k , e autovetores, $\tilde{B}_k$, com $k = 0, 1, 2, \dots, L-1$, por qualquer um dos três métodos discutidos anteriormente: QZ, EVD ou SVD.

Passo 9

Se o cálculo de erros está habilitado, calcula-se os erros:

$$\varepsilon_k = \sqrt{\tilde{B}_k^T \tilde{U}^2 \tilde{B}_k - \mu_k^2}, \quad (2.84)$$

aceitando-se apenas aqueles autovalores cujo erro associado é menor que um certo limite máximo.

Passo 10

Calcula-se as frequências, ω_k , através dos autovalores:

$$\mu_k = e^{-i\tau\omega_k}. \quad (2.85)$$

Normaliza-se os autovetores:

$$\tilde{B}_k^T \tilde{U}^0 \tilde{B}_k = 1. \quad (2.86)$$

Calcula-se as amplitudes:

$$\sqrt{d_k} = \tilde{B}_k^T \tilde{C}, \quad \text{onde} \quad \tilde{C}_j = \sum_{n=0}^{M-1} z_j^{-n} c_n, \quad (2.87)$$

normalizando-as em seguida pela função peso escolhida. Se optarmos pelo método esquematizado na Figura 2.11, despreza-se os autovalores cujas frequências caem fora da parte central da janela (com excessão das primeira e última janelas, que possuem 3/4 ativos).

Passo 11

Este passo é opcional e consiste na implementação de filtros para eliminar-se autovalores espúrios. Por exemplo, filtros baseados nos seguintes critérios:

- (a) Eliminar-se componentes espectrais com amplitudes ($|d_k|$) menores que um certo valor.
- (b) Eliminar-se picos do espectro com $\gamma_k < 0$.

- (c) Eliminar-se picos cujas amplitudes ($|d_k|/\gamma_k$) são menores que um certo valor.
- (d) Eliminar-se picos cuja meia largura (γ_k) é menor que um certo valor.
- (e) Eliminar-se picos cuja meia largura (γ_k) é maior que um certo valor.

Passo 12

Retorna-se ao passo 2, escolhendo-se uma nova janela de frequências ($w = 1$), mantendo os mesmos valores de L , ρ e $\Delta\varphi$, e assim sucessivamente para todas as janelas.

2.8 Exemplos ilustrativos do FDM

2.8.1 Exemplo 1: Resolução

Neste exemplo testaremos a resolução do FDM, isto é, a sua capacidade de resolver linhas lorentzianas próximas mesmo quando o sinal de entrada é severamente truncado. Para que possamos apreciar o exemplo vamos mostrá-lo comparativamente com a DFT.

Tabela 2.4: Parâmetros do sinal temporal mostrado nas Figuras 2.12 e 2.13.

k	f_k	$1/\gamma_k$	$ d_k $	$\theta_k(\text{rad})$
1	10.00	40	2	0
2	10.05	40	4	0
3	10.10	40	8	0
4	10.15	40	8	0
5	10.20	40	4	0
6	10.25	40	2	0

Figura 2.12: Transformadas de Fourier discretas, DFT, do sinal simulado de acordo com os parâmetros da Tabela 2.4 tomando-se o intervalo de amostragem $\tau = 0.02$ (pouco abaixo do limite de Nyquist), para diferentes quantidade de amostragens, N . A coluna à esquerda mostra os sinais a serem transformados (parte real) e a coluna à direita as respectivas transformadas (parte real).

O exemplo consiste de um espectro formado por um sexteto centrado em $f = 10.125$, e os seus parâmetros estão especificados na Tabela 2.4. Para simplificar a visualização e evitar a necessidade de corrigir a fase dos espectros, o tempo morto é tomado igual a zero e todas as seis linhas possuem a mesma constante de decaimento e fase nula; suas amplitudes ocorrem em uma relação simples 1:2:4:4:2:1.

Figura 2.13: Análise espectral feita pelo FDM do sinal simulado de acordo com os parâmetros da Tabela 2.4 tomando-se o intervalo de amostragem $\tau = 0.02$ (pouco abaixo do limite de Nyquist), para diferentes quantidade de amostragens, N . A coluna à esquerda mostra os sinais a serem analisados (parte real) e a coluna à direita os espectros do FDM (parte real).

Primeiramente, mostraremos como a transformada de Fourier discreta deste sinal temporal se comporta com relação ao truncamento do sinal. O resultado deste teste está mostrado na Figura 2.12.

A Figura 2.12 mostra que o espectro de DFT obtido para $N = 32768$ possui resolução suficiente para exibir fidelidade aos parâmetros utilizados na simulação. Para $N = 16384$ nota-se um pequeno desvio nas amplitudes, pois os dois picos centrais não apresentam a mesma amplitude. Para menores

valores de N os resultados obtidos são de baixa resolução, conforme pode-se observar. Para $N = 2048$ nota-se que cada pico é formado por apenas dois pontos e que a amplitude torna-se negativa em algumas regiões, o que é resultado do truncamento.

Estes resultados estão em acordo com o esperado pois, pela Equação 2.54 estima-se que para resolver-se duas linhas separadas por $\Delta f = 0.05$ necessita-se uma quantidade mínima de pontos $N = 1/(\tau \Delta f) = 1000$, lembrando que $\tau = 0.02$. Nesta situação limite, cada linha ficaria representada espectralmente por apenas um ponto.

A intenção agora é apreciar os resultados mostrados na Figura 2.13 que representam o mesmo teste através do FDM.

O teste com o FDM, mostra que para $N = 256$ o resultado de sua análise espectral é fiel aos parâmetros do sinal. Comparando-se com o teste anterior, vemos que o FDM é capaz de obter, neste caso, um espectro de alta resolução com uma quantidade de pontos 64 vezes menor do que aquela necessária para realizar a mesma tarefa através da DFT. Quando o número de pontos é reduzido, abaixo de 256, perde-se resolução mas não se nota deformações excessivas: o espectro é sempre suave e as linhas (em número reduzido) são bem delineadas. Esta é uma característica do FDM pois, uma vez determinada a lista de parâmetros, o espectro pode ser calculado com uma quantidade arbitrária de pontos.

Com relação a esta última observação, poder-se-ia argumentar que a DFT também permite que a quantidade de pontos no espectro seja aumentada sem prejuízo da qualidade do mesmo. Isto diz respeito ao processo pelo qual se estende a dimensão do sinal através de “preenchimento por zeros”, processo este chamado em inglês por *zero-filling*. A diferença é que o preenchimento por zeros não aumenta a resolução do espectro: ele apenas adiciona pontos

interpolados, de acordo com uma regra que nem sempre é satisfatória do ponto de vista dos processos de interpolação numérica.

Este exemplo mostra que, neste caso, o FDM foi capaz de tirar proveito da alta relação sinal/ruído do sinal de entrada, transferindo esta vantagem a um significativo aumento de resolução.

2.8.2 Exemplo 2: Alta densidade de picos

Para demonstrar a habilidade do FDM em identificar e resolver uma grande quantidade de linhas Lorentzianas em uma estreita faixa de frequências, utilizaremos aqui um modelo chamado “escada de Jacó” (ou, *Jacob’s ladder*), adaptado de [6]. O sinal contém 40 tripletos com relação de intensidades 1:2:1, cujas larguras e separações diminuem progressivamente de altas para baixas frequências:

$$c_n = \sum_{k=0}^{39} \left(e^{-2\pi i (0.02n) (19.5-1.0i)0.85^k} + e^{-2\pi i (0.02n) (20.0-1.0i)0.85^k} + e^{-2\pi i (0.02n) (20.5-1.0i)0.85^k} \right) \quad (2.88)$$

A vasta faixa de larguras e separações dos picos, e acúmulo de picos em baixa frequências tornam este problema difícil de ser resolvido através de técnicas paramétricas não relacionadas com a transformada de Fourier.

A Figura 2.14 mostra os resultados obtidos deixando claro que o FDM foi capaz de resolver todas as linhas, inclusive as de mais baixa frequências a partir de um sinal com 32768 pontos. Observa-se em (c) que um espectro de qualidade equivalente não pode ser obtido através da DFT. A separação do tripleto de mais baixa frequência é ≈ 0.000884 enquanto que a resolução estimada para a DFT é de $\delta f = 1/(N\tau) \approx 0.0015$ e que portanto a DFT não possui resolução suficiente quando $N = 32768$.

Figura 2.14: “Escada de Jacó” calculada de acordo com a Equação 2.88, tendo um total de 40 tripletos, sendo o primeiro ($k = 0$) centrado em 20.0, com $\gamma = 1.0$ e separação 0.5, e o último ($k = 39$) centrado em 0.0353, com $\gamma = 0.00177$ e separação 0.000884. O sinal temporal foi calculado utilizando-se $N = 32768$, $\tau = 0.02$ e $t_d = 0$. O segmento de baixa frequência do espectro de absorção da DFT, está mostrado em (c). O espectro de absorção foi calculado por FDM-SVD (utilizou-se $L = 40$, $\rho \approx 0.59$, $W = 959$ e $\Delta\varphi \approx 0.1$) e ele converge corretamente, como está mostrado em (b,d). Para permitir a visualização dos picos de alta frequência, a escala vertical de (b) está truncada. Em (d) é mostrado o segmento de baixas frequências do espectro de FDM.

2.8.3 Exemplo 3: FDM vs. DFT

Já demonstramos que o FDM, utilizando-se da informação prévia de que o sinal contém apenas linhas Lorentzianas, é capaz de fornecer uma resolução além daquela prevista pelo princípio de incerteza da DFT. Entretanto, isto não significa que o FDM fornece uma maior resolução para **todos** os sinais Lorentzianos.

Figura 2.15: Um caso especial em que a DFT funciona melhor que o FDM, apesar do sinal ser composto de apenas linhas Lorentzianas. Esta figura foi copiada de [6].

Na prática, observa-se que é mais vantajoso para o FDM que os picos estejam não uniformemente distribuídos do que uniformemente distribuídos. Para demonstrar este ponto, mostramos na Figura 2.15 uma simulação numérica de um sinal que possui 21 picos igualmente espaçados na faixa

central do intervalo de Nyquist, conforme mostrado na Figura 2.15.

Este exemplo foi copiado de [6] e mostra que, embora o sinal contém apenas linhas Lorentzianas, ele é ótimo para o processamento da DFT, porque a DFT fornece resolução uniforme e então funciona melhor quando os picos estão uniformemente distribuídos.

A Figura 2.15 compara o desempenho da DFT e do FDM aplicados a este sinal. Ela mostra que a DFT é capaz de resolver todos os 21 picos utilizando-se apenas de $N = 60$ pontos, enquanto que o FDM somente converge para $N = 100$. Entretanto, observe que, embora a DFT seja capaz de resolver todas as 21 linhas para $N = 60$, as larguras convergem muito lentamente, enquanto que o FDM é capaz de recuperar todos os parâmetros com precisão a partir de $N = 100$.

2.8.4 Exemplo 4: Ajuste multi-exponencial

Muitos fenômenos físicos são descritos por funções exponenciais, quase sempre, decaimentos exponenciais. Por isso, o problema da decomposição de um transiente em uma soma de exponenciais, chamado de “análise exponencial”, ocorre freqüentemente na Física.

Em termos da transformada de Laplace, uma função $c(t)$ pode ser descrita por uma soma infinita de exponenciais:

$$c(t) = \int_0^\infty g(\gamma) e^{-\gamma t} d\gamma. \quad (2.89)$$

Portanto, o problema da análise exponencial de um transiente, $c(t)$, que significa apenas a determinação de $g(\gamma)$, pode ser resolvido, em princípio, através da transformada inversa de Laplace de $c(t)$.

Usualmente, a transformada inversa de Laplace pode ser calculada facilmente quando a expressão analítica de $c(t)$ é conhecida. Entretanto, isto

raramente acontece na Física Experimental. Na maioria dos casos, a transformada inversa, também conhecida como integral de Bromwich, não pode ser aplicada diretamente tomando-se os dados experimentais (números reais) como argumento da integral discreta, assim como é de costume fazer-se com a transformada de Fourier.

Isto ocorre porque, diferentemente das funções harmônicas, as exponenciais não são ortogonais no eixo real, isto é, a contribuição de cada exponencial ao sinal $c(t)$ não pode ser projetada tomando-se o produto interno definido como uma integral ao longo do eixo real. Isto se reflete no fato de que a transformada de Bromwich é definida como uma integral no plano complexo.

Além disso, quando se trata de observações experimentais, a função sinal, $c(t)$, é amostrada somente ao longo do eixo real. Em vista destes fatos, a função distribuição de Laplace, $g(\gamma)$, ou seja, a transformada inversa de $c(t)$, só pode ser encontrada através da própria definição da transformada direta, Equação 2.89. O problema é que esta equação integral é conhecida por ser “mal posta” (ou como se diz em inglês, *improperly posed*). Este termo significa que a solução, $g(\gamma)$, da Equação 2.89 pode não ser única, pode não existir, e pode não depender continuamente dos dados.

Existe na literatura diversos exemplos que ilustram este problema. O mais famoso foi feito por Lanczos [18, 15], que mostrou que a soma de duas exponenciais pode ser reproduzida, dentro de duas casas decimais, por uma soma de três exponenciais com constantes de tempo e amplitudes completamente diferentes das anteriores. Além disso, a partir do cálculo dos desvios quadráticos mínimos em cada caso, não se pode concluir qual dos dois ajustes é o melhor!

Não vamos entrar em detalhes a respeito deste assunto porque ele foge muito do nosso interesse mas, podemos citar um excelente trabalho de revisão

publicado recentemente que discute com abundância de detalhes os diversos métodos existentes atualmente para a análise exponencial [15].

Tabela 2.5: Parâmetros do sinal temporal mostrado nas Figuras 2.16 e 2.17.

k	f_k	$1/\gamma_k$	$ d_k $	$\theta_k(\text{rad})$
1	0	1	10	0
2	0	2	10	π
3	0	3	10	$\pi/2$
4	0	4	10	$\pi/3$
5	0	5	10	$\pi/4$
6	0	6	10	$\pi/5$
7	0	7	10	$\pi/6$
8	0	8	10	$\pi/7$
9	0	9	10	$\pi/8$
10	0	10	10	$\pi/9$

Tendo em vista que a Equação 2.89 discretizada é um caso particular do HIP (ver a Equação 2.1, fazendo-se $f_k = 0$) vamos a seguir realizar alguns testes, afim de adquirirmos uma pequena intuição do que é que o FDM seria capaz de fazer com relação a este difícil problema.

Consideraremos primeiramente a tabela de parâmetros, Tabela 2.5. Ela representa uma soma de 10 exponenciais, com amplitudes iguais, constantes de tempo igualmente espaçadas e com fases que escolhemos arbitrariamente.

O valor do intervalo de amostragem, τ , foi escolhido de tal forma que ele fosse bem menor que a menor constante de tempo do sinal simulado; e o número de pontos, N , foi escolhido tal que $N\tau$ fosse bem maior que a constante de tempo mais longa.

A fim de enfrentarmos as dificuldades de uma forma gradativa, vamos

Figura 2.16: Análise exponencial feita pelo FDM do sinal simulado de acordo com os primeiros 7 conjuntos de parâmetros da Tabela 2.5, tomando-se o intervalo de amostragem $\tau = 0.05$ e $N = 700$. À esquerda mostra-se o sinal a ser analisado, assim como o módulo do resíduo do ajuste, e na coluna à direita os valores das amplitudes correspondentes a cada um dos 7 valores de $1/\gamma$ encontrados.

primeiramente analisar os resultados do FDM quando se simula um sinal com apenas os primeiros 7 conjuntos de parâmetros da Tabela 2.5. O sinal simulado e o resultado do FDM estão mostrados na Figura 2.16.

Estes resultados mostram que o FDM foi capaz de identificar que o sinal simulado possui 7 componentes, o que é correto (lembrem-se que ninguém lhe deu esta informação) e os erros nos valores encontrados para as amplitudes estão abaixo de 1%. A Figura 2.16(b) não mostra os valores exatos das constantes de tempo e fases, mas podemos dizer ambos foram calculados com razoável precisão.

O resíduo mostrado na Figura 2.16(a) é a diferença entre o sinal de entrada

Figura 2.17: Análise exponencial feita pelo FDM do sinal simulado de acordo com os parâmetros da Tabela 2.5, tomando-se o intervalo de amostragem $\tau = 0.05$ e $N = 1000$. À esquerda mostra-se o sinal a ser analisado, assim como o módulo do resíduo do ajuste, e na coluna à direita os valores das amplitudes correspondentes a cada um dos 10 valores de $1/\gamma$ encontrados.

e o sinal simulado a partir da lista $\{d_k, \omega_k\}$ calculada pelo FDM. Nota-se que o resíduo é cerca de 5 ordens de grandeza menor que o sinal, o que significa que o ajuste foi razoável.

O sinal simulado com todas as 10 componentes da Tabela 2.5 está mostrado na Figura 2.17 juntamente com o resultado do FDM.

Observa-se neste caso um desvio muito maior das amplitudes encontradas com relação aos verdadeiros valores simulados. Não obstante, o FDM foi novamente capaz de distinguir a necessidade de utilizar-se de 10 componentes para simular este sinal! A relação entre o sinal e o resíduo aumentou com relação ao exemplo anterior, sendo agora cerca de 3 ordens de grandeza mas, o ajuste continua razoável.

Acreditamos que a dificuldade encontrada pelo algoritmo numérico foi o fato de que estas 10 exponenciais possuem tempos característicos não muito diferentes entre si. De fato, entre a segunda e a primeira a relação é $\gamma_1/\gamma_2 = 2/1$ mas, entre a última e a penúltima a relação é $\gamma_9/\gamma_{10} = 10/9$, o que significa que as componentes se aproximam muito com o aumento de k .

Tabela 2.6: Parâmetros do sinal temporal mostrado na Figura 2.18.

k	f_k	$1/\gamma_k$	$ d_k $	$\theta_k(\text{rad})$
1	0	1	10	0
2	0	1.5	10	π
3	0	2.25	10	$\pi/2$
4	0	3.4	10	$\pi/3$
5	0	5	10	$\pi/4$
6	0	7.6	10	$\pi/5$
7	0	11.4	10	$\pi/6$
8	0	17	10	$\pi/7$
9	0	25.6	10	$\pi/8$
10	0	38.4	10	$\pi/9$
11	0	57.7	10	0
12	0	86.5	10	π
13	0	130	10	$\pi/2$
14	0	195	10	$\pi/3$
15	0	292	10	$\pi/4$
16	0	438	10	$\pi/5$

Para testar esta última observação elaboramos a outra tabela com 16 conjuntos de parâmetros, a Tabela 2.6, na qual os valores de $1/\gamma$ são equidistantes em uma escala logarítmica, isto é, a relação entre dois valores conse-

cutivos é uma constante, igual a 1.5 neste caso.

Figura 2.18: Análise exponencial feita pelo FDM do sinal simulado de acordo com os parâmetros da Tabela 2.6, tomando-se o intervalo de amostragem $\tau = 0.05$ e $N = 43800$. À esquerda mostra-se o sinal a ser analisado, assim como o módulo do resíduo do ajuste, e na coluna à direita os valores das amplitudes correspondentes a cada um dos 16 valores de $1/\gamma$.

A Figura 2.18 mostra o sinal simulado e o correspondente resultado obtido pelo FDM. Além de “advinhar” que o sinal simulado contém 16 componentes, o FDM foi capaz de calcular todos os parâmetros com precisão razoável! O resíduo mostrado é novamente pequeno e exponencial e a sua forma estranha mostrada no início do decaimento vem do fato de que lá está graficado o módulo do resíduo e, nesta região é onde ele muda de sinal.

Estes resultados são estimulantes mas, todavia, ainda não tivemos tempo para nos aprofundarmos no problema. Talvez o FDM possa vir a ser uma ferramenta útil no desenvolvimento de técnicas para a análise exponencial.

2.8.5 Exemplo 5: Picos não Lorentzianos

A principal hipótese do FDM é que o sinal satisfaça suficientemente bem o modelo Lorentziano proposto pela Equação 2.1. Entretanto, nada impede que o FDM seja aplicado a qualquer outro tipo de sinal, mesmo que este sinal não siga o modelo Lorentziano proposto.

Figura 2.19: Sinal simulado de acordo com a Equação 2.90, para $\tau = 1$ e $N = 2048$, e correspondentes parâmetros espectrais obtidos pelo FDM.

Nesta e na próxima seção vamos analisar duas situações em que o FDM é aplicado a modelos não Lorentzianos. Neste exemplo vamos tratar de uma única componente harmônica com modulação Gaussiana:

$$c(t) = e^{-5 \cdot 10^{-6} t^2} e^{2\pi i t/25}, \quad (2.90)$$

A Figura 2.19 mostra os resultados obtidos. A conclusão é que o FDM necessitou de pelo menos 26 componentes espectrais Lorentzianas para descrever uma única componente espectral Gaussiana. As componentes menos intensas se situam fora da região de frequências mostrada nos gráficos desta figura. A concordância do ajuste com o sinal de entrada é excelente porém são necessárias muitas componentes, centralizadas na frequência do sinal, $1/25 = 0.04$, com diferentes amortecimentos e fases, conforme mostra a figura.

A forma dos dados mostrados na Figura 2.19 é bastante interessante, particularmente àquelas correspondentes ao amortecimento (b) e à amplitude (c). Os pontos relativos ao amortecimento (b) saltam entre duas curvas diferentes para picos consecutivos!

A pergunta que se faz aqui é a seguinte: no caso de sinais não Lorentzianos, como esse, qual seria a utilidade do FDM?

Ainda não sabemos responder esta pergunta!

2.8.6 Exemplo 6: Ruído

Nas aplicações práticas, o sinal a ser analisado possui uma certa quantidade de ruído. Por “ruído” queremos dizer qualquer interferência no sinal que torne a representação da Equação 2.1 inválida. Para implementar exemplos numéricos, utilizamos a forma padrão de ruído gaussiano aditivo, ν_n ,

$$c_n = \sum_{k=0}^{K-1} d_k e^{-in\tau\omega_k} + \lambda\nu_n. \quad (2.91)$$

onde $\text{Re}(\nu_n)$ e $\text{Im}(\nu_n)$ são números aleatórios independentes, uniformemente distribuídos de acordo com uma distribuição normal. Se assumirmos que $\langle \nu_n \rangle = 0$ e $\langle |\nu_n|^2 \rangle = \langle |c_n|^2 \rangle = 1$, então λ mediria a intensidade do ruído.

Devido ao ruído, a quantidade de informação contida na sequência c_n , $n = 0, 1, 2 \dots N-1$, é igual a N e não igual a K como no caso do sinal puro.

Figura 2.20: Um sinal ruidoso analisado pelo FDM e DFT. Os parâmetros utilizados são os da Tabela 2.1 com $N = 1024$, $\tau = 0.02$ e $t_d = 1.0$. O resíduo é a diferença entre o sinal de entrada e o sinal simulado a partir da lista $\{d_k, \omega_k\}$ calculada pelo FDM.

Portanto, o ajuste do sinal de acordo com a expressão 2.1 resultaria $\approx N/2$ pares complexos (d_k, ω_k) , que representariam o espectro do sinal puro e do ruído.

A presença de ruídos de alta intensidade nos sinais a serem analisados pelo FDM representa uma séria dificuldade. Uma análise profunda deste problema demandaria um alongamento excessivo deste texto e, portanto, está fora dos nossos objetivos. Aqui, vamos apenas mostrar um exemplo numérico, a partir do qual o problema poderá ser diagnosticado mais objetivamente sem, entretanto, a necessidade de nos aprofundarmos na matemática do problema

e na proposição de outros recursos do FDM que estão disponibilizados na literatura associada.

A Figura 2.20(a) mostra o sinal a ser analisado, que foi calculado com os parâmetros da Tabela 2.1, com $N = 1024$ pontos, $\tau = 0.02$. O tempo morto foi considerado $t_d = 1.0$, o que corresponde a uma perda dos 50 pontos iniciais.

A Figura 2.20(b) mostra o espectro calculado pela DFT onde se nota que, apesar da presença do ruído, as 6 linhas podem ser facilmente identificadas e as suas amplitudes correspondem de fato às amplitudes esperadas. Isto mostra que, neste caso, a DFT é um bom método para analisar-se este problema, apesar da presença do tempo morto que impossibilita o ajuste da fase do espectro. Por este motivo é mostrado o módulo do espectro de DFT. O ruído do sinal é também transformado pela DFT mas o resultado não é catastrófico, visto que o espectro do ruído se espalha por todo o intervalo de Nyquist, ficando com uma amplitude máxima bem abaixo da amplitude do sinal de interesse.

Tendo em vista estas considerações, concluiríamos que o espectro de DFT é, neste caso, suficiente para a maioria das aplicações e que não haveria a necessidade de nos apoiarmos em outros métodos mais sofisticados, tal como o FDM, a fim de resolver-se novamente um problema que já foi resolvido facilmente pela DFT. Entretanto, utilizaremos o FDM neste problema simplesmente para aprendermos um pouco mais a respeito de suas características e propriedades.

A Figura 2.20(d) mostra o espectro calculado pelo FDM e o resultado obtido, senão desastroso, deixa muito a desejar. Neste espectro mal consegue-se reconhecer as seis linhas Lorentzianas que o sinal simulado possui. Diferentemente da DFT, o ruído no espectro de FDM não é uniformemente dis-

Figura 2.21: Mesmo sinal que o da Figura anterior, porém amostrado com $N = 512$, $\tau = 0.005$ e $t_d = 1.0$.

tribuído pela banda, mas se apresenta na forma de picos espúrios que se confundem com o sinal de interesse. Sem dúvida alguma, este espectro é de muita baixa qualidade se comparado com o da DFT.

Por outro lado, o cálculo dos resíduos (diferença entre o sinal de entrada e o simulado a partir da lista de parâmetros obtidos pelo FDM) mostra que o ajuste do sinal pela Equação 2.1 foi razoável. Isto significa que a lista de parâmetros fornecida pelo FDM não é única. Realizando-se o cálculo novamente com outros parâmetros iniciais do FDM, por exemplo, com outros valores de N , L e ρ , obteríamos um outro bom ajuste (ou baixo resíduo) e um outro espectro bem diferente do anterior. Os dois espectros obtidos para diferentes valores de N , L e ρ diferem fundamentalmente entre si na lista

associada ao picos espúrios, isto é, a lista $\{d_k, \omega_k\}$ é fortemente dependente dos parâmetros iniciais N , L e ρ .

Uma maneira de minimizar estes problemas é utilizar-se de outras técnicas; particularmente eficiente é o método FDM2K, descrito em [7], mas não implementado aqui. Sendo que os artefatos produzidos pelo ruído são dependentes do comprimento do sinal de entrada, utiliza-se neste método um procedimento de média no qual os diferentes espectros obtidos (removendo-se um ou dois ou mais pontos no final do sinal) são somados. Esta soma reduz a amplitude dos picos espúrios e reforça a dos picos genuínos.

Figura 2.22: Mesmo sinal que o da Figura anterior, amostrado com $N = 512$, $\tau = 0.005$ e $t_d = 1.0$, e tratado pelo FDM através do método de truncamento das matrizes por seleção dos valores singulares.

Entretanto, o cálculo do FDM pode, e deve, ser otimizado porque, de fato, ele não é tão ruim assim. O problema que gerou o péssimo espectro do FDM da Figura 2.20(d) é obviamente o excesso de ruído. Entretanto, o efeito da presença de ruído ficou exageradamente acentuado por causa da forma em que o sinal foi amostrado. O sinal da Figura 2.20(a) foi amostrado com $N = 1024$ e $\tau = 0.02$ porque esta é uma situação apropriada para a aplicação da DFT. Para o FDM, que é um processo de ajuste e não uma transformação, este não seria um sinal bem amostrado porque, como o sinal de interesse decai com o tempo, mais da metade do total de pontos está sendo utilizado para amostrar-se somente o ruído. No caso da DFT tenta-se evitar truncamentos do sinal e por isso τ é escolhido o maior possível dentro do limite de Nyquist e N o maior possível também, de tal forma que $N\tau$ seja bem maior que todos os tempos característicos dos decaimentos exponenciais.

Um sinal melhor amostrado para a aplicação do FDM (mas ruim para a DFT) seria o da Figura 2.21(a), com $N = 512$ e $\tau = 0.005$. Vemos que somente a parte inicial do decaimento, na qual a relação sinal a ruído é mais alta, está sendo amostrada. Nesta situação, o resultado do FDM é mais simpático que o anterior e o resíduo é menor, apesar de não calcular corretamente a amplitude dos picos.

Como último recurso, podemos aplicar o método de truncamento das matrizes utilizadas no cálculo, selecionando-se apenas os valores singulares com maior amplitude, conforme foi explicado anteriormente quando descrevemos o método SVD. Observando-se que o espectro obtido na Figura 2.21 e examinando-se o conjunto de L valores singulares obtidos para cada janela, pode-se selecionar somente aqueles de maior amplitude e a partir daí construir-se matrizes truncadas com $L' < L$. Neste caso particular observa-se que $L = 100$ e se tomarmos $L' = 6$ não estaremos fazendo um truncamento e-

xagerado mas, de fato, plausível. O resultado do cálculo do FDM com este truncamento está mostrado na 2.22, onde se nota uma melhora significativa da qualidade do espectro obtido (muito embora ele ainda não consiga reproduzir as amplitudes corretas).

2.9 Conclusões

Neste capítulo apresentamos o método chamado *Filter Diagonalization Method* (FDM), projetado para realizar a análise espectral com alta resolução de sinais no domínio do tempo. Este método resolve o problema do ajuste não linear do *Harmonic Inversion Problem* (HIP) através da álgebra linear, diagonalizando-se pequenas matrizes de dados no domínio da frequência.

Embora o FDM possa ser aplicado a sinais multi-dimensionais, tratamos aqui apenas o caso de sinais uni-dimensionais.

Computacionalmente, o FDM é intrinsecamente estável e eficiente, devido a duas propriedades importantes: ele resolve o HIP através da álgebra linear e a análise espectral local através de uma base do tipo Fourier.

Neste capítulo o método foi aplicado apenas a sinais simulados numericamente, e os resultados foram suficientes para demonstrar que o FDM é capaz de fornecer uma resolução espectral acima do previsto pelo princípio de incerteza tempo-frequência da Transformada de Fourier (FT), desde que o sinal satisfaça suficientemente bem o modelo Lorentziano e que o número de pontos amostrados seja suficientemente longo (embora bem menor que aquele necessário para a aplicação da FT).

Mostramos que o FDM é um procedimento de ajuste de sinais baseado em um modelo Lorentziano e não uma transformação como no caso da FT. Por isso, o resultado do FDM não é representado por um espectro mas sim

por uma lista de parâmetros que descreve o sinal analisado. A existência de uma lista e não de um espectro é uma vantagem do FDM pois facilita a manipulação posterior dos resultados obtidos.

Mostramos também um fato importante: o FDM não substitui a FT. De fato, ficou claro aqui que o FDM é útil apenas nos casos em que a FT não funciona bem, sendo as situações mais típicas aquelas causadas por falta de resolução ou truncamento de sinais.

Capítulo 3

O problema dos autocampos

Passaremos agora a discutir a determinação dos campos ressonantes, também denominados “autocampos”, e para isso vamos considerar um sistema de spins na presença de um campo magnético uniforme, de intensidade h e orientação (θ, φ) com relação a algum sistema de coordenadas. Além dos parâmetros magnéticos que caracterizam o sistema quântico em estudo, o Hamiltoniano do sistema, $\hat{H}$, passa a ser também função de (h, θ, φ) , isto é $\hat{H}(h, \theta, \varphi)$.

O problema consiste em, para uma certa orientação (θ, φ) fixa, encontrar os valores de h para os quais dois autovalores quaisquer de $\hat{H}(h)$ estejam separados pela frequência de operação do espectrômetro, Ω . Matematicamente, esta condição se escreve como:

$$\begin{aligned}\hat{H}(h) |u\rangle &= \nu |u\rangle \\ \hat{H}(h) |v\rangle &= (\nu + \Omega) |v\rangle,\end{aligned}\tag{3.1}$$

onde $|u\rangle$ e $|v\rangle$ são dois estados do sistema de spins separados por uma energia Ω .

A partir daí, os métodos usuais empregam a metodologia que delinearemos a seguir e que está esquematizada na Figura 3.1.

Primeiro, estima-se qual é a região de valores de h em que os campos de ressonância devem se localizar. Na Figura esta região está compreendida no intervalo 0-1500 mT, ou 0-15000 G. Depois, o Hamiltoniano é sistematicamente diagonalizado repetidas vezes para diferentes valores de h , a fim de construir-se o diagrama de níveis de energia mostrado na Figura 3.1(a). Em seguida, constrói-se um segundo gráfico no qual se denota as diferenças de energia entre todos os pares de níveis, conforme está mostrado na Figura 3.1(b). Por último, dada a frequência do espectrômetro, Ω , determina-se no gráfico (b) todos os pontos de intersecção entre uma reta horizontal colocada na posição Ω do eixo vertical com todas as curvas presentes no gráfico.

Figura 3.1: (a) Níveis de energia (em GHz) do íon Tb^{4+} em ThO_2 , utilizado como exemplo dos métodos usuais utilizados para a determinação dos campos ressonantes. (b) Diferença de energia (em GHz) entre todos os pares de níveis mostrados em (a). A linha tracejada denota a frequência de operação do espectrômetro Ω . Os pontos de cruzamento desta linha com as demais fornecem os valores dos campos de ressonância. Exemplo adaptado de [1, 30].

Embora este procedimento seja bastante simples, sua implementação na forma de um algoritmo computacional está longe de ser trivial [30, 31]. Em primeiro lugar, devemos lembrar que, quando se requer a simulação de um espectro de pó, o algoritmo deve ser suficientemente rápido para permitir a repetição do processo inúmeras (talvez milhares de) vezes para diferentes orientações (θ, φ) . Além das exigências de robustez e eficiência, o algoritmo deve ser preciso e quanto a isso, já mencionamos na introdução deste trabalho que diversos métodos sofisticados, porem aproximados, são geralmente utilizados para resolver este problema.

Acrescentamos aqui ainda mais um fato: quase todos os métodos aproximados possuem um ponto fraco. Por exemplo, este é o caso do espectros de RPE em campos magnéticos próximos aos chamados “*crossing points*” e “*anticrossing points*” [12]. As chamadas “*looping transitions*” podem ocorrer no espectro de RPE-CW próximo a estes pontos de cruzamento e envolvem um único par de níveis que estão em ressonância em dois diferentes valores de campo, respectivamente antes e depois do ponto de cruzamento. Quando as distâncias entre estas duas ressonâncias é muito pequena, a forma usual de simular o espectro, que resulta da aproximação linear da dependência da frequência de transição com o campo magnético, falha.

Nossos trabalhos nesta área estão em recente progresso e já podemos encarar os primeiros resultados obtidos com muito entusiasmo, porem ainda falta resolver-se alguns problemas teóricos e computacionais antes de ser possível realizar uma comparação de sua eficiência e precisão com as dos outros métodos existentes, os quais foram construídos para a mesma finalidade.

3.1 A teoria básica dos autocampos

Esta seção está baseada no trabalho de Belford e colaboradores [2] publicado em 1973, no qual uma equação de autocampos é formulada no espaço de Liouville onde o problema dos campos ressonantes pode ser resolvido de uma maneira exata e explícita.

3.1.1 A equação mestra dos autocampos

O procedimento básico assume, a princípio, que o operador Hamiltoniano de spin, $\hat{H}$, pode ser não-hermitiano, embora, é claro que somente os operadores hermitianos é que, neste caso, possuem significado físico. A razão embutida nesta hipótese ficará esclarecida logo adiante.

Além disso, o procedimento assume que o operador Hamiltoniano de spin é uma função linear com a amplitude do campo magnético aplicado, h , isto é:

$$\hat{H} = \hat{F} + h \hat{G}(\theta, \varphi). \quad (3.2)$$

onde $\hat{F}$ e $\hat{G}$ são operadores hermitianos independentes de h , mas $\hat{G}$ carrega a dependência angular do vetor campo magnético. Em geral, os operadores $\hat{F}$ e $\hat{G}$ não comutam entre si.

Uma transição produzida por uma excitação de frequência Ω pode ocorrer entre dois níveis de energia especificados por:

$$\begin{aligned} \hat{H} |u\rangle &= \nu |u\rangle \\ \langle v| \hat{H} &= (\nu - \Omega) \langle v|, \end{aligned} \quad (3.3)$$

onde $|u\rangle$ é um autovetor à direita de $\hat{H}$ com autovalor ν e $\langle v|$ é um autovetor à esquerda com autovalor $(\nu - \Omega)$.

Multiplicando-se a primeira equação à direita por $\langle v|$, a segunda à esquerda por $|u\rangle$ e subtraindo-se as duas, elimina-se a energia ν , ficando:

$$\hat{H} \hat{M} - \hat{M} \hat{H} = \Omega \hat{M} \quad \text{com} \quad \hat{M} = |u\rangle\langle v|. \quad (3.4)$$

A seguir, escolhe-se um base completa de funções ortonormais, $\{|\alpha_i\rangle\}$, com $i = 0 \dots n-1$, onde n é a dimensão do espaço de $\hat{H}$, na qual os autovetores podem ser projetados:

$$|u\rangle = \sum_{k=0}^{n-1} a_k |\alpha_k\rangle \quad \text{e} \quad \langle v| = \sum_{l=0}^{n-1} b_l \langle \alpha_l| \quad (3.5)$$

Substituindo-se estas somas na Equação 3.4 e multiplicando-a à esquerda por $\langle \alpha_i|$ e à direita por $|\alpha_j\rangle$ obtemos a equação matricial:

$$H M - M H = \Omega M, \quad (3.6)$$

onde os elementos de matriz são dados por:

$$\begin{aligned} H_{ij} &= \langle \alpha_i | \hat{H} | \alpha_j \rangle \\ M_{ij} &= a_i b_j \end{aligned} \quad (3.7)$$

Substituindo-se a hipótese formulada na Equação 3.2 se obtém a equação de autocampos desejada, que aqui denominamos de **equação mestra dos autocampos**:

$$\Omega M - FM + MF = h(GM - MG), \quad (3.8)$$

onde o campo de ressonância h é o autovalor; F , G e M são matrizes n -por- n e M é a automatriz.

A equação mestra dos autocampos, Equação 3.8, admite soluções para certos valores de h , que coincidem com os desejados campos de ressonância,

e as correspondentes automatrizes fornecem toda a informação necessária para se calcular os momentos de transição.

Mostraremos adiante que, em geral, a Equação 3.8 pode possuir autovalores complexos, mesmo considerando-se que F e G são hermitianos. De acordo com a hipótese da Equação 3.2 esta situação especifica operadores Hamiltonianos não-hermitianos. É claro que neste contexto, os autovalores complexos são desnecessários e portanto podem ser simplesmente ignorados.

3.1.2 Solução da equação mestra no espaço n^2 .

Até onde temos conhecimento, não existe registro na literatura de tentativas, bem sucedidas ou não, de resolver-se a Equação 3.8 através de algoritmos numéricos. Realizamos algumas tentativas para determinar os autovalores e autovetores desta equação através de métodos iterativos, mas a execução do algoritmo se torna muito lenta, o que inviabiliza sua aplicabilidade.

Pode ser que por este motivo, o trabalho da ref. [2] se utiliza, neste ponto, de uma identidade associada ao produto de Kronecker, válida para quaisquer três matrizes A , B e X ,

$$vec(AXB) = (A \otimes B^T) vec(X), \quad (3.9)$$

onde $\otimes$ denota o produto de Kronecker e $vec(C)$ é um vetor coluna n^2 -por-1 que denota a vetorização da matriz C :

$$vec(C) = [C_{0,0}, C_{0,1}, \dots, C_{0,n-1}, C_{1,0}, \dots, C_{1,n-1}, \dots, C_{n-1,n-1}]^T,$$

de tal forma que a Equação 3.8 pode ser transformada em uma equação de autovalores generalizada:

$$AZ = hBZ, \quad (3.10)$$

onde

$$\begin{aligned} A &= \Omega I \otimes I - F \otimes I + I \otimes F^T \\ B &= G \otimes I - I \otimes G^T, \end{aligned} \quad (3.11)$$

com $Z = \text{vec}(M)$, A , B matrizes n^2 -por- n^2 e I a matriz identidade n -por- n .

Na ref. [2] este resultado é explorado em várias situações práticas mas, aqui apenas discutiremos algumas de suas propriedades e detalhes que serão importantes ao nosso trabalho.

Primeiro, sendo que F e G são hermitianos, também o são A e B . Em geral, B é uma matriz singular mas A quase nunca é singular - A se torna singular somente quando a frequência da radiação, Ω , é exatamente igual a alguma diferença entre dois níveis de energia em campo nulo. Portanto, quando A não é singular, a Equação 3.10 reduz-se a uma simples equação de autovalores:

$$CZ = (1/h)Z, \quad \text{com} \quad C = A^{-1}B. \quad (3.12)$$

Em geral, a matriz C não é hermitiana, mesmo quando os autovalores $1/h$ são todos reais. Neste caso, pode-se realizar uma transformação apropriada que transforma C em uma matriz hermitiana.

Esta transformação, não demonstrada na ref. [2], é executada da seguinte maneira. Primeiro, exprime-se as matrizes F e G em uma base na qual F é diagonal. As novas matrizes F e A são diagonais com elementos reais. Entretanto, a matriz diagonalizada A pode não ser definida-positiva, isto é, ela pode ter autovalores (ou elementos da diagonal) negativos, e isto sempre ocorre quando a frequência da radiação, Ω , é menor que algum desdobramento a campo nulo.

Segundo, transforma-se a matriz A da seguinte maneira:

$$\begin{aligned}
A &= S|A| = S|A|^{1/2}|A|^{1/2} \\
|A|_{ij} &= |A_{ii}| \delta_{ij} \\
|A|_{ij}^{1/2} &= |A_{ii}|^{1/2} \delta_{ij}
\end{aligned} \tag{3.13}$$

onde S é uma matriz diagonal cujos elementos valem $S_{ii} = +1$ ou $S_{ii} = -1$ dependendo se o correspondente A_{ii} é positivo ou negativo, respectivamente e, obviamente, S possui a propriedade: $S^2 = I$. Esta matriz S é conhecida na Álgebra Linear pela designação *signature matrix*.

Terceiro, a partir de 3.10, escrevemos:

$$S|A|^{1/2}Z = h |A|^{-1/2}BZ = h (|A|^{-1/2}B|A|^{-1/2}S) (S|A|^{1/2}Z), \tag{3.14}$$

ou seja,

$$DX = (1/h)X, \tag{3.15}$$

com

$$\begin{aligned}
X &= S|A|^{1/2}Z \\
D &= |A|^{-1/2}B|A|^{-1/2}S
\end{aligned} \tag{3.16}$$

Esta matriz D possui várias propriedades interessantes. Primeiro, a sua adjunta vale $D^H = SDS$ (ou, $SD^H = DS$). Logo, se A for definida-positiva, então $S = I$ e D se torna hermitiana. Portanto, quando A é definida-positiva, os autovalores $1/h$ de D são números reais.

Segundo, ordenando-se os elementos da diagonal de A , tal que os positivos ocupem as primeiras posições e os negativos as últimas, e aplicando-se a mesma troca de linhas e colunas na matriz D , resulta que esta matriz pode ser escrita na forma de blocos:

$$D = \begin{bmatrix} H_1 & H_3 \\ -H_3^H & H_2 \end{bmatrix}, \quad (3.17)$$

onde H_1 e H_2 são matrizes quadradas hermitianas m -por- m e $(n - m)$ -por- $(n - m)$, respectivamente, com m sendo o número de elementos positivos em A ; e H_3 uma matriz retangular m -por- $(n - m)$.

A forma em blocos de D é interessante e implica em várias propriedades. De fato, ela possui uma estrutura similar às matrizes envolvidas nos *saddle point problems*, sobre as quais existe uma vasta literatura disponível [4]. Entretanto, como não estamos interessados em teorizar nesta direção, não vamos discutir este assunto com mais detalhes.

Resumindo-se, podemos dizer que o problema dos autocampos está resolvido exatamente, e para isso basta que se resolva com precisão uma das Equações 3.10, 3.12 ou 3.15.

Em princípio, a formulação básica da teoria dos autocampos e a sua correspondente solução simples no espaço n^2 parece ser direta e prática. Entretanto, se buscarmos por aplicações práticas, encontraremos inúmeros casos de interesse nos quais o número de estados de spin, n , (envolvendo variáveis eletrônicas e nucleares) pode ser muito grande.

Para citar um único exemplo, lembramos o caso apontado na ref. [30] de um íon ^{63}Cu ($S = 1/2$), cercado por quatro núcleos ^{14}N ($I = 1$), levando a uma situação na qual os 628 estados resultam em um problema de autovalores generalizado envolvendo matrizes de dimensão $628^2 = 394384$. Dependendo da aritmética em dupla precisão do ambiente computacional utilizado, mais de 2000 GB de memória seriam necessários para armazenar uma única matriz complexa!

É fato que as matrizes A , B , C e D definidas acima são bastante esparsas, entretanto, parece que este fato não ajuda muito na simplificação

deste problema.

Além disso, o tempo de execução envolvido na solução numérica de um problema de autovalores cresce com o cubo da dimensão das matrizes o que significa, no caso presente, com a sexta-potência do número de estados de spin!

Estas são as razões pelas quais a solução da equação mestra dos autocampos no espaço n^2 tem ficado restrita a problemas de baixa dimensão, tendo sido substituída (ou adaptada), no caso de sistemas de maior dimensão, por métodos aproximados.

3.2 Uma nova proposta

Levando em consideração o que foi concluído na seção anterior, vamos retornar ao espaço de dimensão n , a fim de buscar métodos alternativos que possam solucionar a equação mestra dos autocampos de uma maneira precisa e prática.

3.2.1 Considerações gerais

Uma propriedade da Equação 3.8 que pode ser facilmente demonstrada é a seguinte. Tomando-se o adjunto da Equação 3.8 e trocando-se os sinais dos dois membros, obtém-se:

$$-\Omega M^H - FM^H + M^H F = h^*(GM^H - M^H G), \quad (3.18)$$

Portanto, uma propriedade estrutural da Equação 3.8 é que, se h e M são autovalor e automatriz, respectivamente, então, h^* e M^H também são autovalor e automatriz da equação de autovalores correspondente a uma excitação de frequência $-\Omega$. Considerando que a troca de Ω por $-\Omega$ não muda

a Física do problema, tanto a Equação 3.8 como a 3.18 devem ter o mesmo conjunto de autovalores, logo, devemos acreditar que h^* também é autovalor da Equação 3.8, se h o for. A determinação da automatriz correspondente, assim como uma demonstração rigorosa desta conclusão, deixaremos para um trabalho futuro.

O fato importante é que, além desta última propriedade, a Equação 3.8 possui outras propriedades que não advêm de sua estrutura, mas sim das propriedades de simetria dos operadores $\hat{F}$ e $\hat{G}$. O resultado deste estudo, que deixaremos para relatar no futuro, é que, se h é autovalor da Equação 3.8 então $-h^*$ também é autovalor da mesma equação. Este fato está demonstrado na ref. [2].

Portanto, a conclusão é que cada autovalor real da Equação 3.8 tem a companhia de um outro autovalor real de mesma magnitude mas de sinal contrário, assim, autovalores reais aparecem em duplas, $\pm h$. Além disso, autovalores complexos aparecem como quartetos, do tipo: $\pm h$ e $\pm h^*$. Um outro fato que terá importância mais adiante é que, quando o autovalor é complexo, suas partes real e imaginária possuem magnitudes comparáveis, entretanto, vamos deixar este fato sem demonstração, por enquanto.

Ainda mais, sabemos que, dada a dimensão n do Hamiltoniano, o número máximo de autocampos, ou de transições (permitidas ou não), vale $n_t = (n^2 - n)/2$, que é bem menor que o número máximo de autovalores da Equação 3.8, que é n^2 . Logo, tomando como exemplo o caso em que todos os autovalores são reais, tem que haver no máximo $2n_t$ autovalores não-nulos e, portanto, no mínimo $n^2 - 2n_t = n$ autovalores nulos. Portanto, o espaço nulo deste problema é de dimensão maior ou igual que n . Entretanto, a existência deste espaço nulo não tem, até o momento provocado dificuldades no desenvolvimento teórico que estamos realizando.

Nas próximas seções vamos apresentar uma nova maneira de resolver-se a Equação 3.8 sem a necessidade de resolver-se equações de autovalores.

O novo método é inspirado no Método da Diagonalização Filtrada, FDM, que foi detalhadamente discutido no Capítulo anterior. Sob a luz do FDM, mostraremos que a Equação 3.8 pode ser resolvida no espaço n em diversas situações práticas, sem limitações restritivas sobre o valor de n . As situações em que o novo método se aplica sem restrições são aquelas em que a frequência da radiação, Ω , é maior que todos os desdobramentos de campo nulo. Quando a frequência da radiação é menor que algum dos desdobramentos de campo nulo o método fica limitado a sistema de baixas dimensões, isto é, para n não muito grande.

Acreditamos, baseados nos resultados que obtivemos recentemente, e não relatados aqui, que esta última restrição é puramente matemática e poderá ser removida com facilidade através de técnicas especiais.

3.2.2 A equação diagonal em G

Para o nosso enfoque ao problema dos campos de ressonância estaremos buscando uma equação dinâmica ou, mais especificamente, uma equação diferencial linear de primeira ordem em uma variável $L(x)$, onde L é uma matriz n -por- n que é função de um escalar x , o qual está associado à variável campo magnético.

Esta equação deve ter a propriedade de que: (i) sua solução estacionária seja determinada pela equação mestra, Equação 3.8 e, (ii) a informação espectral contida nos elementos da matriz $L(x)$ deva estar diretamente relacionada com os campos de ressonância, ou autocampos.

Pode parecer a primeira vista que estas intenções não são realistas, entretanto, demonstraremos que tal equação existe e, além disso, ela não é

única.

Em vez de investir esforços em cálculos baseados em primeiros princípios, procuramos inicialmente simplesmente “advinhar” uma equação que satisfaça as nossas necessidades.

A primeira equação que propomos é a seguinte equação diferencial:

$$G \frac{d}{dx} L - \left(\frac{d}{dx} L \right) G = 2\pi i [(\Omega I - F) + LF]. \quad (3.19)$$

Não é difícil mostrar que esta equação admite soluções do tipo:

$$L(x) = \sum_k M_k e^{i \lambda_k x}, \quad (3.20)$$

desde que as matrizes independentes de x , M_k , sejam automatrizes da equação de autovalores

$$\Omega M_k - F M_k + M_k F = \frac{\lambda_k}{2\pi} (G M_k - M_k G), \quad (3.21)$$

e, desde que as matrizes M_k formem um espaço completo, tal que qualquer matriz solução da equação diferencial 3.19 possa ser projetada nas automatrizes.

Esta última suposição pode não ser restritiva, pois, retornando ao formalismo no espaço n^2 e, lembrando a Equação 3.15, vemos que, quando D for hermitiano, automaticamente seus autovetores são ortonormais entre si e formam uma base completa. Quando D não for hermitiano, seus autovetores podem deixar de serem ortogonais entre si, entretanto, espera-se que a matriz D , que neste caso não é uma matriz normal, não se torne patológica (isto é, *defective*). Não sabemos demonstrar esta suposição, mas, acreditamos que ela deve ser verdadeira.

Comparando-se a Equação 3.21 com a equação mestra, pode-se notar que elas se tornam idênticas quando os campos de ressonância são definidos como

$h_k = \lambda_k/(2\pi)$, o que significa que, neste caso, a unidade da variável x é a do recíproco do campo magnético.

Utilizando métodos convencionais, a serem descritos na próxima seção, nós resolvemos a Equação 3.19 em diversas situações práticas e concluimos que, realmente, a partir de suas soluções, todos os campos de ressonância podem ser reproduzidos com precisão. Não apresentaremos aqui estes exemplos porque, como veremos, esta equação, embora correta, não é a que utilizaremos no nosso trabalho.

É importante observar que a Equação 3.19 possui características patológicas que se tornam evidentes mesmo nas situações mais simples. Para entender este fato, é conveniente considerá-la na representação em que G é diagonal:

$$(G_{ii} - G_{jj}) \frac{d}{dx} L_{ij} = 2\pi i \left[\Omega L_{ij} + \sum_k (F_{kj} L_{ik} - F_{ik} L_{kj}) \right]. \quad (3.22)$$

Daqui, conclui-se imediatamente que se qualquer autovalor de G for degenerado (e isto ocorre freqüentemente), esta equação se torna um sistema de equações algébrico-diferenciais. Por outro lado, se esta degenerescência for levantada por pequenas quantidades, simplesmente como acontece quando a interação Zeeman nuclear é incluída no Hamiltoniano de spins, surge a possibilidade de estarmos frente a um sistema *stiff* de equações diferenciais.

Por causa destes “defeitos”, abandonamos esta primeira tentativa, buscando uma maneira mais fácil de resolver-se o problema.

3.2.3 A equação diagonal em F

Uma segunda possibilidade surge com o seguinte sistema de equações diferenciais:

$$(\Omega I - F) \frac{d}{dx} L + \left(\frac{d}{dx} L \right) F = 2\pi i (GL - LG). \quad (3.23)$$

Também não é difícil mostrar que esta equação admite soluções do mesmo tipo que a mostrada em 3.20, desde que as matrizes independentes de x , M_k , sejam automatrizes da equação de autovalores:

$$\frac{\lambda_k}{2\pi} [\Omega M_k - F M_k + M_k F] = (G M_k - M_k G), \quad (3.24)$$

e também, como já discutimos na seção anterior, desde que as matrizes M_k formem um espaço completo, no qual qualquer matriz solução da equação diferencial 3.23 possa ser projetada.

Esta última equação de autovalores é idêntica à equação mestra, Equação 3.8, desde que os campos de ressonância sejam definidos como $h_k = 2\pi/\lambda_k$, o que significa que a variável x tem agora a mesma unidade que o campo magnético. Assim, se o parâmetro λ_k é chamado de “frequência”, o campo de ressonância associado é o “período” correspondente.

Lembrando que o traço de um comutador é nulo, notamos que a matriz $L(x)$, solução da Equação 3.23, possui uma importante propriedade: seu traço é independente de x .

Agora, nossa intenção é resolver a Equação 3.23. Até onde temos conhecimento esta equação não tem solução analítica. Por isso, optamos resolvê-la através de um método numérico. Após tentar algumas opções disponíveis, concluímos que o método de Runge-Kutta de quarta ordem funciona bem. Bem melhor, é o método de Runge-Kutta de quinta ordem com controle dos erros de truncamento por ajuste auto-adaptável do passo de integração. Para maiores detalhes sobre estes métodos, consulte a ref. [28].

Em princípio, a escolha da condição inicial não é um passo crítico no problema. Usualmente escolhemos uma matriz $L_0 = L(0)$ cujos elementos

da diagonal são nulos, o que, através da propriedade de que o traço se conserva, facilita a implementação de um teste para gerenciar a qualidade do processo de integração. Os elementos fora da diagonal são tomados como números complexos, cujas parte real e imaginária são independentemente sorteadas aleatoriamente no intervalo -1.0 a $+1.0$. Embora este tenha sido o nosso procedimento usual, nada impede que outras condições iniciais sejam escolhidas. Aparentemente, os resultados, isto é os valores finais dos autocampos, são pouco dependentes da escolha da condição inicial.

A condição inicial, de fato, determina a estrutura das matrizes $L(x)$. No Exemplo 1, a seguir, mostraremos um caso em que uma escolha apropriada da condição inicial permitiu-se obter matrizes “quase-hermitianas”, termo este que será definido adiante.

Quando se integra a Equação 3.23 através do método mais simples, conhecido como Runge-Kutta de primeira ordem, ou, método de Euler, o problema se reduz em resolver-se a seguinte equação algébrica:

$$(\Omega I - F)L_{k+1} + L_{k+1}F = [(\Omega I - F)L_k + L_k F] + 2\pi i(GL_k - L_k G)\Delta x. \quad (3.25)$$

Nesta expressão, $\Delta x = x_{k+1} - x_k$, $L_{k+1} = L(x_{k+1})$, $L_k = L(x_k)$, com a condição inicial $L_0 = L(0)$. A extensão para as equações de ordem superior é trabalhosa porem trivial.

Analisando-se a Equação 3.25 notamos que ela é do tipo $PX + XQ = R$, com P , Q e R matrizes conhecidas e $X = L_{k+1}$ uma matriz a ser determinada. Quando as matrizes P e Q são triangulares superior, esta equação é conhecida como “equação triangular de Sylvester” e pode ser resolvida através de pacotes computacionais bem estabelecidos e gratuitos.

A matriz F pode sempre ser transformada em triangular superior através da fatorização de Schur, isto é, $F = UTU^H$, onde, T é triangular superior e

U é unitária.

O ponto importante aqui é que, desde que os autovalores de P , denotados por α e os de Q , denotados por β , satisfaçam $(\alpha + \beta) \neq 0$, para todo α e β , a solução da Equação de Sylvester será única! Portanto, para uma dada condição inicial, a solução da Equação 3.23 é sempre única, sendo a única exceção o caso em que a frequência da radiação, Ω , é exatamente igual a um desdobramento de campo nulo.

Em vez de utilizar-se da equação triangular de Sylvester, é mais fácil resolver-se a Equação 3.23 na representação em que F é diagonal:

$$(\Omega - F_i + F_j) \frac{d}{dx} L_{ij} = 2\pi i \sum_k (G_{ik} L_{kj} - G_{kj} L_{ik}) \quad (3.26)$$

Desde que $(\Omega - F_i + F_j) \neq 0$, para todo i e j , esta equação representa um conjunto de equações diferenciais ordinárias de primeira ordem.

Antes de seguir adiante, é instrutivo que elaboremos alguns exemplos que poderão ilustrar o método proposto.

3.2.4 Exemplo 1: o Hamiltoniano diagonal

De agora em diante, vamos considerar o modelo teórico descrito pelo Hamiltoniano de spin mais comum na área da RPE:

$$\hat{H} = \mu_B \vec{H} \cdot \vec{g} \cdot \vec{S} - \mu_n g_n \vec{H} \cdot \vec{I} + \vec{S} \cdot \vec{A} \cdot \vec{I} + \vec{S} \cdot \vec{D} \cdot \vec{S}, \quad (3.27)$$

$\vec{S}$ é o operador spin eletrônico, $\vec{I}$ é o operador spin nuclear, $\vec{H}$ é o vetor campo magnético, μ_B é o magneton de Bohr, μ_n é magneton nuclear, g_n é o fator-g nuclear, $\vec{g}$ é o tensor-g eletrônico, $\vec{A}$ é o tensor de acoplamento hiperfino e $\vec{D}$ é o tensor de desdobramento de campo zero. Os operadores

$\hat{F}$ e $\hat{G}$, assim como suas representações na forma de matrizes, podem ser trivialmente obtidas neste caso.

Um exemplo simples, porem não trivial é caso isotrópico, $S = 1$ e $I = 1/2$, no qual as matrizes F e G são simultaneamente diagonais na base: $|\alpha_0\rangle = |1, +\rangle$, $|\alpha_1\rangle = |1, -\rangle$, $|\alpha_2\rangle = |0, +\rangle$, $|\alpha_3\rangle = |0, -\rangle$, $|\alpha_4\rangle = |-1, +\rangle$ e $|\alpha_5\rangle = |-1, -\rangle$. Os parâmetros do Hamiltoniano são dados por: $g_{zz} = 2.0$, $A_{zz} = 1500$ MHz, $D = 900$ MHz, $E = 0$ ¹ e $\theta = \varphi = 0$. A frequência da radiação vale $\Omega = 9500$ MHz.

Após construir-se as matrizes F e G , a equação diferencial 3.26 foi resolvida para encontrar-se a dependência com x de cada elemento $L_{ij}(x)$. Como condição inicial tomamos uma matriz L_0 na qual todos os elementos são reais e iguais a 1. Neste caso o traço de $L(x)$ será igual a $n = 6$.

A Figura 3.2 mostra a dependência de L_{52} e L_{04} com x . Conforme previsto pela Equação 3.20, cada uma destas funções pode ser ajustada por funções harmônicas, cujos períodos são os campos de ressonância associados. Desenhando-se os níveis de energia, pode-se verificar esta afirmação, isto é, para cada campo calculado (ou período calculado) existe uma possível transição entre dois níveis de energia. Por exemplo, a componente L_{52} oscila com um período que corresponde ao campo de ressonância entre o terceiro e o quinto nível de energia (ver Figura 3.2(A1,B1)).

O período (ou frequência) associado a diferentes elementos de L podem ser diferentes entre si, mas, o fato importante é que todos eles podem ser ajustados a diferenças entre níveis e, portanto, associados a transições possíveis.

Este fato está mostrado na Tabela 3.1, onde todas as transições encontradas estão associadas aos seus respectivos L_{ij} . Pode-se notar que todas

¹ $D_{xx} = -D/3 + E$, $D_{yy} = -D/3 - E$, $D_{zz} = -2D/3$

Figura 3.2: Elementos de matriz $L_{ij} \equiv L(i, j)$ em função do campo magnético x (A1, A2) e correspondentes transições entre níveis de energia (B1, B2) para o sistema de spins $S = 1$ e $I = 1/2$ do Exemplo 1. À esquerda, linhas sólidas correspondem à componente real e tracejadas à componente imaginária. A dependência com x da soma dos elementos da matriz L , $C(x)$, conforme definido no texto, está mostrado em (A3), com as correspondentes transições entre níveis de energia (B3).

as possíveis transições estão assinaladas, exceto aquelas entre níveis hiperfínicos com mesmo m_s , onde a diferença de energia é muito pequena para a frequência da radiação escolhida.

Vale a pena salientar que, embora não se note nos diagramas de energia da Figura 3.2, o número total de níveis de energia é 6, e não 5 como aparentemente demonstra a figura. O fato é que o nível central, quase independente do campo, não é degenerado. Sua degenerescência é ligeiramente levantada pela

Tabela 3.1: Mapa da matriz L do Exemplo 1, mostrando as transições entre níveis de energia para cada elementos de matriz L_{ij} .

		0 $ 1, +\rangle$	1 $ 1, -\rangle$	2 $ 0, +\rangle$	3 $ 0, -\rangle$	4 $ -1, +\rangle$	5 $ -1, -\rangle$
0	$\langle 1, + $			2-5	3-5	0-5	1-5
1	$\langle 1, - $			2-4	3-4	0-4	1-4
2	$\langle 0, + $	1-3	0-3			0-2	1-2
3	$\langle 0, - $	1-2	0-2			0-3	1-3
4	$\langle -1, + $	1-4	0-4	3-4	2-4		
5	$\langle -1, - $	1-5	0-5	3-5	2-5		

presença do termo de interação Zeeman nuclear no Hamiltoniano escolhido inicialmente.

Este particular exemplo é simples porem serve como um guia útil para a elaboração de um método capaz de encontrar todos os campos de ressonância de uma única vez, isto é a partir de um único cálculo e em uma situação mais geral.

Olhando a Tabela 3.1 notamos que a matriz L não é simétrica, nem tampouco hermitiana, entretanto, os elementos que estão acima da diagonal fornecem toda a informação necessária para identificar-se todas as transições possíveis. De fato, todos os elementos abaixo da diagonal estão associados às mesmas transições que aparecem no triângulo superior da matriz e, neste caso, poderiam ser considerados redundantes.

De fato, para cada elemento acima da diagonal existe um outro, abaixo da diagonal, que corresponde ao seu conjugado, por exemplo, $L(0, 5) = L(5, 0)^*$ e $L(0, 4) = L(5, 1)^*$. Para este caso particular, costumamos empregar um termo que não existe na Álgebra Linear: dizemos que estas matrizes $L(x)$

são “quase-hermitianas”.

Esta aparente simetria é casual, consequência do fato de termos considerado um Hamiltoniano que é diagonal na base escolhida e da condição inicial escolhida. Também por este motivo é que cada elemento da matriz L está associado a uma única transição. Em casos mais gerais (não diagonais), cada elemento da matriz estará associado a mais do que uma transição, conforme descreve a Equação 3.20, e, nestes casos, nenhum elemento do triângulo inferior será igual a algum outro elemento do triângulo superior.

Em um caso geral, utilizando-se de uma condição inicial apropriada, é possível realizar-se transformações que gerem uma nova matriz L quase-hermitiana, no sentido que, a cada elemento do triângulo inferior corresponda um elemento conjugado no triângulo superior. Além disso, também é possível realizar-se transformações que gerem uma matriz L de fato hermitiana. Entretanto, não pretendemos utilizar estes recursos neste trabalho e, por isso, não relataremos estes métodos aqui.

Considerando-se que, em uma situação geral, não se espera qualquer simetria presente na matriz L , e o nosso interesse é identificar todos os campos de ressonância associados ao problema, adotamos um procedimento no qual todos os elementos da matriz L são analisados **em um único procedimento**, em vez de analisar-se cada elemento separadamente.

Isto pode ser feito a partir da definição de uma certa função de correlação:

$$C(x) = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} L_{ij}(x) \quad (3.28)$$

A função de correlação do presente exemplo está mostrada na Figura 3.2(A3). De fato, a análise espectral desta função nos fornece toda a informação necessária para identificar-se todos os campos de ressonância e isto pode ser verificado no diagrama de energia da Figura 3.2(B3).

Tabela 3.2: Campos de ressonância calculados para o Exemplo 1. A segunda coluna mostra os valores dos autocampos calculados através da equação de autovalores, 3.10, e a terceira os calculados através da equação diferencial 3.26. Os erros indicados correspondem ao desvio, em ppm, entre a energia da radiação e a separação entre os níveis de energia em cada campo calculado pela Equação 3.26.

Transição	Eq.3.10	Eq.3.26	Error
0-5	1428.9546	1428.9562	0.932
0-4	1695.5941	1695.5950	0.559
1-5	1698.1751	1698.1760	0.555
1-4	1964.8126	1964.8132	0.359
2-5	2804.3234	2804.3236	0.0618
3-5	2808.5954	2808.5956	0.0614
2-4	3335.1086	3335.1087	0.0368
3-4	3340.1814	3340.1815	0.0366
0-3	3442.1174	3442.1175	0.0335
0-2	3447.3530	3447.3531	0.0333
1-3	3983.2109	3983.2110	0.0216
1-2	3989.2787	3989.2788	0.0215

A questão fundamental que surge aqui é: em um caso geral, que tipo de análise espectral deve ser feita, a fim de gerar uma identificação **precisa** de **todos** os campos de ressonância?

O aspecto mais importante relacionado com esta pergunta é: **resolução**.

Sabe-se, por exemplo, que a interação hiperfina pode levar a desdobramentos de níveis de energia tão pequenos quanto da ordem de 100 ppm (ou até menos). Além disso, como já mencionamos na introdução deste capítulo,

existem os chamados “*anticrossing points*” nos diagramas de níveis de energia. Nestes pontos, ou na vizinhança deles, dois níveis de energia, podem ficar muito próximos entre si, sem contudo tornarem-se degenerados.

3.3 Análise espectral da função de correlação

Como ponto de partida, vamos assumir que a função de correlação definida pela Equação 3.28 é equivalente a um sinal complexo discreto, c_p , cuja forma analítica é dada por uma soma de funções harmônicas complexas:

$$C(x_p) = c_p = \sum_{k=0}^{K-1} d_k e^{-i \lambda_k p \delta} \quad \text{com} \quad p = 0, 1, \dots, N-1. \quad (3.29)$$

O intervalo de amostragem, δ , e o índice p definem uma grade finita e equidistante; d_k são as amplitudes complexas e λ_k as frequências complexas. Consideraremos muitas vezes o caso em que λ_k é real, embora esta não seja uma restrição necessária.

Assumindo que a dimensão dos dados, N , é suficientemente grande, o objetivo é determinar-se a lista de parâmetros, $\{d_k, \lambda_k\}$, que melhor se ajusta aos dados c_p .

Este problema foi o objeto de estudo do Capítulo anterior e lá ele foi denominado de “Problema da Inversão Harmônica”, HIP. O método utilizado para resolvê-lo foi o “Método da Diagonalização Filtrada”, FDM.

Iremos agora utilizar o Exemplo 1, para comparar as resoluções do FDM e da Transformada de Fourier Discreta, DFT. No cenário da DFT, a resolução $1/\Delta\lambda$ no domínio da frequência é expressa como $\Delta\lambda = 2\pi/(N\delta)$ (ver Capítulo 1). Em palavras, a menor separação detectável entre duas frequências adjacentes ($\Delta\lambda$) é inversamente proporcional ao tempo de aquisição total

$(N\delta)$. Portanto, a resolução na escala dos autocampos ($h = 2\pi/\lambda$) é dada por:

$$\Delta h = \frac{h^2}{N\delta}. \quad (3.30)$$

Neste contexto, a definição do dubleto centrado em 3986 G (ver a Tabela 3.2), com uma resolução de 100 ppm, necessitaria que N fosse pelo menos igual a 57000, isto é, assumindo que tomemos $\delta = 700$ G, que é o maior intervalo de amostragem permitido pelo critério de Nyquist.

Para resolver-se o caso do Exemplo 1, através do FDM, utilizamos $N = 1024$ e $\delta = 195.5$ G (note que somente uma parte de $C(x)$ está mostrada no gráfico da Figura 3.2(A3)) e os resultados obtidos para os autocampos estão mostrados na Tabela 3.2. Pode-se notar que a diferença entre a energia da radiação e a separação entre os níveis em cada autocampo calculado é menor que 1 ppm no pior caso.

Este resultado valida a hipótese de que o FDM é um método adequado para resolver este problema.

3.4 Exemplo 2: o Hamiltoniano não-diagonal

Neste segundo exemplo, vamos tratar de um caso anisotrópico, com $S = 2$ e $I = 1$, no qual as matrizes F e G não podem ser simultaneamente diagonalizadas. O número total de estados é $n = 15$ e os parâmetros do Hamiltoniano são dados por: $g_{xx} = 1.9$, $g_{yy} = 2.1$, $g_{zz} = 2.0$, $A_{xx} = 300$ MHz, $A_{yy} = 500$ MHz, $A_{zz} = 1500$ MHz, $D = 1000$ MHz, $E = 400$ MHz, $\theta = 35^\circ$, $\varphi = 55^\circ$ e $\Omega = 9500$ MHz.

Após construir-se as matrizes F e G a equação diferencial 3.26 foi resolvida para calcular-se $C(x)$, na faixa 0-600000 G, com $N = 32768$ e $\delta =$

Figura 3.3: Diagrama de níveis de energia mostrando as 90 transições encontradas para o sistema de spins $S = 2$ e $I = 1$ do Exemplo 2.

18.31 G. Os campos de ressonância, um total de 90, se situam na faixa 93-5000 G e puderam ser todos calculados pelo FDM, com um erro máximo de 1 ppm, relativo aos valores encontrados pela Equação 3.10.

Os resultados deste experimento numérico estão mostrados nas Figuras 3.4 e 3.3. Usualmente, após calcular-se a função de correlação, esta é multiplicada por uma função exponencial amortecida, conforme mostra a Figura 3.4. Este procedimento facilita a remoção dos picos espúrios encontrados pelo FDM pois, sendo que a constante de tempo associada a todas as funções harmônicas verdadeiras são conhecidas a priori, pode-se acionar filtros que as selecionem, desprezando as espúrias. Com este procedimento, a lista correta é obtida de uma única execução do FDM.

Figura 3.4: Componente real da função de correlação $C(x)$, modulada por um amortecimento exponencial, $d = 150000$, para o sistema de spins $S = 2$ e $I = 1$ do Exemplo 2.

3.5 Exemplo 3: o caso singular

Quando a frequência da microonda é exatamente igual a um desdobramento de campo nulo, um ou mais coeficientes, $(\Omega - F_i + F_j)$, no lado esquerdo da Equação 3.26, serão nulos. Neste caso, o sistema de equações representa um sistema de equações algébrico-diferenciais, que não podem ser resolvidas com os métodos simples que adotamos aqui.

Em vez de usarmos procedimentos mais complexos, capazes de resolver este tipo de problema, iremos adotar um procedimento simples, que utiliza-se do mesmo procedimento anterior e é baseado em um “truque” válido, que

consiste em modular-se a frequência da radiação.

Para o primeiro passo de Runge-Kutta tomamos um valor para a frequência, Ω_1 , que é ligeiramente acima da frequência desejada Ω , tal que $\Omega_1 = \Omega + \Delta_1$. Para $\Omega \approx 9500$ MHz, uma boa escolha pode ser Δ_1 na faixa 50-100 MHz.

Para o segundo passo, a frequência é alterada para $\Omega_2 = \Omega - \Delta_2$, e o valor de Δ_2 é calculado para cada uma das equações do sistema de equações 3.26, de acordo com a seguinte receita:

$$\begin{aligned}
 & 1/(\Omega_1 - F_i + F_j) + 1/(\Omega_2 - F_i + F_j) = 2/(\Omega - F_i + F_j) \\
 & \quad \text{se } \Omega - F_i + F_j \neq 0 \\
 & \Omega_1 = \Omega + \Delta_1 \quad \text{e} \quad \Omega_2 = \Omega - \Delta_1 \\
 & \quad \text{se } \Omega - F_i + F_j = 0
 \end{aligned} \tag{3.31}$$

Depois, o procedimento é repetido para todos os outros passos, isto é, aplica-se Ω_1 nos passos ímpares e Ω_2 nos passos pares.

O resultado desta manipulação é excelente e para demonstrá-lo utilizamos os mesmos parâmetros magnéticos do Exemplo 1, ligeiramente alterados pela mudança do parâmetro $D = 8750$ MHz. Nesta situação existe uma transição em campo nulo.

A solução da Equação 3.26, tomando-se $\Delta_1 = 100$ MHz, é estável e a aplicação do FDM à função de correlação encontrada fornece as 10 transições corretas, mostradas na Figura 3.5. É claro que a transição de campo nulo não pode ser calculada por este método, mas ela pode ser facilmente determinada a priori. O erro entre a energia da radiação e as correspondentes separações entre níveis é da ordem de alguns ppm, no pior caso.

Para sermos sinceros, devemos dizer que este problema, aqui aparentemente resolvido, não é tão simples assim. O motivo pelo qual ele se tornou

Figura 3.5: Diagrama de níveis de energia e transições calculadas para o sistema de spins $S = 1$ e $I = 1/2$ do Exemplo 3. A transição em campo nulo não está mostrada.

simples é que, para o conjunto de parâmetros magnéticos utilizados, os autovalores da Equação 3.10 são todos reais, tanto quando a frequência da radiação vale Ω_1 , como quando ela vale Ω_2 , apesar da matriz A não ser definida-positiva no segundo caso.

Quando este fato casual não ocorre, encontraremos os problemas que serão discutidos na próxima seção.

3.6 Exemplo 4: o caso não-definido-positivo

Nesta seção iremos discutir alguns problemas que limitam a aplicação do método de solução do problema dos autocampos que foi proposto.

O método utilizado para resolver-se a equação mestra no espaço n^2 nos mostrou que, se a matriz A não é definida-positiva, alguns dos autocampos podem resultar complexos. Este caso ocorre somente quando $(\Omega - F_i + F_j)$ resultar negativo para algum par (i, j) .

A realidade é que, sendo que o novo método proposto é resolvido no espaço n , através das soluções de um sistema de equações diferenciais, a existência de autocampos complexos causa sérias dificuldades.

Quando algum dos autocampos é complexo, a solução geral para $L(x)$, dada pela Equação 3.20, representará funções harmônicas complexas, cujas amplitudes podem decair, ou crescer, conforme x aumenta. Isto ocorre porque algumas das frequências, λ_k , se tornam números complexos.

Entretanto, nada impede que o sistema de equações diferenciais possa também ser resolvido nesta situação. Para exemplificar, consideraremos um exemplo numérico dado pelos seguintes parâmetros: $S = 3/2$, $I = 0$, $g_{xx} = 2.0$, $g_{yy} = 1.9$, $g_{zz} = 2.1$, $D = 8750$ MHz, $E = 500$ MHz, $\theta = 35^\circ$, $\varphi = 55^\circ$ e $\Omega = 7000$ MHz.

A Equação 3.26 foi resolvida para este exemplo, no intervalo 0-40000 G, tomando-se $N = 4096$ e $\delta = 9.768$ G, e o resultado está mostrado na Figura 3.6(a). Nota-se neste gráfico que as curvas estão interrompidas em torno de $x = 20000$.

Isto ocorre porque o controle da qualidade da integração de 3.26, feito através do monitoramento do traço das matrizes L , avisou, que neste ponto, o cálculo estava sendo efetuado com erros significativos e interrompeu o processo. Esta instabilidade numérica ocorre porque, como se pode notar no

Figura 3.6: Funções de correlação calculadas para o sistema de spins $S = 1$ e $I = 0$ do Exemplo 4. (a) Solução da Equação 3.26. (b) Solução da Equação 3.32, para $d = 1000$ G. (c) Função de correlação resultante.

gráfico, os elementos de matriz de L crescem exponencialmente, atingindo os limites de *roundoff* e *overflow*.

As curvas que se notam na Figura 3.6(a) são somente aquelas correspondentes às componentes harmônicas moduladas por exponenciais crescentes. As componentes que estamos interessados, aquelas que correspondem aos verdadeiros autocampos, ou frequências λ_k reais, não se notam porque elas se mantêm estacionárias com amplitudes da ordem da unidade.

Este é o problema! Como fazer para ter-se precisão e resolução, analisando-

se dados obscurecidos por outros, quando a relação entre os últimos e os primeiros é da ordem de 10^6 , ou maior?

Ainda não resolvemos este problema de uma forma definitiva, para casos mais gerais, mas, estamos a caminho. No caso particular deste exemplo, o problema pode ser resolvido de uma maneira bem simples.

Em primeiro lugar, re-escrevemos a Equação 3.26 de uma maneira um pouco diferente:

$$\frac{d}{dx} L_{ij} = \frac{2\pi i}{\Omega - F_i + F_j} \sum_k (G_{ik} L_{kj} - G_{kj} L_{ik}) - d L_{ij}, \quad (3.32)$$

onde d é um fator de amortecimento. O termo que contém d não afeta a parte real das frequências das componentes harmônicas contidas na solução da equação mas introduz um amortecimento exponencial em todas elas. Ajustando-se o valor de d pode-se controlar o crescimento daquelas componentes moduladas por exponenciais crescentes, sem contudo perder precisão numérica.

O resultado deste cálculo está mostrado na Figura 3.6(b), para $d = 1000$ G. Nota-se que foi possível integrar-se a equação até o limite pré-estabelecido, mantendo-se $C(x)$ próximo da unidade. O problema é que as componentes associadas aos autocampos que nos interessam ficam fortemente atenuadas exponencialmente, tal que para $x \approx 10d$, elas praticamente se extinguem!

Este fato limita a resolução do FDM, na determinação dos autocampos. Todavia, no exemplo em estudo esta limitação não é restritiva. A Figura 3.6(c) ilustra os resultados obtidos pela aplicação do FDM aos dados de 3.6(b).

Sendo que a constante de decaimento das componentes que nos interessa é conhecida a priori (isto é, ela vale d), estas podem ser filtradas, eliminando-se

aquelas componentes que não nos interessam. O FDM é capaz de recuperar o sinal correspondente às componentes estacionárias, mostradas na Figura 3.6(c), mas, devido ao amortecimento imposto, elas decaem rapidamente.

Figura 3.7: Diagrama de níveis de energia e transições calculadas para o sistema de spins $S = 1$ e $I = 0$ do Exemplo 4.

Os autocampos calculados pelo FDM neste exemplo continuam sendo precisos, apesar do amortecimento imposto, em 1 ppm, conforme ilustrado na Figura 3.7.

Este procedimento funciona bem no caso de problemas com baixas dimensão. Quando a dimensão, n , é grande somente pode-se alcançar resolução suficiente aumentando-se o produto $N\delta$. Neste caso, o amortecimento exagerado impede que o aumento deste parâmetro implique em aumento de

resolução.

Por este motivo, se faz necessário o desenvolvimento de métodos mais eficazes para resolver-se este problema.

3.7 Conclusões

O novo método apresentado tem o objetivo de determinar todos os campos de ressonância de um dado Hamiltoniano de spin, linear com o campo magnético aplicado. Sua formulação não é baseada em hipóteses restritivas, o que o torna um método que pode ser aplicado com bastante generalidade.

O método é capaz de determinar precisamente, em um único procedimento, todos os campos de ressonância, sem a necessidade de resolver equações de autovalores, ou algoritmos sofisticados.

Mostramos aqui apenas alguns exemplos, escolhidos com o objetivo de testar a precisão do método, mas, estudamos muitos outros casos que não foram mostrados aqui. Em todos eles, o método foi capaz de calcular os autocampos com uma precisão da ordem de alguns ppm.

Devido a uma dificuldade matemática, que acreditamos ser transponível, o método está sujeito a limitações quando a frequência da radiação aplicada não é maior que os desdobramentos de campo nulo. Mas, em todos os outros casos, o método é robusto e tem potencial para se tornar uma ferramenta de rotina na área.

Sendo que até o momento estivemos prioritariamente preocupados com o estabelecimento do algoritmo e com a precisão dos resultados, ainda não otimizamos as rotinas de cálculo numérico, a fim que pudéssemos comparar a eficiência do método proposto, com aquela dos outros pacotes computacionais disponíveis na comunidade.

Todavia, esperamos que ao sugerir um novo método de ataque ao problema, estejamos contribuindo para o aumento do conhecimento e da versatilidade dos métodos de simulação de espectros na área da RPE de onda contínua.

Apêndice A

Álgebra linear

Nossa intenção neste apêndice não é escrever um texto sobre a Álgebra Linear mas sim, apenas, discutir com um pouco mais de detalhes alguns assuntos que são especificamente importantes para o nosso trabalho. Não temos a pretensão de sermos matematicamente rigorosos, nem tampouco preocupados com a demonstração de todas as afirmações.

O objetivo é ajudar os leitores, principalmente os estudantes, que estão pela primeira vez entrando contato com os assuntos abordados nesta tese. Neste caso, este apêndice serve apenas como uma pequena revisão dos conceitos que consideramos imprescindíveis ao entendimento do nosso trabalho. Para maiores detalhes deve-se consultar os livros textos tradicionais da matéria.

A.1 Definições

Uma matriz é dita simétrica se ela é igual à sua transposta:

$$A = A^T \quad \text{ou} \quad A_{ij} = A_{ji} \quad (\text{A.1})$$

O adjunto de um operador $\hat{Q}$ é o operador $\hat{Q}^H$ que satisfaz:

$$\langle q|\hat{Q}|p\rangle = \langle p|\hat{Q}^H|q\rangle^* \quad (\text{A.2})$$

No caso em que o operador $\hat{Q}$ é auto-adjunto ou hermitiano:

$$\langle q|\hat{Q}|p\rangle = \langle p|\hat{Q}|q\rangle^* \quad (\text{A.3})$$

A matriz adjunta, ou hermitiana conjugada, de A , A^H , é igual ao complexo-conjugado de sua transposta:

$$A^H = (A^T)^* = (A^*)^T \quad \text{ou} \quad [A^H]_{ij} = A_{ji}^* \quad (\text{A.4})$$

Uma matriz é dita auto-adjunta, ou hermitiana, se ela for igual à sua adjunta:

$$A^H = A \quad \text{ou} \quad [A^H]_{ij} = A_{ij} = A_{ji}^* \quad (\text{A.5})$$

Uma matriz real é dita ortogonal se a sua transposta é a sua inversa:

$$A^T A = A A^T = I \quad (\text{A.6})$$

Uma matriz A é unitária se a sua transposta-conjugada é a sua inversa:

$$A^H A = A A^H = I \quad (\text{A.7})$$

Uma matriz A é dita normal se ela comuta com a sua adjunta:

$$A^H A = A A^H \quad (\text{A.8})$$

Quando uma matriz é hermitiana seus autovalores são reais. Por outro lado, os autovetores não degenerados de uma matriz normal são ortogonais

entre si e formam uma base completa. Quando a matriz não é normal, seus autovetores não são ortogonais (exceto casualmente) mas, usualmente eles formam uma base completa (mas, não sempre).

A.2 O produto interno

A mecânica quântica convencional considera somente operadores (usualmente Hermitianos) atuando no espaço de Hilbert. Este espaço é caracterizado por um **produto interno hermitiano**, isto é, para duas funções $|\alpha\rangle$ e $|\beta\rangle$ define-se o produto interno Hermitiano por:

$$\langle\alpha|\beta\rangle = \langle\beta|\alpha\rangle^*, \quad (\text{A.9})$$

onde o asterisco representa o complexo conjugado. Esta definição tem muitas conseqüências úteis, por exemplo, para uma função $|\alpha\rangle$ pode-se definir a norma, $\langle\alpha|\alpha\rangle$, que é sempre um número real positivo e, portanto, os autovalores de um operador Hermitiano serão sempre números reais.

Entretanto, estas propriedades não são sempre úteis para descrever-se o comportamento de sistemas dissipativos, para os quais costuma-se definir o **produto interno complexo-simétrico** em um espaço não-Hilbert:

$$(\alpha|\beta) = (\beta|\alpha). \quad (\text{A.10})$$

Aqui, utilizamos parênteses a fim de diferenciar este tipo de produto interno do anterior. Note que, com esta nova definição, $(\alpha|\alpha)$ não é necessariamente real, o que significa que a norma não está necessariamente definida neste espaço abstrato não-Hilbert. Em particular, e antecipando que a afirmação seguinte não representa uma situação corriqueira e desejável, $(\alpha|\alpha)$ pode até ser nulo mesmo que $|\alpha\rangle$ seja não-nulo.

A.3 Mudança de base

A escolha de uma base para representar um operador em um dado espaço não é única e, dadas duas bases que representam um mesmo operador, a nossa intenção será determinar qual a relação que existe entre elas. Tendo em vista que, no nosso trabalho, as duas definições anteriores para o produto interno serão utilizadas, vamos discutir a questão da mudança de base considerando cada uma destas duas situações separadamente.

Utilizando-se destes conceitos, iremos em seguida analisar as diferentes formas de decomposição de matrizes que serão utilizadas no nosso trabalho.

A.3.1 Mudança de base no espaço de Hilbert

Sejam duas bases ortonormais completas $\{|\alpha\rangle\}$ e $\{|\beta\rangle\}$. A ortonormalidade das bases implica que:

$$\langle\alpha_i|\alpha_j\rangle = \delta_{ij} \quad , \quad \sum_i |\alpha_i\rangle\langle\alpha_i| = \hat{I}, \quad (\text{A.11})$$

e

$$\langle\beta_i|\beta_j\rangle = \delta_{ij} \quad , \quad \sum_i |\beta_i\rangle\langle\beta_i| = \hat{I}, \quad (\text{A.12})$$

com $i, j = 1 \dots n$. Sendo que as bases são completas, podemos expressar qualquer *ket* $|\beta_i\rangle$ como uma combinação linear dos *kets* da base $\{|\alpha_i\rangle\}$ (e também vice-versa). Isto é,

$$|\beta_i\rangle = \hat{I}|\beta_i\rangle = \sum_j |\alpha_j\rangle\langle\alpha_j|\beta_i\rangle = \sum_j |\alpha_j\rangle U_{ji}, \quad (\text{A.13})$$

onde, os elementos da matriz transformação U são definidos por:

$$[U]_{ij} \equiv U_{ij} = \langle \alpha_i | \beta_j \rangle. \quad (\text{A.14})$$

Na notação matricial a Equação A.13 se escreve como:

$$\beta = \alpha U, \quad (\text{A.15})$$

onde α e β são vetores linha $(1, n)$ e U é uma matriz quadrada (n, n) .

Realizando a transformação na direção oposta temos:

$$|\alpha_i\rangle = \hat{I}|\alpha_i\rangle = \sum_j |\beta_j\rangle \langle \beta_j | \alpha_i \rangle = \sum_j |\beta_j\rangle [U^*]_{ij} = \sum_j |\beta_j\rangle [U^H]_{ji}, \quad (\text{A.16})$$

onde os símbolos U^* e U^H denotam a matriz conjugada de U e a matriz transposta-conjugada (ou adjunta) de U , respectivamente.

A Equação A.16 escrita na notação matricial fica:

$$\alpha = \beta U^H. \quad (\text{A.17})$$

Multiplicando-se esta equação à direita por U e usando a Equação A.15 obtemos:

$$U^H U = I. \quad (\text{A.18})$$

Analogamente, multiplicando-se a Equação A.15 à direita por U^H e usando a Equação A.17 obtemos:

$$U U^H = I. \quad (\text{A.19})$$

Portanto, as duas bases estão relacionadas entre si por uma **transformação unitária**, o que é uma consequência direta da ortonormalidade das bases.

Vamos agora determinar como estão relacionadas entre si as representações matriciais de um certo operador $\hat{O}$ nas duas bases ortonormais. Seja A a representação de $\hat{O}$ na base $\{|\alpha_i\rangle\}$ e B a representação de $\hat{O}$ na base $\{|\beta_i\rangle\}$:

$$\hat{O}|\alpha_i\rangle = \sum_j |\alpha_j\rangle \langle \alpha_j | \hat{O} | \alpha_i \rangle = \sum_j |\alpha_j\rangle A_{ji}, \quad (\text{A.20})$$

$$\hat{O}|\beta_i\rangle = \sum_j |\beta_j\rangle \langle \beta_j | \hat{O} | \beta_i \rangle = \sum_j |\beta_j\rangle B_{ji}. \quad (\text{A.21})$$

Para encontrar a relação entre A e B utilizamos a técnica familiar de introduzir o operador identidade:

$$\begin{aligned} B_{ij} &= \langle \beta_i | \hat{O} | \beta_j \rangle = \langle \beta_i | 1 \hat{O} 1 | \beta_j \rangle \\ &= \sum_{kl} \langle \beta_i | \alpha_k \rangle \langle \alpha_k | \hat{O} | \alpha_l \rangle \langle \alpha_l | \beta_j \rangle = \sum_{kl} [U^H]_{ik} A_{kl} U_{lj} \\ &= \sum_k [U^H]_{ik} [AU]_{kj} = [U^H AU]_{ij}. \end{aligned} \quad (\text{A.22})$$

Logo,

$$B = U^H AU. \quad (\text{A.23})$$

Multiplicando-se à esquerda por U e à direita por U^H encontra-se a relação inversa:

$$A = UBU^H. \quad (\text{A.24})$$

Estas duas últimas relações mostram que as representações A e B estão relacionadas entre si por uma transformação unitária. Um fato importante a respeito de tal transformação é que para qualquer operador Hermitiano cuja

representação na base $\{|\alpha_i\rangle\}$ seja não-diagonal, é sempre possível encontrar-se uma outra base $\{|\beta_i\rangle\}$ na qual a representação deste mesmo operador seja diagonal. Este assunto será discutido adiante com mais detalhes.

A demonstração anterior pode ser generalizada para o cálculo dos elementos de matriz de um operador $\hat{O}$ entre duas bases quaisquer. Sejam as três bases ortonormais $\{|\alpha_i\rangle\}$, $\{|\beta_i\rangle\}$ e $\{|\gamma_i\rangle\}$, sendo que as duas últimas são derivadas a partir da primeira pelas transformações:

$$\beta = \alpha U \quad \text{e} \quad \gamma = \alpha V. \quad (\text{A.25})$$

Assim,

$$\begin{aligned} C_{ij} &= \langle \gamma_i | \hat{O} | \beta_j \rangle = \sum_{kl} [V^*]_{li} \langle \alpha_l | \hat{O} | \alpha_k \rangle U_{kj} \\ &= \sum_{kl} [V^H]_{il} A_{lk} U_{kj} = \sum_l [V^H]_{il} [AU]_{lj} = [V^H AU]_{ij}, \end{aligned} \quad (\text{A.26})$$

ou seja,

$$C = V^H AU. \quad (\text{A.27})$$

Sendo que U e V são ambos unitários, a relação inversa fica:

$$A = VCU^H. \quad (\text{A.28})$$

Este tipo de transformação será utilizada adiante no método da decomposição de uma matriz nos seus valores singulares.

A.3.2 Mudança de base no espaço não-Hilbert

Neste caso, vamos assumir que as bases $\{|\alpha_i\rangle\}$ e $\{|\beta_i\rangle\}$ não são patológicas, isto é, que elas são completas, ortonormais e que a identidade continua

definida:

$$(\alpha_i|\alpha_j) = \delta_{ij} \quad , \quad \sum_i |\alpha_i)(\alpha_i| = \hat{I}, \quad (\text{A.29})$$

e

$$(\beta_i|\beta_j) = \delta_{ij} \quad , \quad \sum_i |\beta_i)(\beta_i| = \hat{I}. \quad (\text{A.30})$$

Utilizando-se da definição do produto escalar complexo-simétrico, Equação A.10, é fácil mostrar que:

$$\alpha = \beta U \quad , \quad \beta = \alpha U^T, \quad (\text{A.31})$$

e

$$B = U^T A U \quad , \quad A = U B U^T \quad , \quad C = V^T A U, \quad (\text{A.32})$$

com

$$U^T U = U U^T = I. \quad (\text{A.33})$$

Assim, repete-se os mesmos resultados do caso anterior desde que se faça a troca do símbolo “H” por “T”, isto é, em vez de transpor e conjugar os elementos de matriz simplesmente os transpõe.

A.4 Autovalores e autovetores

Quando um operador $\hat{O}$ atua em um vetor $|x\rangle$, o vetor resultante é, em geral, distinto de $|x\rangle$. Se $\hat{O}|x\rangle$ é simplesmente uma constante vezes $|x\rangle$, isto é,

$$\hat{O} |x\rangle = \lambda_x |x\rangle, \quad (\text{A.34})$$

dizemos que $|x\rangle$ é um autovetor do operador $\hat{O}$ com autovalor λ_x . Sem perda de generalidade, o autovetor pode ser normalizado: $\langle x|x\rangle = 1$.

Em geral, o autovetor $|x\rangle$ pode ser decomposto em uma base completa $\{|\alpha_i\rangle\}$:

$$|x\rangle = \sum_i x_i |\alpha_i\rangle. \quad (\text{A.35})$$

Substituindo A.35 na A.34 e multiplicando à esquerda por $\langle \alpha_j|$ obtemos:

$$\begin{aligned} \sum_i \langle \alpha_j | \hat{O} | \alpha_i \rangle x_i &= \lambda_x \sum_i x_i \langle \alpha_j | \alpha_i \rangle \\ \sum_i A_{ji} x_i &= \lambda_x \sum_i B_{ji} x_i, \end{aligned} \quad (\text{A.36})$$

ou seja:

$$A x = \lambda_x B x, \quad (\text{A.37})$$

onde $A_{ij} = \langle \alpha_i | \hat{O} | \alpha_j \rangle$ e $B_{ij} = \langle \alpha_i | \alpha_j \rangle$.

Esta última equação é conhecida por “**equação de autovalores generalizada**” e é a representação matricial da equação de autovalores (Equação A.34) na base $\{|\alpha_i\rangle\}$. Se a base for ortonormal, então $B_{ij} = \delta_{ij}$ e a equação de autovalores generalizada se reduz a uma **equação de autovalores simples**:

$$A x = \lambda_x x. \quad (\text{A.38})$$

Se a dimensão da base é n , A é uma matriz quadrada (n, n) e x é uma matriz coluna $(n, 1)$ que contem as componentes de $|x\rangle$ na base $\{|\alpha_i\rangle\}$. Esta última equação representa um conjunto de n equações lineares homogêneas

e, através do seu polinômio característico, n valores possíveis de λ_x poderão ser determinados, os quais denotaremos por λ_i com $i = 1 \dots n$. Em geral, autovalores podem ser nulos, reais ou complexos, e este último caso pode ocorrer mesmo quando a matriz A é real. Além disso, os autovalores podem ter multiplicidade menor ou igual a n . Como podem existir até n autovalores distintos, e como cada autovalor corresponde a um único autovetor, costuma-se escrever a Equação A.38 como:

$$A x_i = \lambda_i x_i \quad , \quad i = 1, 2, \dots n. \quad (\text{A.39})$$

Definindo $\lambda = \text{diag}(\lambda_i)$ por uma matriz quadrada (n, n) cujos elementos fora da diagonal são nulos e os da diagonal iguais aos autovalores λ_i , a Equação A.39 pode ser escrita numa forma mais compacta:

$$A X = X \lambda, \quad (\text{A.40})$$

onde X é uma matriz quadrada na qual a coluna j é o autovetor correspondente ao autovalor λ_j , isto é, X_{ij} é a componente i do autovetor correspondente ao autovalor λ_j . Explicitamente:

$$\begin{aligned} & \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{bmatrix} \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nn} \end{bmatrix} = \\ & = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nn} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}. \quad (\text{A.41}) \end{aligned}$$

A.5 Autovalores à esquerda e diagonalização de matrizes

A matriz A definida anteriormente por $A_{ij} = \langle \alpha_i | \hat{O} | \alpha_j \rangle$ é a representação do operador $\hat{O}$ na base $\{|\alpha_i\rangle\}$. Esta matriz não é necessariamente diagonal e, por isso, é interessante procurar-se por uma outra base $\{|\beta_i\rangle\}$ na qual o operador representa-se por uma matriz, B , diagonal, isto é, $B_{ij} = \lambda_i \delta_{ij}$. Em outras palavras, procura-se diagonalizar a matriz A .

Na seção anterior, os autovetores eram vetores coluna multiplicados à direita da matriz A . Estes autovetores são denominados de “autovetores à direita” e os correspondentes autovalores de “autovalores à direita”. Entretanto, é possível também tentar-se encontrar autovetores linha que multiplicam-se à esquerda da matriz A e satisfazem:

$$yA = \lambda y \quad (\text{A.42})$$

Estes são os chamados “autovetores à esquerda” e os correspondentes autovalores de “autovalores à esquerda”. Com relação a esta nova definição, pode-se provar o seguinte: (i) os autovalores à esquerda e à direita de A são idênticos. (ii) Se a matriz A é hermitiana, os autovetores à esquerda são os hermitianos conjugados, ou transposto-conjugados, dos da direita. (iii) Se os autovalores são não-degenerados, cada autovetor à esquerda é ortogonal a todos os autovetores à direita, exceto com o seu correspondente de mesmo autovalor, e, vice-versa.

Da mesma maneira que fizemos na seção anterior:

$$\begin{aligned} A X &= X \operatorname{diag}(\lambda_1 \dots \lambda_n) = X \lambda \\ Y A &= \operatorname{diag}(\lambda_1 \dots \lambda_n) Y = \lambda Y \end{aligned} \quad (\text{A.43})$$

Pode-se mostrar também que as matrizes X e Y são inversas uma da outra. Portanto,

$$Y A X = X^{-1} A X = \text{diag}(\lambda_1 \dots \lambda_n) = \lambda \quad (\text{A.44})$$

Assim, a matriz de autovalores, λ , pode ser obtida por uma transformação de similaridade sobre a matriz A . Isto significa que toda matriz com um conjunto completo de autovetores, o que inclui todas as matrizes normais (e algumas não-normais) pode ser diagonalizada por uma transformação de similaridade, isto é, as colunas da matriz de transformação que afetam a diagonalização são os autovetores à direita, e, que as linhas de sua inversa são os autovetores à esquerda.

Se a matriz A for hermitiana então $X^{-1} = X^H$ e a transformação é dita unitária. Portanto, o problema de diagonalizar uma matriz hermitiana A é equivalente ao de encontrar-se uma matriz unitária X que converte A em uma matriz diagonal:

$$X^H A X = \lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}. \quad (\text{A.45})$$

A matriz X é denominada de “matriz dos autovetores” e a matriz λ é denominada de “matriz dos autovalores”. Invertendo-se a Equação A.45 obtemos:

$$A = X \lambda X^H. \quad (\text{A.46})$$

Esta equação define o método denominado “**decomposição de uma matriz nos seus autovalores**”, ou em inglês, *eigenvalue decomposition of*

a *matrix*.

Não é difícil mostrar que se as bases estão definidas em um espaço não-Hilbert, onde o produto escalar se exprime pelo produto interno complexo-simétrico (ver a primeira seção deste apêndice) a decomposição da matriz A simétrica se faz pela transformação:

$$A = X \lambda X^T, \quad (\text{A.47})$$

com X satisfazendo $XX^T = X^T X = I$.

A.6 Ortonormalização da base

Algumas bases utilizadas no nosso trabalho, como por exemplo a base de Krilov, são normalizadas e formam um conjunto completo, mas não são ortogonais. Por isso, e considerando o que foi discutido anteriormente (ver Equação A.37), as equações envolvendo autovalores não podem neste caso serem colocadas na forma usual da Equação A.40. Em geral as equações resultam em equações denominadas “equações de autovalores generalizadas” e possuem a forma $AX = BX\lambda$. A fim de simplificar-se a solução destes problemas procura-se por um método capaz de ortogonalizar a base. O procedimento aqui utilizado está baseado no texto indicado em [32].

Se o conjunto de funções $\{|\alpha_i\rangle\}$ não é ortogonal, isto é,

$$\langle \alpha_i | \alpha_j \rangle = B_{ij}, \quad (\text{A.48})$$

então, sempre será possível encontrar-se uma matriz de transformação V (não unitária) tal que o conjunto transformado $\{|\beta_i\rangle\}$, dado por

$$|\beta_i\rangle = \sum_j V_{ji} |\alpha_j\rangle, \quad (\text{A.49})$$

forme um conjunto ortonormal, isto é,

$$\langle \beta_i | \beta_j \rangle = \delta_{ij}. \quad (\text{A.50})$$

A fim de obter-se as propriedades de V substituímos a Equação A.49 na Equação A.50 para obter:

$$\begin{aligned} \langle \beta_i | \beta_j \rangle &= \sum_{kl} [V^*]_{li} \langle \alpha_l | \alpha_k \rangle V_{kj} = \sum_{kl} [V^H]_{il} B_{lk} V_{kj} \\ &= \sum_l [V^H]_{il} [BV]_{lj} = [V^H BV]_{ij} = \delta_{ij}. \end{aligned} \quad (\text{A.51})$$

Esta equação pode então ser escrita na forma matricial

$$V^H B V = I, \quad (\text{A.52})$$

que define a relação que a matriz V deve satisfazer se a base transformada forma um conjunto ortonormal. Conforme veremos adiante, será exigido que V seja não-singular, isto é, que exista a inversa V^{-1} .

A matriz B é Hermitiana e pode ser diagonalizada por uma transformação unitária:

$$U^H B U = b, \quad (\text{A.53})$$

onde b é a matriz diagonal dos autovalores de B .

Uma maneira de se obter uma base ortonormal é através do método chamado **ortonormalização canônica**. Ele usa uma transformação definida por:

$$V = U b^{-1/2} \quad \text{ou} \quad V_{ij} = \frac{U_{ij}}{b_j^{1/2}}. \quad (\text{A.54})$$

Substituindo esta definição na Equação A.52 resulta:

$$V^H B V = (U b^{-1/2})^H B (U b^{-1/2}) = b^{-1/2} (U^H B U) b^{-1/2} = b^{-1/2} b b^{-1/2} = I, \quad (\text{A.55})$$

o que mostra que A.54 é de fato uma transformação que ortogonaliza a base.

Um problema numérico com este tipo de transformação pode ocorrer quando algum dos autovalores de B se aproxima de zero (ou que o módulo do complexo b_i é muito pequeno), pois neste caso a transformação A.54 irá produzir uma instabilidade numérica.

Neste caso, é sempre possível, por inspeção dos n autovalores encontrados, e ordenando-os em ordem decrescente ($|b_1| > |b_2| \dots > |b_n|$), decidir-se quantos m autovalores são muito pequenos e então utilizar-se de uma matriz transformação truncada:

$$\tilde{V} = \begin{bmatrix} U_{1,1}/b_1^{1/2} & U_{1,2}/b_2^{1/2} & \cdots & U_{1,n-m}/b_{n-m}^{1/2} \\ U_{2,1}/b_1^{1/2} & U_{2,2}/b_2^{1/2} & \cdots & U_{2,n-m}/b_{n-m}^{1/2} \\ \vdots & \vdots & \ddots & \vdots \\ U_{n,1}/b_1^{1/2} & U_{n,2}/b_2^{1/2} & \cdots & U_{n,n-m}/b_{n-m}^{1/2} \end{bmatrix}, \quad (\text{A.56})$$

onde as últimas m colunas de V foram eliminadas para dar origem à matriz truncada $\tilde{V}$ de dimensão $(n, n-m)$. Este assunto está discutido com outros detalhes no Capítulo que discute o método FDM.

Para mostrar a utilidade deste método vamos considerar o problema da equação de autovalores generalizada mencionada no início desta seção:

$$A X = B X \lambda, \quad (\text{A.57})$$

onde A , B , X e λ são matrizes (n, n) , X é a matriz dos autovetores e λ a matriz dos autovalores. Se A ou B for não-singular então esta equação pode ser reduzida a uma equação de autovalores simples,

$$X = (A^{-1}B) X \lambda \quad \text{ou} \quad (B^{-1}A) X = X \lambda, \quad (\text{A.58})$$

entretanto, esta não é uma situação que ocorre freqüentemente.

Quando A e B são singulares, o problema generalizado pode ser reduzido ao simples através do método da ortonormalização discutido anteriormente. Para isso é primeiro necessário encontrar-se uma base de funções na qual a matriz B é diagonal, o que é feito primeiro através da decomposição de B nos seus autovalores, $B = UbU^H$, onde b é a matriz diagonal com os autovalores de B , e U é a matriz dos autovetores de B . Na base em que B é diagonal, a matriz A se transforma em $a = UAU^H$, com a relação inversa dada por $A = U^H a U$. Note que a não é necessariamente diagonal (a será diagonal somente quando A e B comutarem entre si).

Utilizando a decomposição de B a Equação A.57 se transforma em:

$$\begin{aligned} A X &= (UbU^H) X \lambda \\ A X &= (Ub^{1/2})(b^{1/2}U^H X) \lambda \\ (b^{-1/2}U^H A U b^{-1/2})(b^{1/2}U^H X) &= (b^{1/2}U^H X) \lambda \\ (V^H A V)(V^{-1}X) &= (V^{-1}X) \lambda \\ \bar{A} \bar{X} &= \bar{X} \lambda. \end{aligned} \quad (\text{A.59})$$

Portanto, chega-se em uma equação de autovalores simples na qual os novos autovetores se exprimem como $\bar{X} = (b^{1/2}U^H)X = V^{-1}X$ e a nova matriz A se transforma em $\bar{A} = (b^{-1/2}U^H)A(Ub^{-1/2}) = V^H A V$.

Assim, o problema de autovalores generalizado pode ser reduzido ao simples através de uma transformação canônica da base. Denominamos este método por **EVD**, sigla inspirada no termo inglês *EigenValue Decomposition*.

A.7 Decomposição nos valores singulares

Seja uma matriz A que é uma representação do operador $\hat{O}$ na base $\{|\alpha_i\rangle\}$ de tal forma que $A_{ij} = \langle\alpha_i|\hat{O}|\alpha_j\rangle$. Vimos anteriormente (Equação A.27) que, dadas duas transformações unitárias R e L , representações de $\hat{O}$ nas bases $\{|\beta_i\rangle\}$ e $\{|\gamma_i\rangle\}$, respectivamente, o operador pode ser decomposto da seguinte forma:

$$A = L D R^H, \quad (\text{A.60})$$

com

$$D_{ij} = \langle\gamma_i|\hat{O}|\beta_j\rangle, \quad LL^H = L^H L = I \quad \text{e} \quad RR^H = R^H R = I. \quad (\text{A.61})$$

D é uma nova representação de $\hat{O}$ neste produto misto entre as duas novas bases. As letras L e R estão sendo aqui utilizadas para facilitar a visualização e memorização das equações posto que L é a inicial de *left* e R é a inicial de *right*; posições estas que as respectivas matrizes ocupam na equação anterior. Um caso particular de tal transformação, e de extrema importância, é quando a matriz D é uma matriz diagonal (o que explica porque utilizamos a letra D) na qual os elementos da diagonal são números reais positivos, isto é, $D = \text{diag}(d_i)$ com $d_i \geq 0$. Neste caso, os elementos da diagonal de D são chamados de **valores singulares** da matriz A . Normalmente os valores singulares são ordenados em ordem decrescente, $d_1 > d_2 \dots > d_n$, de tal forma que o maior deles ocupa o canto superior esquerdo da matriz D .

Quando a matriz A é singular, alguns dos valores singulares serão nulos e o número de valores singulares não-nulos determina o *rank*, r , da matriz A (em geral, $r \leq n$) e o número de valores singulares nulos determina a dimensão do “espaço nulo” da matriz A .

Embora não seja uma condição necessária, somente consideraremos aqui matrizes quadradas (n, n) . Em geral, pode ser demonstrado que qualquer matriz retangular (n, m) pode também ser decomposta nos seus valores singulares. Para uma descrição mais detalhada do método e de suas aplicações, é interessante consultar os livros textos de Álgebra linear. O nosso interesse aqui é apresentar o método da decomposição nos valores singulares como uma solução alternativa do problema da equação de autovalores generalizada, $AX = BX\lambda$, que foi introduzido na seção anterior. Para isso, primeiro decompos $B = LDR^H$ e em seguida realizamos o seguinte procedimento algébrico:

$$\begin{aligned}
 AX &= (LDR^H) X \lambda \\
 AX &= (LD^{1/2}) (D^{1/2}R^H X) \lambda \\
 (D^{-1/2}L^H A R D^{-1/2}) (D^{1/2}R^H X) &= (D^{1/2}R^H X) \lambda \\
 \bar{A} \bar{X} &= \bar{X} \lambda.
 \end{aligned} \tag{A.62}$$

Novamente, chega-se em uma equação de autovalores simples na qual os novos autovetores se exprimem como $\bar{X} = (D^{1/2}R^H)X$ e a nova matriz A se transforma em $\bar{A} = (D^{-1/2}L^H)A(RD^{-1/2})$.

Vemos que este procedimento é bastante similar ao da seção anterior, onde resolveu-se o mesmo problema utilizando-se da decomposição nos autovalores, técnica que foi denominada de EVD. Por analogia, denominamos este novo método por **SVD**, a partir da nomenclatura em língua inglesa *Singular Value Decomposition*.

Do ponto de vista numérico, a técnica SVD é mais eficiente que a EVD por um motivo bastante sutil. Para entender isso, lembremo-nos que foi mencionado na seção anterior que as matrizes A e B são, em grande parte das situ-

ações, patológicas (tecnicamente, diz-se que as matrizes são *ill-conditioned*). Em outras palavras, estas matrizes são singulares, ou quase-singulares, sendo que este último caso representa uma dificuldade numérica ainda mais grave que a do primeiro.

Quando as matrizes A e B são singulares, ou quase-singulares, é sempre possível, por inspeção, decidir-se quantos m valores singulares são muito pequenos a ponto de gerarem instabilidades numéricas e, a partir daí realizar-se o truncamento adequado das matrizes de transformação L e R de forma similar à realizada na seção anterior. Aqui, a nossa experiência adquirida ao tratar muitos exemplos práticos nos diz que esta “inspeção” é mais simples e mais precisa quando se trata de selecionar valores singulares nulos ou quase-nulos, do que selecionar-se autovalores (complexos) de módulo nulo ou quase-nulo. Particularmente, quando os elementos da matriz estão afetados por ruído, a matriz deixa de ser (exatamente) singular para se tornar quase-singular. Em muitos casos práticos, o efeito deste processo faz com que os autovalores nulos deixem de ser nulos e dificilmente distinguíveis daqueles não-nulos da matriz original sem ruído. Por outro lado, os valores singulares são mais sensíveis a esta “imperfeição” da matriz, o que propicia um truncamento mais correto das matrizes de transformação.

Bibliografia

- [1] A. Abragam and B. Bleaney. *Electron paramagnetic resonance of transition ions*. Dove, 1986. p. 337.
- [2] G. G. Belford, R. L. Belford, and J. F. Burkhalter. Eigenfields: a practical direct calculation of resonance fields and intensities for field-swept fixed-frequency spectrometers. *J. Magn. Reson.*, 11:251–265, 1973.
- [3] R. L. Belford and M. J. Nilges. Computer simulation of powder spectra. *EPR Symposium, 21st Rock Mountain Conference, Denver, Colorado*, 1979.
- [4] M. Benzi, G. H. Golub, and J. Liesen. Numerical solution of saddle point problems. *Acta Numerica*, pages 1–137, 2005.
- [5] C. F. Bracciali, X. Li, and A. S. Ranga. Real orthogonal polynomials in frequency analysis. *Mathematics of computation*, 74:341–362, 2004.
- [6] J. Chen. *Nonlinear Methods for High Resolution Spectral Analysis and Their Applications in Nuclear Magnetic Resonance Experiments*. PhD thesis, University of California, Irvine, 2002.
- [7] J. Chen, V. A. Mandelshtan, and A. J. Shaka. Regularization of the two-dimensional filter diagonalization method: FDM2K. *J. Magn. Res.*, 146:363–368, 2000.

- [8] R. Chen and H. Guo. Efficient calculation of matrix elements in low storage filter diagonalization. *J. Chem. Phys.*, 111:464–471, 1999.
- [9] J. Demmel and B. Kagstrom. The generalized schur decomposition of an arbitrary pencil $a - zb$: robust software with error bounds and applications. Part I: Theory and algorithms. *ACM Trans. Math. Softw.*, 19:160–174, 1993.
- [10] J. Demmel and B. Kagstrom. The generalized schur decomposition of an arbitrary pencil $a - zb$: robust software with error bounds and applications. Part II: Software and applications. *ACM Trans. Math. Softw.*, 19:175–201, 1993.
- [11] S. Van Doorslaer, G. A. Sierra, and A. Schweiger. Dead time-dependent line distortions in absolute-value electron spin echo envelope modulation spectra. *J. Magn. Reson.*, 136:152–158, 1999.
- [12] B. J. Gaffney and H. J. Siverstone. Simulation methods for looping transitions. *Journal of Magnetic Resonance*, 134:57–66, 1998.
- [13] K. E. Gates, M. Griffin, G. H. Hanson, and K. Burrage. Computer simulation of magnetic resonance spectra employing homotopy. *J. Magn. Reson.*, 135:103–112, 1998.
- [14] M. Griffin, A. Muys, C. Noble, D. Wang, C. Eldershaw, K. E. Gates, K. Burrage, and G. R. Hanson. Xsophe, a computer simulation software suite for the analysis of electron paramagnetic resonance spectra. *Mol. Phys. Rep.*, 26:60–84, 1999.
- [15] A. A. Istratov and O. F. Vyvenko. Exponential analysis in physical phenomena. *Review of Scientific Instruments*, 70:1233–1257, 1999.

- [16] G. Jeschke, V. A. Mandelshtam, and A. J. Shaka. Pure absorption electron spin echo envelope modulation spectra by using de filter diagonalization method for harmonic inversion. *J. Magn. Reson.*, 137:221–230, 1999.
- [17] W. B. Jones, W. J. Thron, O. Njastad, and H. Waadeland. Szegő polynomials applied to frequency analysis. *Journal of Computational and Applied Mathematics*, 46:217–228, 1993.
- [18] C. Lanczos. *Applied Analysis*. Prentice-Hall, Englewood Cliffs, NJ, 1959. p. 272.
- [19] C. J. Magon, J. F. Lima, R. R. Ribeiro, and M. J. Martins. The harmonic inversion of the field-swept fixed-frequency resonance spectrum. *J. Magn. Reson.*, 184:176–183, 2007.
- [20] V. A. Mandelshtam. FDM: the filter diagonalization method for data processing in NMR experiments. *Progr. NMR Spectrosc.*, 38:159–196, 2001.
- [21] V. A. Mandelshtam, T. P. Grozdanov, and H. S. Taylor. Bound states and resonances of the hydroperoxyl radical HO₂. An accurate quantum mechanical calculation using filter diagonalization. *J. Chem. Phys.*, 103:10074–10084, 1995.
- [22] V. A. Mandelshtam and H. S. Taylor. Harmonic inversion of time signals and its applications. *J. Chem. Phys.*, 107:6756–6769, 1997.
- [23] V. A. Mandelshtam and H. S. Taylor. A low storage filter diagonalization method for quantum eigenenergy calculation or for spectral analysis of time signals. *J. Chem. Phys.*, 106:5085–5090, 1997.

- [24] S. K. Misra. Evaluation of spin-hamiltonian parameters from EPR data by the method of least-squares fitting. *J. Magn. Reson.*, 23:403–410, 1976.
- [25] C. B. Moler and G. W. Stewart. An algorithm for generalized matrix eigenvalue problems. *SIAM J. Num. Anal.*, 10:241–256, 1973.
- [26] D. Neuhauser. Bound state eigenfunctions from wave packets - time-energy resolution. *J. Chem. Phys.*, 93:2611–2616, 1990.
- [27] M. J. Nilges. *Electron Paramagnetic Resonance Studies of Low Symmetry Nickel (I) and Molybdenum (V) Complexes*. PhD thesis, University of Illinois, Urbana, IL., 1979.
- [28] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, second edition, 1992.
- [29] M. I. Scullane, L. K. White, and N. D. Chasteen. An efficient approach to computer simulation of EPR spectra of high-spin Fe(III) in rhombic ligand fields. *J. Magn. Reson.*, 47:383–397, 1982.
- [30] S. Stoll. *Spectral simulations in solid-state Electron Paramagnetic Resonance*. PhD thesis, Eidgenössische Technische Hochschule Zürich, 2003.
- [31] S. Stoll and A. Schweiger. Easyspin, a comprehensive software package for spectral simulation and analysis in EPR. *J. of Magn. Reson.*, 178:42–55, 2006.
- [32] S. Szabo and N. S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. McGraw-Hill, New York, first edition, 1982. p. 142-145.

- [33] P. C. Taylor, J. F. Baugher, and H. M. Kritz. Magnetic resonance spectra in polycrystalline solids. *Chem. Rev.*, 75:203–240, 1975.
- [34] M. R. Wall and D. Neuhauser. Extraction, through filter-diagonalization, of general quantum eigenvalues or classical normal mode frequencies from a small number of residues or a short-time segment of a signal. I. Theory and application to a quantum-dynamics model. *J. Chem. Phys.*, 102:8011–8022, 1995.