

**Autovalores de Polinômios  
Matriciais:  
Sensibilidade, Computação e  
Aplicações <sup>1</sup>**

**Fermín S. V. Bazán  
Departamento de Matemática  
Universidade Federal de Santa Catarina  
88040-900 Florianópolis – SC, Brasil  
email: [fermin@mtm.ufsc.br](mailto:fermin@mtm.ufsc.br)**

<sup>1</sup>Notas de Minicurso: 24<sup>o</sup>CBM



# Prefácio

O material apresentado neste texto, visa principalmente, difundir a teoria básica do problema de autovalor polinomial matricial, enfatizando uma série de aplicações em problemas de interesse contemporâneo, bem como chamar atenção às propriedades intrínsecas do problema, relacionadas com a sensibilidade de autovalores a pequenas perturbações nos coeficientes. A motivação para sua elaboração foi que, apesar da teoria sobre polinômios matriciais ser atualmente bem desenvolvida e muito explorada em análise/controlado de sistemas dinâmicos, o problema de sensibilidade é um tópico em desenvolvimento e com poucas publicações a respeito. Assim, além de difundir a teoria fundamental do problema de autovalor polinomial matricial, um dos objetivos do texto é prover resultados sobre sensibilidade de autovalores, apropriados para uma série de problemas nos quais são requeridos apenas uns poucos autovalores e/ou o subespaço invariante associado. Os problemas a serem discutidos envolvem: problemas inversos HR (harmonical Retrieval), problemas de Identificação de Sistemas Dinâmicos MIMO (Vibratórios, por exemplo) e problemas de realização de sistemas. O texto é fortemente influenciado por notas de dois bem sucedidos Minicursos apresentados no XXIII CNMAC (2000) e XXIV CNMAC (2001), escritos em parceria com os Professores M. I. Cardoso e L. H. Bezerra, respectivamente, nos quais foram apresentadas algumas aplicações envolvendo o problema de autovalor para o caso escalar.

O texto está organizado como segue. O Capítulo 1 descreve conceitos gerais sobre polinômios matriciais e a teoria espectral básica de polinômios matriciais mônicos. Ênfase é dada ao caso polinomial quadrático. O material apresentado aqui é de fácil acesso e encontrado nas referências clássicas. O Capítulo 2, destina-se à descrição de aplicações correntes nas quais o problema de autovalor polinomial matricial cumpre um papel fundamental. Especial ênfase é dada a problemas inversos de identificação, mas muitas outras aplicações são comentadas, indicando ao leitor numerosas fontes bibliográficas. A discussão do problema de sensibilidade de autovalores de polinômios matriciais gerais é baseada no conceito de condicionamento e apresentado Capítulo 3. O Capítulo 4, apresenta uma discussão exaustiva do problema de sensibilidade de autovalores de polinômios matriciais provenientes de aplicações em problemas HR. Muitos dos resultados apresen-

tados aqui são inéditos. O Capítulo 5 destina-se à descrição de métodos numéricos para computação de autovalores de matrizes, dando ênfase, em particular, ao problema de calcular um número seletivo de valores singulares e subespaços associados correspondentes, de matrizes provenientes de problemas HR. Métodos numéricos para o problema de autovalor polinomial matricial são em geral, mais sofisticados do que os métodos para o problema de autovalor matricial padrão, e requerem, portanto, certa experiência e treinamento para compreendê-los. Por causa disso, eles são apenas comentados no texto, mas numerosas fontes bibliográficas são indicadas. O texto termina com um apêndice que inclui um resumo de conceitos e resultados de álgebra linear numérica, fortemente utilizados ao longo do texto.

Finalmente, gostaria de expressar meu profundo agradecimento aos colegas, Professores Mário C. Zambaldi e Marcio R. Fernandes, pela colaboração na elaboração deste trabalho. Agradecimentos também vão para o pessoal administrativo do Departamento de Matemática pelo apoio técnico prestado.

Florianópolis SC, Junho 2003

Fermín S. V. Bazán

A meu pai, José Maria, *in memoriam*.



# Conteúdo

<b>1</b>	<b>Generalidades</b>	<b>3</b>
1.1	Introdução . . . . .	3
1.2	Linearização . . . . .	6
1.3	Pares de Jordan . . . . .	10
1.4	Espectro de Polinômios Matriciais Quadráticos . . . . .	15
1.4.1	Solução de Sistemas de Segunda Ordem . . . . .	17
<b>2</b>	<b>Aplicações Correntes</b>	<b>21</b>
2.1	Recuperação de Harmônicos . . . . .	21
2.1.1	Abordagem Polinomial . . . . .	25
2.1.2	Métodos de Subespaço . . . . .	28
2.1.3	Recuperação de Harmônicos em Situações Reais . . . . .	31
2.2	Problemas com Restrições de Igualdade . . . . .	32
2.3	Identificação de Sistemas Vibratórios . . . . .	33
2.3.1	Introdução . . . . .	33
2.3.2	Relações Entrada-Saída e o Problema de Identificação . . . . .	35
2.3.3	Abordagem Polinomial Matricial e de Subespaço . . . . .	38
2.4	Problema de Realização em Sistemas Dinâmicos . . . . .	42
2.4.1	Realização a partir da Resposta ao Impulso Unitário . . . . .	43
2.4.2	Realização a partir de Informação Input-Output . . . . .	45
2.5	Miscelânea . . . . .	46
2.5.1	Solução de EDP's . . . . .	47
2.5.2	Processamento de Sinais . . . . .	48
2.5.3	Aplicações Diversas . . . . .	50
<b>3</b>	<b>Sensibilidade de Autovalores</b>	<b>51</b>
3.1	Noções de Condicionamento . . . . .	51
3.1.1	Raízes de um Polinômio . . . . .	54
3.1.2	Solução de um Sistema de Equações Lineares . . . . .	57
3.1.3	Autovalores de uma Matriz . . . . .	58
3.2	Condicionamento de Autovalores de Polinômios Matriciais . . . . .	59

<b>4</b>	<b>Análise de Sensibilidade e de Erro em problemas HR</b>	<b>65</b>
4.1	Sensibilidade de Autovalores: Caso Polinomial Escalar . . . . .	65
4.1.1	Condicionamento de Matrizes de Vandermonde . . . . .	65
4.1.2	Análise de Sensibilidade: Abordagem da Matriz Companheira . . .	73
4.1.3	Análise de Sensibilidade: Abordagem da Matriz Companheira Pro- jetada . . . . .	74
4.2	Sensibilidade de Autovalores: Caso Polinomial Matricial . . . . .	75
4.2.1	Análise de Sensibilidade: Um Caso Real . . . . .	79
4.3	Análise de erro: Caso Escalar . . . . .	80
4.3.1	Resultados Numéricos . . . . .	85
4.4	Análise de Erro: Caso Polinomial Matricial . . . . .	87
4.4.1	Estimativas para $\ A_{\mathcal{P}} - \tilde{A}_{\mathcal{P}}\ _2$ . . . . .	87
4.4.2	Análise de Erro: Um Caso Real (Mini-Mast) . . . . .	90
<b>5</b>	<b>Computação de Autovalores</b>	<b>93</b>
5.1	Introdução . . . . .	93
5.2	Autovalores Matriciais . . . . .	95
5.2.1	Transformações de Möbius . . . . .	98
5.3	Métodos de Iteração Simultânea . . . . .	98
5.3.1	Métodos de Arnoldi . . . . .	99
5.3.2	Arnoldi com Recomeço . . . . .	100
5.4	Método QR . . . . .	100
5.4.1	Matrizes Simétricas Reais . . . . .	103
5.5	Computação de Valores Singulares Dominantes . . . . .	104
5.5.1	O Procedimento de Rayleigh-Ritz . . . . .	104
5.5.2	Iteração em Subespaços . . . . .	107
5.5.3	Iteração em Subespaços com Aceleração de Ritz . . . . .	108
5.6	Método de Lanczos . . . . .	109
5.6.1	Método de Lanczos Básico . . . . .	109
5.6.2	Método de Lanczos com Reortogonalização Completa . . . . .	112
5.6.3	Lanczos com Semiortogonalização . . . . .	113
5.6.4	Método de Lanczos com Recomeços Implícitos . . . . .	114
5.7	Taxa de Convergência . . . . .	125
5.7.1	Estimativas para $\sigma_d(H(\ell))$ . . . . .	125
5.8	Resultados Numéricos Sobre IRL . . . . .	127
5.9	Métodos para Autovalores de Polinômios Matriciais . . . . .	129
	<b>Apêndice</b>	<b>131</b>

<b>A</b>	<b>Conceitos Básicos da Álgebra Linear</b>	<b>131</b>
A.1	Notação . . . . .	131
A.2	Autovalores e Autovetores . . . . .	132
A.2.1	Condicionamento dos Autovalores . . . . .	134
A.3	Decomposição em Valores Singulares (SVD) . . . . .	136
A.3.1	Sensibilidade dos Valores Singulares . . . . .	138
A.4	Pseudo Inversa . . . . .	138
A.5	Problemas de Quadrados Mínimos Lineares . . . . .	139
A.5.1	Problemas de Quadrados Mínimos Totais . . . . .	140
A.6	Ângulo entre Subespaços . . . . .	142



# Lista de Figuras

1.1	Sistema vibratório massa mola amortecedor . . . . .	15
1.2	Sistema vibratório massa mola . . . . .	16
2.1	Estimativa dos parâmetros $r_l$ e $\lambda_l$ pelo Método de Prony a partir de um sinal $\tilde{s}_k = s_k + \epsilon_k$ . ('o': valores exatos, '+' :valores aproximados). (a): $\lambda$ 's, (b): $r$ 's. . . . .	23
2.2	Sinais exato (linha contínua), perturbado (linha pontilhada), e reconstruído (linha tracejada) usando 64 amostras do sinal. (a): 14 componentes, (b): 20 componentes. . . . .	24
2.3	(a): Sinais $\mathbf{s}_1(t)$ e $\mathbf{s}_2(t)$ . (b): Mesmos sinais de (a) em escala semilogaritmica	24
2.4	Um sistema genérico . . . . .	33
2.5	Protótipo de um prédio de 3 andares . . . . .	35
3.1	Raízes do polinômio de Wilkinson. o: raízes exatas, $\cdot$ : raízes do polinômio com coeficientes $\tilde{c}_j$ , obtidos perturbando os coeficientes $c_j$ do polinômio de Wilkinson. . . . .	56
3.2	Esquerda: Autovalores de $P_2(\lambda)$ . Direita: $\kappa(\lambda, C_A)$ , $\epsilon = 0.1125$ . . . . .	63
3.3	Esquerda: Autovalores de $P_2(\lambda)$ . Direita: $\kappa(\lambda, C_A)$ , $\epsilon = 11.25$ . . . . .	63
4.1	Comportamento de $\ \hat{f}_M\ ^2$ . . . . .	71
4.2	Comportamento de $D_M^2$ . . . . .	72
4.3	Limitante superior para $\kappa_2(W_M)$ , obtida no teorema 4.8: linha sólida, e $\kappa_2(W_M)$ : linha pontilhada, em escala logarítmica. . . . .	72
4.4	$\ E\ /\sigma_d(H(l))$ : linha contínua, $\text{sen } \theta$ : linha tracejada, $2 \text{sen } \frac{\theta}{2}$ : linha tracejada e pontilhada, em escala logarítmica. . . . .	86
4.5	Limitante (4.32): linha contínua, $\ T - \tilde{T}\ $ : linha tracejada e pontilhada, e $ \lambda_l - \tilde{\lambda}_l $ para $l = 2, 4, 5$ : linhas pontilhada, tracejada e contínua-pontilha, respectivamente. . . . .	86

4.6	Resultados utilizando ruído baixo: (a) $\sigma_d(H)$ (linha tracejada) e valor médio de $\ E\ _2$ (linha sólida) como funções da dimensão da matriz $H$ . (b) Valor médio do erro máximo em valor absoluto de $\lambda_j$ (linha sólida) e valor médio da estimativa dada por $\sin(\theta)$ (linha tracejada). . . . .	90
4.7	Resultados utilizando ruído alto: (a) $\sigma_d(H)$ (linha tracejada) e valor médio de $\ E\ _2$ (linha sólida) como funções da dimensão da matriz $H$ . (b) Valor médio do erro máximo em valor absoluto de $\lambda_j$ (linha sólida) e valor médio da estimativa dada por $\sin(\theta)$ (linha tracejada). . . . .	91
5.1	Imagem original . . . . .	94
5.2	k=40 . . . . .	94
5.3	k=80 . . . . .	94
5.4	Em 'o' os 11 valores singulares de $H$ , em 'x' e '+' os 15 maiores valores singulares de $\tilde{H}$ para $\sigma = 5$ e $\sigma = 15$ , respectivamente. . . . .	128
A.1	Interpretação Geométrica do PQML . . . . .	139



# Capítulo 1

## Generalidades

### 1.1 Introdução

O estudo de equações polinomiais remonta a quase 5 séculos e tem tido grande influência no progresso da matemática [91]. Entretanto, a preocupação com equações polinomiais envolvendo matrizes e/ou polinômios com coeficientes matriciais é relativamente recente. Alguns dos primeiros trabalhos neste contexto foram publicados em 1928 por Sylvester [101]. Uma outra referência envolvendo polinômios matriciais, fortemente motivada por problemas relacionados com sistemas vibratórios, é Frazer, Duncan e Collar [48], publicada em 1938.

Um polinômio matricial numa variável complexa  $\lambda$  (também chamado de  $\lambda$ -matriz), é uma função de valor matricial da forma

$$P_m(\lambda) = A_m\lambda^m + A_{m-1}\lambda^{m-1} + \cdots + A_1\lambda + A_0, \quad (1.1)$$

em que os coeficientes  $A_j$  ( $j = 0 : m$ ) são matrizes  $q \times q$  ( $q \geq 1$ ). Ou seja,  $P_m(\lambda)$  é uma matriz  $q \times q$  cujas entradas são polinômios escalares de grau menor ou igual a  $m$ . Se  $A_m$  é a matriz identidade  $q \times q$ ,  $P_m(\lambda)$  é dito polinômio mônico.

Dois problemas destacam no contexto de polinômios matriciais. Um deles dedicado à procura de matrizes  $X \in \mathbb{C}^{q \times q}$  tal que

$$P(X) = A_mX^m + A_{m-1}X^{m-1} + \cdots + A_1X + A_0 = 0, \quad (1.2)$$

e outro preocupado com a busca de escalares complexos  $\lambda$  tal que  $\det(P(\lambda)) = 0$ .

As soluções  $X$  para o primeiro problema, caso elas existam, são chamadas de *solventes* (solvents) do polinômio matricial. A dificuldade com este problema é que a existência de solventes não é garantida pelo teorema fundamental da álgebra (válido para polinômios escalares). O polinômio  $P_2(\lambda) = I\lambda^2 + A_0$  ilustra esta dificuldade. A existência de

solventes neste caso depende da matriz  $A_0$ , pois a equação polinomial  $X^2 = -A_0$  pode ou pode não ter solução quando  $A_0$  é singular, veja, por exemplo, Horn [63, Seção 6.4]. O problema de determinar a existência de solventes data aproximadamente da segunda década de 1900 (veja, Roth [102]) mas hoje em dia é um assunto bem resolvido. A caracterização precisa do número de solventes, no entanto, é um tópico de pesquisa em pleno desenvolvimento e com aplicações em importantes áreas como análise numérica e teoria de controle, entre outras [39, 40, 44]. Algumas outras referências sobre solventes de polinômios matriciais são [57, Seções 4.1 e 4.2] e [27, 98, 61, 102, 46].

Quanto ao segundo problema, há teoria bem desenvolvida e nenhuma dificuldade em relação a existência de soluções  $\lambda$ 's as quais são chamadas de *autovalores* do polinômio matricial. Analogamente ao caso matricial, para cada autovalor  $\lambda$  associa-se um vetor  $x \in \mathbb{C}^q$ ,  $x \neq 0$ , tal que

$$P_m(\lambda)x = 0, \quad (1.3)$$

chamado *autovetor à direita* de  $P_m(\lambda)$  e  $\{\lambda, x\}$  é chamado de *auto-par* (eigenpair) de  $P_m(\lambda)$ . Similarmente, um *autovetor à esquerda* de  $P_m(\lambda)$  associada ao autovalor  $\lambda$ , é um vetor  $y \in \mathbb{C}^{q \times q}$ ,  $y \neq 0$ , tal que

$$y^*P_m(\lambda) = 0. \quad (1.4)$$

O símbolo  $*$  denota conjugação complexa seguida de transposição. O problema de encontrar os auto-pares do polinômio é conhecido como o problema de autovalor polinomial matricial.

Da definição (1.3) segue que se  $m = 1$ , e  $q > 1$ , o problema de autovalor polinomial matricial é simplesmente um problema de autovalor generalizado, ou seja, um problema do tipo  $Ax = \lambda Bx$ , o qual reduz-se ao problema de autovalor matricial padrão quando  $B = I$ . Excelentes referências para o problema de autovalor generalizado são [59, 131, 108]. No caso  $q = 1$ , o problema é aquele de calcular os zeros de um polinômio com coeficientes complexos cuja solução é bem explicada pelo teorema fundamental da álgebra.

Independente do valor de  $q$ , os autovalores de um polinômio matricial podem ser definidos como segue.

**Definição 1.1.1** Seja  $P_m(\lambda)$  o polinômio matricial dado em (1.1). Se  $\det(P_m(\lambda))$  não é identicamente nulo,  $P_m(\lambda)$  é dito *regular*, caso contrário ele é dito *singular*. Se  $P_m(\lambda)$  é regular,  $\mathbf{p}(\lambda) = \det(P_m(\lambda))$  é um polinômio escalar de grau menor ou igual a  $m \cdot q$  chamado de *polinômio característico* de  $P_m(\lambda)$ , e os autovalores de  $P_m(\lambda)$  são definidos como

- (a) As raízes do polinômio característico  $\mathbf{p}(\lambda)$ ,
- (b)  $\infty$ , com multiplicidade  $m \cdot q - \text{grau}(\mathbf{p}(\lambda))$  se  $\text{grau}(\mathbf{p}(\lambda)) < m \cdot q$ .

Autovalores com multiplicidade algébrica igual a 1 são chamados de *autovalores simples*. O conjunto de todos os autovalores é conhecido como *espectro* de  $P_m(\lambda)$  e será denotado por  $\Lambda(P_m(\lambda))$ .

Alguns exemplos que ilustram a definição acima são dados a seguir.

**Exemplo 1.1.2** Seja

$$P_1(\lambda) = A_1\lambda + A_0 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \lambda + \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}.$$

Então, é imediato que  $\det(P_1(\lambda)) \doteq 0$  para todo  $\lambda$ , portanto, o polinômio é singular. Neste caso, todos os números complexos são considerados como autovalores de  $P_1(\lambda)$ . Polinômios singulares aparecem em aplicações práticas mas o estudo deles foge do escopo desta apresentação não sendo, portanto, abordados ao longo deste texto.

O seguinte exemplo ilustra a presença de autovalores  $\infty$ .

**Exemplo 1.1.3** Considere o polinômio matricial

$$P_1(\lambda) = A_1\lambda + A_0 = \begin{bmatrix} -2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \lambda + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} -2\lambda + 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -\lambda \end{bmatrix}.$$

Então  $p(\lambda) = \det P_1(\lambda) = (2\lambda - 1)\lambda$ . Como neste caso  $m = 1$ ,  $q = 3$ , segue então que o espectro do polinômio é:

$$\Lambda(P_1(\lambda)) = \{1/2, 0, \infty\}.$$

Vale destacar que autovalores igual a  $\infty$  aparecem em aplicações práticas; veja K. Brennan [28] para algumas dessas aplicações no contexto de sistemas de equações diferenciais ordinárias com restrições.

**Observação 1.1.4** Diferente ao problema de autovalor matricial padrão, no caso polinomial matricial podem aparecer autovalores distintos possuindo o mesmo autovetor. O seguinte exemplo ilustra este fenômeno.

**Exemplo 1.1.5** Seja  $P_2(\lambda) = A_2\lambda^2 + A_1\lambda + A_0$  tal que

$$A_2 = \begin{bmatrix} 0 & 6 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad A_1 = \begin{bmatrix} 1 & -6 & 0 \\ 2 & -7 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_0 = I.$$

O polinômio característico de  $P_m(\lambda)$  é

$$p(\lambda) = \det(P_m(\lambda)) = -6\lambda^5 + 11\lambda^4 - 12\lambda^3 + 12\lambda^2 - 6\lambda + 1,$$

e os auto-pares  $\{\lambda_k, x_k\}$  são dados na Tabela 1.1

k	1	2	3	4	5	6
$\lambda_k$	1/3	1/2	1	$i$	$-i$	$\infty$
$x_k$	$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

Tabela 1.1: Autovalores diferentes com autovetores iguais.

## 1.2 Linearização

Um aspecto interessante em relação a polinômios matriciais é que o problema de autovalores pode ser reduzido a um problema de autovalor generalizado envolvendo matrizes de ordem  $m \cdot q$ . Para tal, observe que a cada polinômio matricial mônico  $P_m(\lambda)$  pode associar-se uma matriz companheira em blocos definida por

$$C_1 = \begin{bmatrix} 0 & I_q & 0 & \cdots & 0 \\ 0 & 0 & I_q & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & \cdots & I_q \\ -A_0 & -A_1 & & \cdots & -A_{m-1} \end{bmatrix}. \quad (1.5)$$

Um resultado importante que justifica esta associação é que se  $\{\lambda, x\}$  é um auto-par de  $P_m(\lambda)$ , e se  $\hat{x} \in \mathbb{C}^{m \cdot q}$  é um vetor em blocos definido por

$$\hat{x} = \begin{bmatrix} x \\ \lambda x \\ \vdots \\ \lambda^{m-2}x \\ \lambda^{m-1}x \end{bmatrix}, \quad (1.6)$$

então  $\{\lambda, \hat{x}\}$  é um auto-par de  $C_1$ . De fato,

$$C_1 \hat{x} = \begin{bmatrix} \lambda x \\ \lambda^2 x \\ \vdots \\ \lambda^{m-1} x \\ -\sum_{k=1}^m A_{k-1} \lambda^{k-1} x \end{bmatrix} = \begin{bmatrix} \lambda x \\ \lambda^2 x \\ \vdots \\ \lambda^{m-1} x \\ \lambda^m x \end{bmatrix} = \lambda \hat{x}.$$

Portanto, os autovalores de  $P_m(\lambda)$  podem ser calculados a partir da matriz companheira em bloco utilizando-se métodos numéricos para o problema de autovalor matricial padrão. O polinômio linear  $C_1 - \lambda I$  é dito uma *linearização* de  $P_m(\lambda)$ .

Uma outra linearização do polinômio matricial mônico  $P_m(\lambda)$  pode ser construída a partir da matriz companheira em blocos  $C_2$  definida como

$$C_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 & -A_0 \\ I_q & 0 & \cdots & 0 & -A_1 \\ 0 & I_q & \cdots & 0 & -A_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_q & -A_{m-1} \end{bmatrix}. \quad (1.7)$$

Uma relação entre autovetores de  $P_m(\lambda)$  e aqueles da matriz  $C_2$  também pode ser deduzida. Para tal, seja  $\{\lambda, \hat{x}\}$ , um auto par de  $C_2$ . Decompondo  $\hat{x}$  em blocos tal que

$$\hat{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}, \quad x_k \in C^q, \quad k = 1 : m,$$

segue que

$$C_2 \hat{x} = \lambda \hat{x} \iff \begin{cases} -A_0 x_m & = \lambda x_1 \\ x_1 - A_1 x_m & = \lambda x_2 \\ \vdots & \\ x_{m-2} - A_{m-2} x_m & = \lambda x_{m-1} \\ x_{m-1} - A_{m-1} x_m & = \lambda x_m \end{cases} \quad (1.8)$$

A última das equações em (1.8) garante então que  $x_m \neq 0$ , caso contrário  $\hat{x}$  seria o vetor nulo. Utilizando as  $m$  equações em (1.8) é visto facilmente que o autovetor  $\hat{x}$  pode

ser escrito como

$$\widehat{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{m-1} \\ x_m \end{bmatrix} = \begin{bmatrix} (A_1 + A_2\lambda + A_3\lambda^2 + \cdots + A_{m-1}\lambda^{m-2} + \lambda^{m-1}I_q)x_m \\ (A_2 + A_3\lambda + \cdots + A_{m-1}\lambda^{m-3} + \lambda^{m-2}I_q)x_m \\ \vdots \\ (A_{m-1} + \lambda I_q)x_m \\ x_m \end{bmatrix}. \quad (1.9)$$

A primeira equação em (1.8) com  $x_1$  dado pela equação (1.9) garante então que

$$P_m(\lambda)x_m = 0.$$

Portanto, se  $\{\lambda, \widehat{x}\}$  é um auto-par de  $C_2$ , então a última coluna bloco de  $\widehat{x}$  é um autovetor à direita de  $P_m(\lambda)$  associado ao autovalor  $\lambda$ .

Linearizações para o caso em que  $P_m(\lambda)$  não é mônico, ou seja,  $A_m \neq I_q$ , também são possíveis. Se  $A_m$  é não singular, pode-se utilizar o polinômio mônico com coeficientes  $\check{A}_k = A_m^{-1}A_k$ ,  $k = 1 : m-1$ , em cujo caso qualquer uma das linearizações descritas acima pode ser utilizada. Se  $A_m$  é singular, o problema de autovalor polinomial matricial pode ser resolvido através do problema de autovalor generalizado  $Ax = \lambda Bx$ , envolvendo a matriz  $C_1$ , definido como

$$A - \lambda B = \begin{bmatrix} 0 & I_q & 0 & \cdots & 0 \\ 0 & 0 & I_q & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & I_q & 0 \\ -A_0 & -A_1 & \cdots & -A_{m-1} & 0 \end{bmatrix} - \lambda \begin{bmatrix} I_q & 0 & 0 & \cdots & 0 \\ 0 & I_q & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & I_q & 0 \\ 0 & 0 & 0 & \cdots & A_m \end{bmatrix}. \quad (1.10)$$

É claro que os auto-pares deste polinômio linear são da forma  $\{\lambda, \widehat{x}\}$  com  $\widehat{x}$  como em (1.6) e que os  $\lambda$ 's encontrados são autovalores do polinômio matricial  $P_m(\lambda)$ . Uma linearização análoga pode ser construída utilizando-se a matriz  $C_2$ .

A exposição acima sugere que para descrever de maneira precisa os autovalores do polinômio  $P_m(\lambda)$ , basta caracterizar os autovalores do problema de autovalor generalizado  $Ax = \lambda Bx$ . Isto é visto no próximo teorema, cuja prova pode ser encontrada em [42, Cap. 4, p. 174]

**Teorema 1.2.1** *Seja  $Q(\lambda) = A - \lambda B \in C^{n \times n}$  um polinômio regular. Se  $B$  é não singular, todos os autovalores de  $Q(\lambda)$  são finitos e iguais àqueles de  $B^{-1}A$  (ou de  $AB^{-1}$ ). Se  $B$  é singular,  $Q(\lambda)$  tem autovalores  $\infty$  com multiplicidade  $n - \text{rank}(B)$ . Se  $A$  é não*

singular, os autovalores de  $Q(\lambda)$  são os recíprocos dos autovalores de  $A^{-1}B$  (ou de  $BA^{-1}$ ), e autovalores nulos de  $A^{-1}B$  correspondem a autovalores  $\infty$  de  $Q(\lambda)$ .

A caracterização dos autovalores de  $P_m(\lambda)$  depende portanto do coeficiente  $A_m$ , já que dele depende o posto da matriz  $B$  dada em (1.10). O teorema acima sugere que o problema de autovalor generalizado  $Ax = \lambda Bx$  pode ser resolvido através do seguinte método, quando  $B$  é não singular.

- (a) Resolva o sistema  $BC = A$ ,
- (b) Resolva o problema autovalor matricial padrão associado à matriz  $C$  (usando o algoritmo QR)<sup>1</sup>, por exemplo.

Se  $A$  é não singular, o mesmo método pode ser aplicado com  $C$  a solução de  $AC = B$ . Uma desvantagem desta abordagem é que os autovalores calculados podem ser imprecisos se o sistema a ser resolvido é mal-condicionado. Um exemplo que ilustra esta dificuldade pode ser encontrado em [59, Cap. 7, p. 376]. Uma das maneiras mais eficientes de se contornar dificuldades numéricas é através de uma decomposição especial obtida usando matrizes ortogonais, descrita como segue.

**Teorema 1.2.2** [Decomposição de Schur Generalizada] *Seja  $P(\lambda) = A - \lambda B \in \mathbb{C}^{n \times n}$ . Então existem matrizes ortogonais  $Q$  e  $Z$ , e matrizes triangulares superiores  $T$  e  $S$  tal que*

$$Q^*AZ = T, \quad \text{e} \quad Q^*BZ = S.$$

*Além disso, se para algum  $k$ ,  $t_{k,k} = s_{k,k} = 0$ , então  $\Lambda(P(\lambda)) = \mathbb{C}$ , caso contrario*

$$\Lambda(P(\lambda)) = \{t_{i,i}/s_{ii}, s_{ii} \neq 0\}.$$

**Demonstração:** A prova do teorema pode ser encontrada em Golub [59, Cap. 7, p. 377].

□

Aspectos sobre a implementação numérica da decomposição generalizada de Schur são descritos em [59, Cap. 7].

---

<sup>1</sup>Algoritmos para o problema de autovalor matricial padrão são descritos em Capítulo 5

### 1.3 Pares de Jordan

É bem conhecido que toda matriz  $A \in \mathbb{C}^{n \times n}$  tem uma decomposição de Jordan dada como  $A = XJX^{-1}$ , com  $J = \text{diag}(J_1, J_2, \dots, J_t)$ , em que

$$J_k = \begin{bmatrix} \lambda_k & 1 & \cdots & 0 \\ 0 & \lambda_k & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & \lambda_k \end{bmatrix}, k = 1, \dots, t, \quad (1.11)$$

é um bloco de Jordan de ordem  $m_k$ ,  $m_k$  é a multiplicidade do autovalor  $\lambda_k$ , e

$$m_1 + m_2 + \cdots + m_t = n.$$

A matriz  $X$  contém os autovetores e vetores generalizados de  $A$ . A decomposição de Jordan acima fornece uma descrição completa de qualquer polinômio mônico linear  $P_1(\lambda) = I\lambda + A = X(I+J)X^{-1}$ . A pergunta natural então é se alguma fatoração análoga é possível a qual descreva completamente um polinômio matricial de grau  $m$ . A resposta é afirmativa e dada em termos de pares de Jordan associados a  $P_m(\lambda)$ , os quais são definidos de maneira análoga ao caso matricial padrão.

**Definição 1.3.1** Dizemos que uma seqüência de vetores  $x_0, x_1, \dots, x_k$ ,  $x_0 \neq 0$ , é uma *Cadeia de Jordan* de comprimento  $k + 1$  correspondente ao autovalor  $\lambda_0$  de  $P_m(\lambda)$  se

$$\begin{aligned} P_m(\lambda_0)x_0 &= 0 \\ P_m(\lambda_0)x_1 + P_m^{(1)}(\lambda_0)x_0 &= 0 \\ P_m(\lambda_0)x_2 + P_m^{(1)}(\lambda_0)x_1 + \frac{1}{2}P_m^{(2)}(\lambda_0)x_0 &= 0 \\ &\vdots \\ P_m(\lambda_0)x_k + P_m^{(1)}(\lambda_0)x_{k-1} + \cdots + \frac{1}{j!}P_m^{(j)}(\lambda_0)x_{k-j} + \cdots + \frac{1}{k!}P_m^{(k)}(\lambda_0)x_0 &= 0, \end{aligned} \quad (1.12)$$

em que  $P_m^{(j)}(\lambda)$  denota a  $j$ -ésima derivada de  $P_m(\lambda)$ .

**Observação 1.3.2** A definição acima garante que  $x_0$ , algumas vezes dito *vetor líder*, é um autovetor à direita de  $P_m(\lambda)$  associado ao autovalor  $\lambda_0$ . Os vetores restantes da seqüência não precisam ser linearmente independentes (ou seja, algum deles pode ser o vetor nulo), e são conhecidos como *autovetores generalizados*.

Deve ficar claro que dependendo da multiplicidade geométrica do autovalor  $\lambda$ ,<sup>2</sup> difer-

<sup>2</sup>Diz-se que um autovalor  $\lambda_0$  tem *multiplicidade geométrica* igual a  $r$  se

$$\dim[\text{Ker}(P_m(\lambda_0))] = r.$$

entes cadeias de Jordan associadas com diferentes comprimentos podem ser encontradas. O exemplo a seguir ilustra o uso de (1.12) no cálculo de cadeias de Jordan.

**Exemplo 1.3.3** Seja

$$P_2(\lambda) = A_2\lambda^2 + A_1\lambda + A_0 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ -1 & 0 & -1 \end{bmatrix} \lambda^2 + \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \lambda + \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$

O polinômio característico de  $P_2(\lambda)$  é  $p(\lambda) = t^3(t+1)(t-0.5)$ . Os autovalores diferentes de  $P_2(\lambda)$  são portanto:  $\lambda_1 = 0$  com multiplicidade 3,  $\lambda_2 = -1$  com multiplicidade 1, e  $\lambda_3 = 0.5$  com multiplicidade 1. Como

$$P_2(\lambda_1) = P_2(0) = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix},$$

e  $\dim[\text{Ker}(P_2(0))] = 2$ , existem dois autovetores linearmente independentes associados a  $\lambda_1 = 0$ . Portanto, pelo menos duas cadeias de Jordan diferentes podem ser encontradas. Usando a primeira das equações em (1.12), vem

$$P_2(0)x_0 = 0 \implies A_0x_0 = 0.$$

Este sistema têm soluções  $x_0 = [a, b, -a]^T$  para  $a \neq 0$  ou  $b \neq 0$ . Agora, usando a segunda das equações em (1.12) segue que

$$P_2(0)x_1 + P_2^{(1)}(0)x_0 = 0 \implies A_0x_1 + A_1x_0 = 0. \quad (1.13)$$

Se  $x_1 = [c, d, e]^T$ , o sistema acima pode ser escrito como

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} c \\ d \\ e \end{bmatrix} = \begin{bmatrix} -b \\ a \\ a \end{bmatrix}.$$

Assim, se  $b \neq -a$ , o sistema resultante é incompatível, portanto não existe qualquer autovetor generalizado  $x_1$  e a cadeia de Jordan associada ao vetor  $\lambda_1$  com vetor líder  $x_0 = [a, b, -a]^T$  é de comprimento 1. Mas, se  $b = -a$ , o sistema tem soluções da forma  $x_1 = [z, t, a - z]^T$ ,  $t, z$  arbitrários e  $a \neq 0$ . Verifica-se que não existe nenhuma terna de vetores  $x_0, x_1, x_2$  que satisfaça a terceira equação em (1.12). Portanto as cadeias de Jordan associadas ao autovalor  $\lambda_1$  são:

$$x_0 = [a, b, -a]^T, \quad a \neq 0, b \neq 0, b \neq -a \quad (\text{comprimento } 1)$$

$x_0 = [a, -a, -a]^T$ ,  $x_1 = [z, t, a - z]^T$ ,  $z, t \in \mathbb{C}$ ,  $z, t$  arbitrários (comprimento 2).

A cadeia de Jordan associada ao autovalor  $\lambda_2 = -1$  é

$$x_0 = [a, -2a, 0]^T, \quad a \in \mathbb{C}, \quad a \neq 0 \quad (\text{comprimento } 1),$$

e a cadeia de Jordan associada ao autovalor  $\lambda_3 = -0.5$  é

$$x_0 = [v, -0.8v, 0.6v]^T, \quad v \in \mathbb{C}, \quad v \neq 0, \quad (\text{comprimento } 1).$$

**Definição 1.3.4** Seja  $\lambda_0 \in \Lambda(P_m(\lambda))$ . Dizemos que  $\mathcal{J}_1 = \{x_{1,0}, x_{1,1}, \dots, x_{1,s_1-1}\}$ ,  $\mathcal{J}_2 = \{x_{2,0}, x_{2,1}, \dots, x_{2,s_2-1}\}, \dots, \mathcal{J}_{r_0} = \{x_{r_0,0}, x_{r_0,1}, \dots, x_{r_0,s_{r_0}-1}\}$ , é um conjunto canônico de cadeias de Jordan associadas ao autovalor  $\lambda_0$  se

- (a)  $\mathcal{J}_1$  é de comprimento máximo com  $v_{1,0} \in \text{Ker}(P_m(\lambda_0))$ ,
- (b)  $\mathcal{J}_i$ , é de comprimento máximo com  $v_{i,0} \notin \text{span}\{v_{1,0}, v_{2,0}, \dots, v_{i-1,0}\}$  para  $i = 2: r_0$ , e
- (c)  $r_0 = \dim[\text{Ker}(P_m(\lambda_0))]$ .

**Observação 1.3.5** Um resultado importante em relação a conjuntos canônicos de cadeias de Jordan é que se  $m_0$  é a multiplicidade algébrica de  $\lambda_0$ , então (veja, [57, Cap. 1, Prop. 1.15])

$$s_1 + s_2 + \dots + s_{r_0} = m_0.$$

Os números  $s_1, \dots, s_{r_0}$  são determinados unicamente (não dependem dos autovetores escolhidos) e conhecidos como *multiplicidades parciais* do autovalor  $\lambda_0$ .

**Definição 1.3.6** Seja  $P_m(\lambda)$  um polinômio mônico com  $t$  autovalores distintos  $\lambda_1, \dots, \lambda_t$  cujas multiplicidades algébricas são  $m_1, \dots, m_t$ , ou seja  $m_1 + m_2 + \dots + m_t = m \cdot q$ . Seja  $\mathcal{J}_1^j, \dots, \mathcal{J}_{r_j}^j$  um conjunto canônico de cadeias de Jordan associado ao autovalor  $\lambda_j$ . Seja  $X_j$  a matriz  $m \times m_j$  definida por

$$X_j = [\mathcal{J}_1^j, \dots, \mathcal{J}_{r_j}^j], \quad 1 \leq j \leq t.$$

Seja  $J_j = \text{diag}(J_{r_{1,j}}, \dots, J_{r_{j,j}})$ , com  $J_{r_{i,j}}$ ,  $1 \leq i \leq j$ , um bloco de Jordan associado ao autovalor  $\lambda_j$ , de ordem  $r_{i,j}$ , com  $r_{i,j}$  o comprimento da cadeia de Jordan  $\mathcal{J}_i^j$ ,  $1 \leq i \leq r_j$ . Um *par de Jordan* associado ao autovalor  $\lambda_j$  é a dupla  $\{X_j, J_j\}$ , enquanto que um *par de Jordan* associado ao polinômio  $P_m(\lambda)$  é a dupla  $\{X, J\}$ , com  $X = [X_1, \dots, X_t] \in \mathbb{C}^{q \times (m \cdot q)}$ ,  $J = \text{diag}(J_1, \dots, J_t) \in \mathbb{C}^{(m \cdot q) \times (m \cdot q)}$ .

**Exemplo 1.3.7** Seja  $P_3(\lambda)$  definido por

$$P_3(\lambda) = \begin{bmatrix} \lambda^3 & \sqrt{2}\lambda^2 - \lambda \\ \sqrt{2}\lambda^2 + \lambda & \lambda^3 \end{bmatrix}.$$

O polinômio característico de  $P_3(\lambda)$  é  $p(\lambda) = \lambda^2(\lambda - 1)^2(\lambda + 1)^2$ , portanto, tem-se que

$$\Lambda(P_3(\lambda)) = \{\lambda_1, \lambda_2, \lambda_3\} = \{0, 1, -1\}.$$

Como  $P_3(\lambda_1) = P_3(0) = 0$ , existem dois autovetores linearmente independentes associados a  $\lambda_1 = 0$ . Escolhendo  $v_{1,0} = [1, 0]^T$ , o mesmo procedimento do Exemplo 1.3.3 assegura que

$$v_{1,0} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad v_{1,1} = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

é um conjunto canônico de cadeias de Jordan associado a  $\lambda_1 = 0$  (já que  $\lambda_1$  é de multiplicidade algébrica dois, veja 1.3.5), portanto, um par de Jordan é

$$X_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad J_1 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Consideremos agora o autovalor  $\lambda_2 = 1$ . Como

$$P_3(\lambda_2) = P_3(1) = \begin{bmatrix} 1 & \sqrt{2} - 1 \\ \sqrt{2} + 1 & 1 \end{bmatrix},$$

utilizando a primeira das equações em (1.12), segue que  $v_{2,0}$  é

$$v_{2,0} = \begin{bmatrix} -\sqrt{2} + 1 \\ 1 \end{bmatrix}.$$

Agora, utilizando a segunda das equações em (1.12) para calcular o primeiro vetor generalizado  $v_{2,1}$ , segue que  $v_{2,1}$  pode ser escrito como

$$v_{2,1} = \begin{bmatrix} \sqrt{2} - 2 \\ 0 \end{bmatrix}.$$

Portanto, um par de Jordan associado ao autovalor  $\lambda_2$  é

$$X_2 = \begin{bmatrix} -\sqrt{2} + 1 & \sqrt{2} - 2 \\ 1 & 0 \end{bmatrix}, \quad J_2 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Desenvolvendo o mesmo procedimento, pode-se verificar que um par de Jordan associado ao autovalor  $\lambda_3 = -1$  é

$$X_3 = \begin{bmatrix} \sqrt{2} + 1 & \sqrt{2} + 2 \\ 1 & 0 \end{bmatrix}, \quad J_3 = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}.$$

Finalmente, um par de Jordan associado ao polinômio  $P_3(\lambda)$  é:

$$X = [X_1, X_2, X_3] = \begin{bmatrix} 1 & 0 & -\sqrt{2} + 1 & \sqrt{2} - 2 & \sqrt{2} + 1 & \sqrt{2} + 2 \\ 0 & 1 & 10 & 0 & 1 & 0 \end{bmatrix},$$

$$J = \begin{bmatrix} 0 & 1 & & & & \\ & 0 & & & & \\ & & 1 & 1 & & \\ & & & 1 & & \\ & & & & -1 & 1 \\ & & & & & -1 \end{bmatrix}.$$

Algumas propriedades de pares de Jordan, cujas provas podem ser encontradas em [57, Cap. 1, p. 45], são descritas no próximo teorema.

**Teorema 1.3.8** *Seja  $\{X, J\}$  um par de Jordan do polinômio mônico  $P_m(\lambda)$ . Então:*

- (a) *As colunas de  $X$  são linearmente independentes se e somente se  $P_m(\lambda)$  é linear.*
- (b)  $XJ^m + A_{m-1}XJ^{m-1} + \dots + A_1XJ + A_0X = 0$ .
- (c) *Seja  $S$  a matriz  $(m \cdot q) \times (m \cdot q)$  definida por*

$$S = \begin{bmatrix} X \\ XJ \\ \vdots \\ XJ^{m-1} \end{bmatrix} = \begin{bmatrix} X_1 & X_2 & \dots & X_t \\ X_1J_1 & X_2J_2 & \dots & X_tJ_t \\ \vdots & \vdots & \dots & \vdots \\ X_1J_1^{m-1} & X_2J_2^{m-1} & \dots & X_tJ_t^{m-1} \end{bmatrix}. \quad (1.14)$$

*Então  $S$  é não singular e a matriz companheira  $C_1$  definida em (1.5) satisfaz:*

$$C_1 = SJS^{-1}. \quad (1.15)$$

A equação (1.15) é uma decomposição de Jordan da matriz  $C_1$ . Assim, todas as propriedades do polinômio mônico  $P_m(\lambda)$  concentram-se na matriz  $S$  e a matriz de Jordan  $J$ . Isto responde a questão colocada no início da seção.

Finalmente, uma *tripla de Jordan* do polinômio mônico  $P_m(\lambda)$  é uma terna  $\{X, J, Y\}$ , com  $Y$  de ordem  $(m \cdot q) \times q$  definido por

$$Y = S^{-1} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ I_q \end{bmatrix} \quad (1.16)$$

Triplas de Jordan são utilizadas para resolver sistemas de equações diferenciais ordinárias com coeficientes constantes [57], detalhes para o caso quadrático são dados na próxima seção.

Quando o polinômio não é mônico e o coeficiente  $A_m$  é singular, o par de Jordan  $\{X, J\}$  deve ser formado por um par de Jordan associado aos autovalores finitos, digamos  $\{X_F, J_F\}$ , e por outro par de Jordan associado aos autovalores  $\infty$ , digamos  $\{X_\infty, J_\infty\}$ , em que  $J_\infty$  é um bloco de Jordan com autovalor  $\lambda = 0$ . Detalhes para o caso quadrático podem ser encontrados em [122].

## 1.4 Espectro de Polinômios Matriciais Quadráticos

Polinômios matriciais quadráticos aparecem de maneira natural em conexão com sistemas de equações diferenciais ordinárias do tipo

$$M\ddot{u}(t) + C\dot{u}(t) + Ku(t) = f(t), \quad (1.17)$$

em que  $M$ ,  $C$  e  $K$  são matrizes  $n \times n$ , e  $u(t) = [u_1(t), \dots, u_n(t)]^T$ ,  $f(t) = [f_1(t), \dots, f_n(t)]^T$  são funções vetoriais de valor real. Na área de sistemas vibratórios em engenharia mecânica, por exemplo,  $M$ ,  $C$  e  $K$  são simétricas,  $M$  é conhecida como matriz de massa,  $C$  como matriz de amortecimento, e  $K$  como matriz de rigidez. Num certo sentido, o modelo (1.17) descreve a equação de movimento de um sistema que generaliza o sistema massa-mola-amortecedor descrito na Figura 1.1.

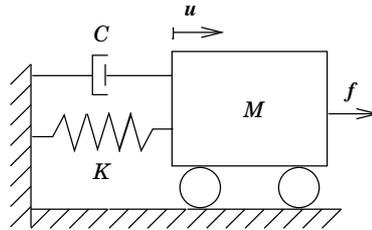


Figura 1.1: Sistema vibratório massa mola amortecedor

A título de exemplo, considere o sistema mecânico da Figura 1.2. Ele consiste de dois blocos em movimento com massas  $m_1$  e  $m_2$ , conetados por três molas cujas constantes são  $k_1, k_2$  e  $k_3$ . Para efeitos de simplificação, assuma que não existe fricção entre os blocos e o piso.

Aplicando a lei de Newton à força  $f_1$  vem

$$f_1 - k_1 u_1 - k_2(u_1 - u_2) = m_1 \ddot{u}_1,$$

ou equivalentemente

$$m_1 \ddot{u}_1 + (k_1 + k_2)u_1 - k_2 u_2 = f_1.$$

Procedendo analogamente com a força  $f_2$  vem

$$m_2 \ddot{u}_2 - k_2 u_1 + (k_1 + k_2)u_2 = f_2.$$

As duas equações acima podem ser colocadas matricialmente como

$$\begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \begin{bmatrix} \ddot{u}_1 \\ \ddot{u}_2 \end{bmatrix} + \begin{bmatrix} k_1 + k_2 & -k_2 \\ -k_2 & k_1 + k_2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix},$$

que é um caso particular do sistema descrito em (1.17).

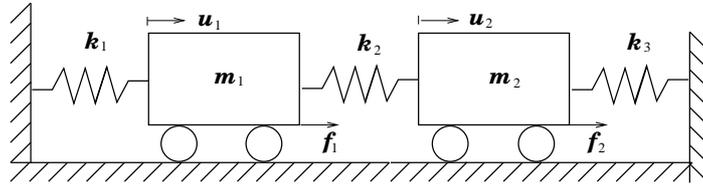


Figura 1.2: Sistema vibratório massa mola

O problema de autovalor polinomial matricial quadrático aparece quando procura-se soluções do problema homogêneo associado ao sistema (1.17) da forma  $u(t) = e^{\lambda t}x$ , tal que  $\lambda \in \mathbb{C}$  e  $x \in \mathbb{C}^n$ ,  $x \neq 0$ , não dependem de  $t$ . De fato, como a função  $u$  escolhida deve satisfazer a equação diferencial

$$M\ddot{u}(t) + C\dot{u}(t) + Ku(t) = 0,$$

é imediato que  $\lambda$  e  $x$  devem satisfazer a equação

$$(M\lambda^2 + C\lambda + K)x = 0. \quad (1.18)$$

Esta é uma equação autovalor autovetor associada ao polinômio quadrático

$$P(\lambda) = M\lambda^2 + C\lambda + K. \quad (1.19)$$

Aplicações do modelo (1.17) envolvendo matrizes de diferentes tipos são encontradas em muitas outras áreas, uma vasta lista de referências sobre o assunto pode ser encontrada em [122]. Quanto ao espectro de  $P(\lambda)$ , ele depende das matrizes envolvidas, e é formado por  $2n$  autovalores (finitos ou infinitos) com até  $2n$  autovetores associados. A existência de mais do que  $n$  autovetores implica necessariamente que eles são linearmente dependentes. Uma caracterização do espectro e autovetores do polinômio quadrático como função das características das matrizes envolvidas é apresentada na Tabela 1.2.

### 1.4.1 Solução de Sistemas de Segunda Ordem

O objetivo aqui é utilizar a informação espectral concentrada numa tripla de Jordan para resolver o sistema descrito em (1.17). Assuma que  $M$  é não singular. Seja  $\{X, J, Y\}$  uma tripla de Jordan associada ao polinômio  $P_m(\lambda)$ . A parte (b) do Teorema 1.3.8 implica então que

$$MXJ^2 + CXJ + KX = 0. \quad (1.20)$$

Quanto a matriz  $Y$ , ela satisfaz a equação (veja (1.16)).

$$Y = \begin{bmatrix} X \\ XJ \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ I \end{bmatrix} M^{-1}. \quad (1.21)$$

Utilizando a definição de  $X$  e  $Y$ , prova-se que

$$\begin{aligned} YCX + YMXJ + JYMX &= I, \\ XYM = 0 \quad XJYM &= I. \end{aligned} \quad (1.22)$$

Usando (1.20) verifica-se facilmente que a função

$$u_h(t) = Xe^{Jt}x_0,$$

com  $x_0$  um vetor constante em  $\mathbb{C}^{2n}$ , é a solução geral da equação diferencial homogênea associada.

Defina  $u_p(t)$  por

$$u_p(t) = Xe^{Jt} \int_0^t e^{-Js} Y f(s) ds.$$

Diferenciando esta função segue que

$$\dot{u}_p(t) = XJe^{Jt} \int_0^t e^{-Js} Y f(s) ds,$$

e

$$\ddot{u}_p(t) = XJ^2e^{Jt} \int_0^t e^{-Js}Yf(s)ds + XJYf(t).$$

Substituindo  $u_p(t)$ ,  $\dot{u}_p(t)$ , e  $\ddot{u}_p(t)$  em (1.17), a segunda equação em (1.22) assegura então que

$$M\ddot{u}_p(t) + C\dot{u}_p(t) + K\ddot{u}_p(t) = f(t),$$

e assim,  $u_p(t)$  é uma solução particular do sistema (1.17). Portanto, a solução geral do sistema (1.17) pode ser escrita como

$$u(t) = u_h(t) + u_p(t) = Xe^{Jt} \left( x_0 + \int_0^t e^{-Js}Yf(s)ds \right).$$

Detalhes adicionais sobre a solução do problema quando  $M$  é singular podem ser encontrados em Gohberg [57, Cap. 8].

Uma outra forma de resolver o problema é através de alguma linearização do polinômio quadrático. Algumas linearizações bem como métodos numéricos para resolver o problema de autovalores são comentados no Capítulo 5.

Matrizes Envolvidas	Tipo de Autovalores	Tipo de Autovetores
$M$ não singular	$2n$ $\lambda$ 's finitos	
$M$ singular	$\lambda$ 's finitos e infinitos	
$M, C, K$ reais	$\lambda$ 's reais ou em pares complexos conjugados $(\lambda, \bar{\lambda})$	Se $\{\lambda, x\}$ é auto-par $\{\bar{\lambda}, \bar{x}\}$ é auto-par
$M, C, K$ Hermitianas	$\lambda$ 's reais ou em pares complexos conjugados $(\lambda, \bar{\lambda})$	Se $\{\lambda, x\}$ é auto-par $\{\bar{\lambda}, x\}$ é auto-par com $x$ autovetor à esq.
$M, C, K$ Hermitianas $M > 0, C, K \geq 0$	$\text{Re}(\lambda) \leq 0$	
$M, K$ Hermitianas $M > 0, C = -C^*$	$\lambda$ 's imaginários puros ou em pares $(\lambda, -\bar{\lambda})$	Se $\{\lambda, x\}$ é auto-par $\{-\bar{\lambda}, x\}$ é auto-par com $x$ autovetor à esq.
$M, K$ Reais simétricas $M > 0, K > 0, C = -C^*$	$\lambda$ 's imaginários puros	

Tabela 1.2: Descrição do espectro do polinômio quadrático  $P(\lambda) = M\lambda^2 + C\lambda + K$  como função das características das matrizes envolvidas.  $M > 0$  ( $\geq 0$ ) significa  $M$  definida positiva (semidefinida positiva)



# Capítulo 2

## Aplicações Correntes

Este capítulo destina-se a apresentar algumas aplicações envolvendo polinômios matriciais com o intuito de salientar a relevância do estudo teórico do problema de autovalor polinomial matricial bem como de técnicas para calcular o espectro total ou parcial dos polinômios matriciais envolvidos.

### 2.1 Recuperação de Harmônicos

Seja  $s_k$ ,  $k = 0, 1, 2, \dots$ , um conjunto de amostras de um sinal discreto no tempo, modelado como uma soma de exponenciais complexas

$$s_k = \sum_{j=1}^d r_j \lambda_j^k = \sum_{j=1}^d r_j e^{(\alpha_j + i\omega_j)\Delta t k}, \quad (2.1)$$

em que os coeficientes  $r_j$  são números complexos,  $i^2 = -1$ ,  $\alpha_j \leq 0$ ,  $\omega_j \in \mathbb{R}$ , e  $\Delta t$  é a taxa de amostragem. Sinais deste tipo são geralmente associados à resposta impulsiva de sistemas invariantes no tempo, e o significado dos parâmetros envolvidos depende do sistema em estudo ou da origem do sinal. Se  $s_k$  é a resposta livre ou a resposta ao impulso unitário de um sistema mecânico, por exemplo, os  $\alpha_j$  fornecem amortecimentos do sistema, os  $\omega_j$  descrevem as frequências naturais de vibração e os  $r_j$  o grau de participação de cada modo do sistema.

Dada uma seqüência finita de amostras  $\tilde{s}_k = s_k + \epsilon_k$ ,  $k = 0, 1, \dots, L$ , em que  $\epsilon_k$  denotam incertezas no sinal, o problema de *recuperação de harmônicos* consiste em calcular o inteiro  $d$  e estimativas dos coeficientes  $r_j$  e as exponenciais  $\lambda_j$ . As incertezas  $\epsilon_k$  são de diversas naturezas (imprecisões de aparelhos de medição, erros de arredondamento, fatores externos, etc) e chamadas simplesmente de *ruídos*.

O problema data de 1795 e foi inicialmente resolvido por de R. de Prony [100]. Hoje é muito freqüente em áreas como, comunicações, localização de objetos por radar, processa-

mento de sinais sísmicos, identificação paramétrica de sistemas dinâmicos, análise modal de estruturas mecânicas, ressonância magnética nuclear, análise de compartimentos, diagnósticos médicos, etc [6, 9, 19, 20, 32, 66, 71, 100, 118].

A idéia chave de R. de Prony foi provar que o sinal  $s_k$  satisfaz uma equação de diferenças. De fato, seja  $P(\lambda)$  um polinômio de grau  $d$ , com zeros  $\lambda_l$ ,  $l = 1, \dots, d$

$$P(\lambda) = \prod_{l=1}^d (\lambda - \lambda_l) = \sum_{i=0}^d c_i \lambda^i, \quad c_d = 1. \quad (2.2)$$

Fazendo o somatório  $\sum_{j=0}^d c_j s_{j+i}$ , para  $1 \leq i \leq L - d$ , e usando a equação (2.1), tem-se:

$$\begin{aligned} \sum_{j=0}^d c_j s_{j+i} &= \sum_{j=0}^d c_j \sum_{l=1}^d r_l \lambda_l^{j+i} = \sum_{j=0}^d c_j \sum_{l=1}^d r_l \lambda_l^j \lambda_l^i \\ &= \sum_{l=1}^d r_l \lambda_l^i \sum_{j=0}^d c_j \lambda_l^j = \sum_{l=1}^d r_l \lambda_l^i P(\lambda_l) = 0, \end{aligned}$$

pois os  $\lambda_l$  são raízes de  $P(\lambda)$ . Assim, o sinal  $s_j$  satisfaz a equação de diferenças

$$\sum_{j=0}^d c_j s_{j+i} = 0. \quad (2.3)$$

Reciprocamente, prova-se que se  $\gamma_j$ ,  $j = 1 : d - 1$  satisfaz (2.3), então o polinômio  $P(\lambda)$  (2.2) tem  $d$  raízes que são exatamente os  $d$ - $\lambda$ 's.

Assumindo o número  $d$  conhecido, o método de Prony para calcular as exponenciais  $\lambda$ 's pode ser resumido a:

1. Resolver o sistema (2.4):

$$\begin{bmatrix} s_1 & s_2 & \cdots & s_d \\ s_2 & s_3 & \cdots & s_{d+1} \\ \vdots & \vdots & \cdots & \vdots \\ s_{L-d} & s_{L-d+1} & \cdots & s_{L-1} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{d-1} \end{bmatrix} = - \begin{bmatrix} s_{d+1} \\ s_{d+2} \\ \vdots \\ s_L \end{bmatrix}, \quad L = 2d, \quad (2.4)$$

e construir o polinômio  $P(\lambda)$ .

2. Determinar as raízes do polinômio  $P(\lambda)$  de (2.2).

Os parâmetros  $r_l$  são estimados facilmente pelo método dos quadrados mínimos lineares após a determinação dos  $\lambda$ 's [7, 9].

O sistema acima é conhecido como as equações de Yule-Walker [130]. Curiosamente, Prony apresentou esta equação matricial mais de um século antes de Yule (1927) e Walker (1931).

A grande dificuldade do método de Prony é que, em geral, as raízes de um polinômio podem ser muito sensíveis a pequenas perturbações nos coeficientes. Para ilustrar este fenômeno, o método de Prony foi aplicado para recuperar os parâmetros de um sinal com 8 componentes (ou seja,  $d = 8$ ) as quais são descritas na Tabela 2.1.

$\text{real}(\lambda_l)$	$\text{imag}(\lambda_l)$	$\text{real}(r_l)$	$\text{imag}(r_l)$
0.2500	$\pm 0.8800$	0	$\pm 0.8000$
0.3000	$\pm 0.8500$	-0.7300	$\pm 0.9800$
0.7800	$\pm 0.5800$	1.2500	$\pm 0.4800$
-0.6250	$\pm 0.4600$	1.0000	$\pm 1.3750$

Tabela 2.1: Parâmetros de um sinal real com 8 componentes.

A Figura 2.1 mostra os resultados obtidos pelo método a partir de um conjunto de dados  $\tilde{s}_k = s_k + \epsilon_k$  tal que  $\|\epsilon\|/\|s\| \approx 0.03$  em que  $\epsilon$  e  $s$  denotam vetores contendo respectivamente as amostras do ruído (Gaussiano com média zero) e do sinal. Ao todo foram considerados 64 amostras, porém para aplicar o método, foram utilizadas somente as 16 primeiras. Observe nesse gráfico como os parâmetros  $\lambda$ 's mais próximos são mais sensíveis a perturbações.

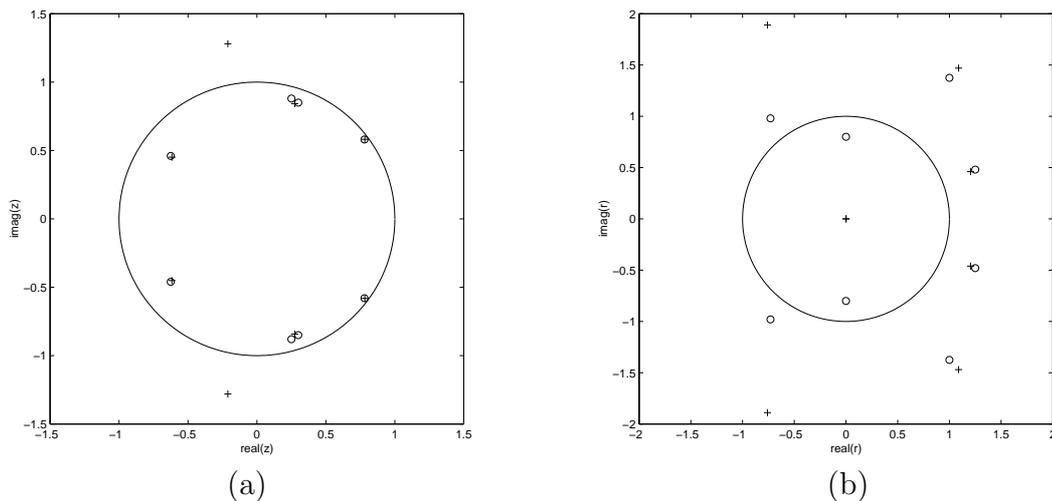


Figura 2.1: Estimativa dos parâmetros  $r_l$  e  $\lambda_l$  pelo Método de Prony a partir de um sinal  $\tilde{s}_k = s_k + \epsilon_k$ . ('o': valores exatos, '+' : valores aproximados). (a):  $\lambda$ 's, (b):  $r$ 's.

Outra dificuldade associada ao problema é que sinais do tipo exponencial contami-

nados por ruídos podem ser descritos muito bem utilizando um número arbitrário de componentes. Isto é ilustrado na Figura 2.2 que mostra que um mesmo sinal pode ser bem “ajustado” utilizando um número arbitrário de termos.

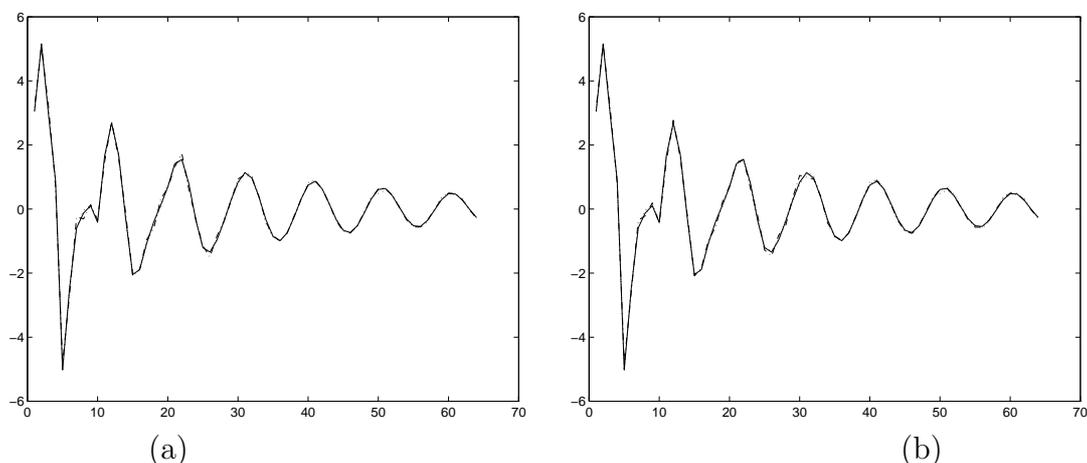


Figura 2.2: Sinais exato (linha contínua), perturbado (linha pontilhada), e reconstruído (linha tracejada) usando 64 amostras do sinal. (a): 14 componentes, (b): 20 componentes.

Finalmente, outro exemplo que ilustra que sinais compostos por exponenciais contendo parâmetros muito diferentes podem fornecer curvas muito próximas umas das outras é apresentado na Figura (2.3). Neste caso, trata-se de dois sinais  $s_1(t)$  (com 3 exponenciais) e  $s_2(t)$  (com 2 exponenciais) definidos por

$$\begin{aligned} s_1(t) &= 0.305e^{-t/0.633} + 2.202e^{-t/0.225} \\ s_2(t) &= 0.0951e^{-t} + 0.8607e^{-t/0.333} + 1.557e^{-t/0.2}. \end{aligned} \quad (2.5)$$

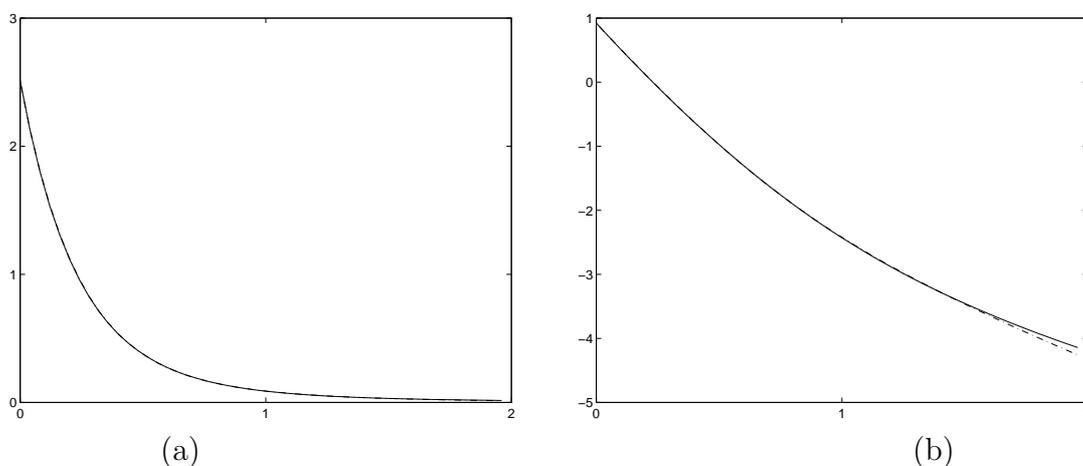


Figura 2.3: (a): Sinais  $s_1(t)$  e  $s_2(t)$ . (b): Mesmos sinais de (a) em escala semilogaritmica

As dificuldades associadas ao problema de recuperação de harmônicos podem ser resumidas a:

- O número  $d$  é desconhecido.
- As raízes de um polinômio podem ser muito sensíveis a pequenas perturbações nos coeficientes.

### 2.1.1 Abordagem Polinomial

Como os  $r$ 's entram linearmente no modelo e os parâmetros embutidos nas exponenciais entram não linearmente (veja, (2.1)), o problema pode ser abordado pelo método dos *quadrados mínimos* para problemas não-lineares, usando o método de Gauss-Newton, por exemplo, na sua versão para problemas separáveis [113]). Entretanto, vale enfatizar que, a menos que o número de componentes do sinal seja conhecido a priori, e que o algoritmo utilizado seja inicializado adequadamente, a solução construída por esses algoritmos pode não ser satisfatória [113, 124].

As dificuldades acima podem ser contornadas elegantemente através de técnicas da álgebra linear computacional. A idéia básica é que o número  $d$  pode ser detectado pela inspeção do posto de uma versão sobredeterminada da matriz do sistema de Yule-Walker (2.4), chamada de *matriz observação*, e que os  $\lambda$ 's podem ser estimados através da solução de um problema de autovalor (ou via extração das raízes de um polinômio).

Começamos mostrando a conexão entre o posto da matriz observação com o número de componentes do sinal. Seja  $H(\ell)$  a matriz de ordem  $M \times N$  definida por

$$H(\ell) = \begin{bmatrix} s_\ell & s_{\ell+1} & \cdots & s_{\ell+N-1} \\ s_{\ell+1} & s_{\ell+2} & \cdots & s_{\ell+N} \\ \vdots & \vdots & \cdots & \vdots \\ s_{\ell+M-1} & s_{\ell+M} & \cdots & s_{\ell+M+N-2} \end{bmatrix}, \quad \ell \geq 0. \quad (2.6)$$

**Teorema 2.1.1** *Seja  $H(\ell)$  a matriz de Hankel definida acima. Então*

$$H(\ell) = V_M \Lambda^l R V_N^T, \quad (2.7)$$

com  $V_M$  a matriz de Vandermonde

$$V_M = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_d \\ \lambda_1^2 & \lambda_2^2 & \cdots & \lambda_d^2 \\ \vdots & \vdots & \cdots & \vdots \\ \lambda_1^{M-1} & \lambda_2^{M-1} & \cdots & \lambda_d^{M-1} \end{bmatrix}, \quad (2.8)$$

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_d), \quad \text{e} \quad R = \text{diag}(r_1, \dots, r_d). \quad (2.9)$$

Conseqüentemente, sempre que  $\min(M, N) \geq d$ , vale

$$\text{rank}(H(\ell)) = d. \quad (2.10)$$

**Demonstração:** A fatoração (2.7) é uma conseqüência imediata da estrutura da matriz  $H(\ell)$  e do modelo (2.1). A propriedade (2.10)<sup>1</sup> segue do fato de que o subespaço coluna da matriz  $H(\ell)$ , chamado *subespaço sinal*, é gerado pelas colunas de  $V_M$  (veja, (2.7)).

□

A abordagem polinomial para o problema utiliza modelos de predição linear progressivos e regressivos. Um modelo de predição linear progressivo assume que  $s_{\ell+N}$  tem a forma:

$$c_0 s_\ell + c_1 s_{\ell+1} + \dots + c_{N-1} s_{\ell+N-1} = s_{\ell+N}, \quad \ell \geq 0. \quad (2.11)$$

Nesta formulação, a ordem do modelo é  $N$  e os coeficientes  $c$ 's, chamados *parâmetros preditores*, são estimados através da solução do sistema de equações de predição :

$$H(\ell)c = H(\ell+1)e_N, \quad (2.12)$$

em que  $e_N$  é o vetor canônico em  $\mathbb{R}^N$ . O sistema acima é consistente e admite infinitas soluções. O lema a seguir, garante que os  $\lambda$ 's podem ser extraídos sempre a partir das raízes de um polinômio de grau  $N \geq d$  ( $d$  é o número de componentes do sinal):

$$P_f(t) = t^N + c_{N-1}t^{N-1} + \dots + c_1t + c_0, \quad (2.13)$$

o qual é chamado *polinômio preditor progressivo*.

**Lema 2.1.2** *Seja  $P_f(t)$  definido em (2.13). Então, sempre que  $N \geq d$ , os  $\lambda_j$ 's são zeros de  $P_f(t)$ .*

**Demonstração:** Considere a matriz companheira associada a  $P_p(t)$ :

$$C_f = \begin{bmatrix} 0 & 0 & \dots & 0 & -c_0 \\ 1 & 0 & \dots & 0 & -c_1 \\ 0 & 1 & \dots & 0 & -c_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -c_{N-1} \end{bmatrix}_{N \times N}. \quad (2.14)$$

---

<sup>1</sup>rank( $\cdot$ ) denota o posto da matriz ( $\cdot$ )

Agora note que, por causa da estrutura da matriz de Hankel  $H(\ell)$ , o sistema (2.12) pode ser reescrito como

$$H(l+1) = H(l)C_f.$$

Daí, usando a fatoração (2.7), segue que

$$V_N C_f = \Lambda V_N^T,$$

a qual garante que as linhas de  $V_N^T$  são autovetores à esquerda de  $C_f$ , associados aos autovalores  $\lambda_1, \dots, \lambda_d$ . Como os autovalores de  $C_f$  são as raízes do polinômio associado, isto prova que os  $\lambda$ 's são raízes de  $P_f(t)$ .

□

Observe que  $d$  zeros do polinômio  $P_f(t)$ , chamados *zeros do sinal*, são da forma  $\lambda_l = e^{(\alpha_l + i\omega_l)\Delta t}$ ,  $l = 1, \dots, d$ .

Um modelo de predição linear na ordem inversa (ou regressiva) é da forma:

$$d_0 h_\ell + d_1 s_{\ell+1} + \dots + d_{N-1} s_{\ell+N-1} = s_{\ell-1}, \quad \ell \geq 1.$$

Os parâmetros  $\lambda_j$  são obtidos, agora, dos zeros de polinômios preditores regressivos:

$$P_b(t) = t^N + d_0 t^{N-1} + \dots + d_{N-2} t + d_{N-1}, \quad (2.15)$$

cujos coeficientes são obtidos pela solução de

$$H(\ell)d = H(\ell-1)e_1, \quad \ell \geq 1, \quad (2.16)$$

em que  $e_1$  denota o primeiro vetor canônico em  $\mathbb{R}^N$ . Neste caso, o problema de calcular os zeros do polinômio  $P_b(t)$  é equivalente ao problema de calcular os autovalores da matriz companheira  $\mathcal{C}_b$ , dada por:

$$\mathcal{C}_b = \begin{bmatrix} -d_0 & 1 & 0 & \dots & 0 \\ -d_1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -d_{N-2} & 0 & 0 & \dots & 1 \\ -d_{N-1} & 0 & 0 & \dots & 0 \end{bmatrix}_{N \times N}.$$

Com um raciocínio análogo ao anterior, obtém-se que  $d$  autovalores de  $\mathcal{D}$ , e portanto,  $d$  zeros de  $P_b(t)$ , são da forma  $\lambda_l^{-1} = e^{-(\alpha_l + i\omega_l)\Delta t}$ .

Em ambos os casos, existem  $N - d$  zeros (*zeros espúrios*) sem significado físico e que decorrem de usar um polinômio de grau maior do que o necessário. A razão de utilizar

$N \geq d$  é porque em problemas práticos o número de componentes do sinal é desconhecido. A separação dos  $d$  zeros desejados não é imediata uma vez que os  $(N - d)$  zeros espúrios não dependem da escolha dos coeficientes  $c_i$  ( $d_i$ ), feita dentre as infinitas soluções do sistemas (2.12) e (2.16).

O seguinte teorema fornece informação sobre os zeros espúrios.

**Teorema 2.1.3** *Se os coeficientes  $c_i$  e  $d_i$  dos polinômios  $P_f(t)$  e  $P_b(t)$  são respectivamente, as componentes da solução de norma mínima dos sistemas (2.12) e (2.16), então os zeros espúrios de ambos os polinômios ficam localizados no interior do círculo unitário. Ainda mais, os zeros espúrios de  $P_f(t)$  são os complexos conjugados daqueles de  $P_b(t)$ .*

**Demonstração:** Veja [7].

□

Visto que os zeros do sinal associados ao polinômio  $P_b(t)$  ficam no exterior do círculo unitário, pois  $|\lambda_j^{-1}| > 1$ , e os zeros espúrios ficam dentro, durante muito tempo o problema de recuperar exponenciais foi abordado a partir do cálculo das raízes desse polinômio, apesar do teorema ser válido quando o sinal é livre de ruído. Um método neste contexto é o método de Tufts e Kumaresan (KT), onde os coeficientes são calculados usando a técnica da SVD truncada (TSVD) [78]. Outro método pode ser encontrado em [9], onde os coeficientes são calculados através do método dos gradientes conjugados. A dificuldade encontrada nos métodos existentes, que utilizam o polinômio regressivo, é que a separação desejada torna-se complicada para sinais com alto nível de ruído.

## 2.1.2 Métodos de Subespaço

Os métodos do tipo subespaço podem ser divididos em duas classes [125]: métodos de ajuste de subespaço [126] e métodos que exploram a propriedade da matriz Hankel ser invariante ao deslocamento<sup>2</sup> [79], [6], [71], [69],[125].

Os métodos da primeira classe utilizam técnicas de otimização para resolver um problema não-linear separável, em que algumas das incógnitas entram linearmente e outras não-linearmente. Se não for dado um bom chute inicial, o método pode não convergir ou encontrar mínimos locais que não correspondem à solução procurada. Devido a estas dificuldades, abordaremos apenas alguns métodos invariantes a deslocamento. Os métodos invariantes ao deslocamento não usam explicitamente polinômios na sua formulação, eles são considerados aqui porque sua precisão depende dos coeficientes de um polinômio preditor. Isto será discutido no próximo capítulo.

---

<sup>2</sup>Dizer que  $H(\ell)$  é invariante ao deslocamento significa que o espaço coluna de  $H(\ell + 1)$  está contido no espaço coluna de  $H(\ell)$ .

Seja  $V_M$  a matriz de Vandermonde definida em (2.8), cujas colunas geram o espaço sinal de  $H(\ell)$ . Se  $\mathcal{A}$  é uma matriz formada por todas as linhas da matriz  $V_M$ , exceto última, e  $\mathcal{B}$  é formada por todas as linhas de  $V_M$ , exceto pela primeira, então  $\mathcal{A}Z = \mathcal{B}$ . A propriedade da invariância a deslocamentos da matriz de Hankel, assegura que se  $\mathcal{V}$  é outra matriz, cujas colunas geram o espaço coluna de  $H(\ell)$ , e se  $A$  e  $B$  são obtidas a partir de  $\mathcal{V}$  analogamente como  $\mathcal{A}$  e  $\mathcal{B}$  foram obtidas a partir de  $V_M$ , então existe uma matriz  $T$   $d \times d$ , chamada *matriz de transição*, tal que

$$AT = B. \quad (2.17)$$

**Teorema 2.1.4** *Os autovalores de  $T$  são  $\lambda_l = e^{(\alpha_l + i\omega_l)\Delta t}$ ,  $l = 1, \dots, d$ .*

**Demonstração:** Como as colunas de  $V_M$  e as colunas de  $\mathcal{V}$  geram  $\mathcal{S}_M$ , o subespaço sinal, existe uma matriz  $F \in \mathbb{C}^{d \times d}$ , não singular, tal que

$$\mathcal{V} = V_M F. \quad (2.18)$$

Decomponha  $\mathcal{V}$  e  $V_M$  como

$$\mathcal{V} = \begin{bmatrix} A \\ x^* \end{bmatrix} = \begin{bmatrix} y^* \\ B \end{bmatrix}, \quad V_M = \begin{bmatrix} V_{M-1} \\ e^* \Lambda^{M-1} \end{bmatrix} = \begin{bmatrix} e^* \\ V_{M-1} \Lambda \end{bmatrix},$$

onde  $x$  é a última linha de  $\mathcal{V}$  e  $y$  a primeira linha de  $\mathcal{V}$ . Usando esta decomposição e (2.18), segue que  $B = V_{M-1} \Lambda F$  e  $A = V_{M-1} F$ . Substituindo  $A$  e  $B$  em (2.17), vem que

$$(V_{M-1} F)T = V_{M-1} \Lambda F.$$

Como  $V_{N-1}^\dagger V_{N-1} = I$  (veja Apêndice A.4), da equação acima segue que  $T = F^{-1} \Lambda F$ . Isto mostra que os autovalores de  $T$  são os  $\lambda_l$ ,  $l = 1, 2, \dots, d$ .

□

O mesmo raciocínio pode ser feito usando uma matriz  $U \in \mathbb{C}^{N \times d}$ , cujas colunas geram o espaço linha de  $H(\ell)$  (chamado também de *subespaço sinal*). Neste caso, a equação (2.17) envolve uma matriz de transição semelhante a  $\Lambda^*$ .

As diferentes escolhas de  $\mathcal{V}$  e as diferentes formas de resolver (2.17), resultam em vários métodos do tipo subespaço. A seguir, são apresentados dois destes métodos.

## Método de Kung

O primeiro é um método bastante conhecido, o método de Kung [79], o qual foi criado em 1978, e é conhecido em Ressonância Magnética Nuclear, como HSVD. O método

está baseado na SVD de  $H(l)$ , a matriz  $\mathcal{V}$  é a matriz dos vetores singulares à esquerda, associados aos valores singulares não nulos. E a solução de (2.17) é dada no sentido dos quadrados mínimos lineares, isto é, a matriz de transição é calculada via pseudo inversão. Assim:

$$T_K = A^\dagger B. \quad (2.19)$$

Na prática, a pseudo inversa é eficientemente calculada sem ser necessário fazer nenhuma inversão de matriz. De fato, suponha que a SVD de  $H(\ell)$  seja dada por:

$$H(l) = U\Sigma V^* = [U_1 \ U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^* \\ V_2^* \end{bmatrix} = U_1 \Sigma_1 V_1^*, \quad (2.20)$$

onde  $U_1$ , e  $V_1$  são as matrizes dos vetores singulares associados aos valores singulares não nulos,  $\Sigma_1$  é uma matriz diagonal, cujos elementos da diagonal são  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_d > 0$ . Agora, escreva

$$U_1 = \begin{bmatrix} u^* \\ A \end{bmatrix},$$

onde  $u^*$  é a primeira linha de  $U_1$ , e observe que  $I = U_1^* U_1 = uu^* + A^* A$ . Usando este resultado e o Lema de Sherman Morrison (veja [59]), segue que

$$A^\dagger = (A^* A)^{-1} A^* = \left( I + \frac{uu^*}{1 - u^* u} \right) A^*, \quad (2.21)$$

o qual mostra que a matriz de transição em (2.19) pode ser calculada usando a relação (2.21).

## Método HTLS

Se a equação (2.17) for resolvida no sentido dos quadrados mínimos totais, obtém-se o algoritmo chamado **HTLS**, o qual extrai os autovalores do sinal da matriz:

$$T_S = -\hat{V}_{12}(\hat{V}_{22})^{-1}, \quad (2.22)$$

em que  $\hat{V}_{12}$ ,  $\hat{V}_{22}$ , são matrizes  $d \times d$ , obtidas da SVD da matriz aumentada

$$[A \ B] = \hat{U} \hat{\Sigma} \hat{V}^*, \quad \text{com } \hat{V} = \begin{bmatrix} \hat{V}_{11} & \hat{V}_{12} \\ \hat{V}_{21} & \hat{V}_{22} \end{bmatrix}.$$

Para verificar a relação (2.22) é suficiente notar que  $T_S$  é solução do sistema linear  $\hat{A}X = \hat{B}$ , em que

$$\hat{A} = \hat{U} \hat{\Sigma} \hat{V}^* [e_1 \dots e_d], \quad \text{e} \quad \hat{B} = \hat{U} \hat{\Sigma} \hat{V}^* [e_{d+1} \dots e_{2d}].$$

O método foi desenvolvido por S. Van Huffel, [71], e é muito utilizado em problemas de Ressonância Magnética Nuclear.

### 2.1.3 Recuperação de Harmônicos em Situações Reais

Os métodos descritos foram desenvolvidos a partir do sinal exato, isto é, na situação ideal onde não há ruídos. Na prática, os sinais são perturbados por erros nos instrumentos de medição, variação de temperatura, erros de arredondamento, etc. Ou seja, o problema deve ser resolvido a partir de  $\tilde{s}_j = h_l + \epsilon_j$ , e a matriz de informações é  $\tilde{H}(\ell) = H(\ell) + E$ . Nestas condições, nem os coeficientes preditores  $c_j$  (ou  $d_j$ ) e nem a matriz  $\mathcal{V}$  utilizada pelos métodos de subespaço podem ser calculados exatamente. No caso dos coeficientes preditores na direção inversa, o primeiro problema a ser resolvido é construir uma solução aproximada para a solução  $d^+ = H(\ell)^\dagger b$ , com  $b = H(\ell - 1)e_1$  do sistema (2.16) (veja o Teorema 2.1.3), a partir do sistema perturbado

$$\tilde{H}x = \tilde{b}, \quad (2.23)$$

em que  $\tilde{b} = \tilde{H}(\ell - 1)e_1 = b + \epsilon$ . Uma maneira de se resolver este problema é através da SVD de  $\tilde{H}(\ell)$ . A saber, se a SVD

$$\tilde{H}(\ell) = [\tilde{U}_1 \ \tilde{U}_2] \text{diag}(\tilde{\Sigma}_1, \tilde{\Sigma}_2) [\tilde{V}_1 \ \tilde{V}_2]^*,$$

é disponível, usando a melhor aproximação de posto  $d$  da matriz  $\tilde{H}(\ell)$  (veja [59, Teor. 2.5.3]), obtém-se que uma solução aproximada para o problema em questão é

$$\tilde{d}^+ = \tilde{V}_1 \tilde{\Sigma}_1^{-1} \tilde{U}_1^* \tilde{b}.$$

Prova-se que (veja Hansen [65], por exemplo)

$$\frac{\|d^+ - \tilde{d}^+\|_2}{\|d^+\|_2} \leq \frac{\kappa(H(\ell))}{1 - \|E\|_2 \|H^\dagger(\ell)\|_2} \left[ \frac{\|E\|_2}{\|H(\ell)\|_2} + \frac{\|\epsilon\|_2}{\|b\|_2} \right] + \|E\|_2 \|H^\dagger(\ell)\|_2, \quad (2.24)$$

em que  $\kappa(H(\ell)) = \sigma_1(H(\ell))/\sigma_d(H(\ell))$  é o número de condição da matriz  $H(\ell)$  (informações adicionais sobre condicionamento são dadas no Capítulo 3). De (2.24) segue que o erro relativo nos coeficientes  $d_j$  estimados pela SVD depende tanto do número de condição  $\kappa(H(\ell))$  quanto do nível de ruído nos dados. Se  $\kappa(H(\ell))$  é grande, as estimativas dos  $d_j$  podem ser dramáticas. Outra dificuldade é que a pseudo-inversa  $H^\dagger$  não depende continuamente dos dados, isto é, pequenas variações nos dados podem produzir grande perturbações na pseudo-inversa estimada (ver o exemplo da página 12 de

Bjork [26]). O problema de estimar o erro no cálculo de aproximações da matriz  $\mathcal{V}$  utilizadas pelos métodos do tipo subespaço, a partir da matriz de informações, será discutido mais adiante.

Como em ambas as classes de métodos apresentadas, Predição Linear e Métodos de Subespaço, o problema central a ser resolvido é um problema de autovalor, surge a pergunta natural: o que ocorre com os autovalores calculados nestas condições? Ou seja, qual a sensibilidade dos  $\lambda$ 's em função do erro  $E$ ? Respostas para esta questão serão dadas no próximo capítulo.

## 2.2 Problemas com Restrições de Igualdade

Problemas de otimização com restrições aparecem em aplicações como ajuste de superfícies, tomografia, sistemas mecânicos, etc., veja por exemplo, [49, 50, 51]. O objetivo aqui é apresentar um problema desse tipo e ilustrar que sua solução envolve um problema de autovalor polinomial matricial. Para tanto, sejam  $A \in \mathbb{R}^{n \times n}$  simétrica e  $b \in \mathbb{R}^n$ . Considere o seguinte problema de otimização

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \{x^T Ax - 2b^T x\}, \\ \text{Sujeito a } x^T x = a > 0. \end{aligned} \quad (2.25)$$

Mostraremos que (2.25) pode ser reduzido a um problema de autovalor polinomial quadrático através da técnica dos multiplicadores de Lagrange [55]. De fato, seja

$$\phi(x, \lambda) = x^T Ax - 2b^T x - \lambda(x^T x - a).$$

Calculando as derivadas em relação a  $x$  e  $\lambda$  segue que

$$Ax - \lambda x = b, \quad \text{e} \quad a = x^T x. \quad (2.26)$$

Se  $\lambda$  não é um autovalor de  $A$ , a primeira das equações em (2.26) pode ser escrita como

$$x = (A - \lambda I)^{-1}b.$$

Introduzindo a variável  $y = (A - \lambda I)^{-2}b$ , a restrição  $x^T x = a$  implica então que

$$x^T x = b^T (A - \lambda I)^{-2} b = b^T y,$$

e assim, tem-se que

$$(A - \lambda I)^2 y = b \iff (\lambda^2 + -2\lambda A + A^2) = b. \quad (2.27)$$

Agora como  $b^T y/a = 1$ , o vetor  $b$  pode ser escrito como

$$b = a^{-1} b b^T y,$$

portanto, a equação (2.27) é equivalente ao problema de autovalor polinomial quadrático

$$P_2(\lambda)y = (I\lambda^2 + A_1\lambda + A_0)y = 0, \quad (2.28)$$

em que

$$A_1 = -2A, \quad A_0 = (A^2 - a^{-1} b b^T).$$

A solução do problema de minimização (2.25) é dada então por  $x = (A - \lambda I)^{-1} b$ , sendo  $\lambda$  o menor autovalor do polinômio quadrático  $P_2(\lambda)$ , detalhes podem ser encontrados em [49].

## 2.3 Identificação de Sistemas Vibratórios

### 2.3.1 Introdução

Intuitivamente, o termo *sistema* refere-se a uma entidade (física, econômica, etc) capaz de produzir alguma reação  $y$  (*saída/output*) como consequência de um estímulo  $f$  (*entrada/input*), conforme Figura 2.4.

No texto, trataremos apenas de sistemas dinâmicos lineares descritos por equações de estado do tipo

$$\mathcal{S} : \begin{cases} \dot{x}(t) &= A_c x(t) + B_c u(t) \\ y(t) &= C_c x(t) + D_c u(t), \end{cases} \quad (2.29)$$

no caso contínuo, ou por

$$\mathcal{S} : \begin{cases} \mathbf{x}_{k+1} &= A_d \mathbf{x}_k + B_d \mathbf{f}_k \\ \mathbf{y}_k &= C_d \mathbf{x}_k + D_d \mathbf{f}_k, \end{cases} \quad (2.30)$$

no caso discreto. Nas equações acima,  $A_c, A_d$  são matrizes de ordem  $n$  (*matrizes do sistema*),  $B_c, B_d$  são  $n \times q$  (*matrizes de controle*),  $C_c, C_d$  são  $p \times n$  (*matrizes de observação*),

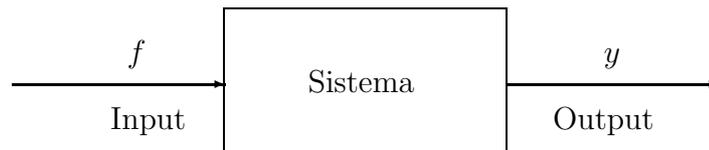


Figura 2.4: Um sistema genérico

e  $D_c, D_d$  são  $p \times q$  (matrizes de transmissão). O número  $n$  é conhecido como *ordem* do sistema. Por sua vez,  $x(t), y(t), f(t), t \geq 0$ , são funções vetoriais de dimensão compatível com as operações envolvidas, e  $\mathbf{x}_k \in \mathbb{R}^n, \mathbf{f}_k \in \mathbb{R}^q$  e  $\mathbf{y}_k \in \mathbb{R}^p$ . Assumiremos daqui em diante que os sistemas em consideração são estáveis, ou seja, que os autovalores  $\lambda_i(A)$  satisfazem a propriedade  $\text{real}(\lambda_i) < 0$ , no caso contínuo, e  $|\lambda_i| < 1$  para o caso discreto.

Um problema na análise de sistemas é identificar um sistema dinâmico a partir de informações provenientes de um par de seqüências finitas de sinais discretos de entrada e saída,  $\mathbf{f}_k \in \mathbb{R}^q$  e  $\mathbf{y}_k \in \mathbb{R}^p$ . Em certos casos, o problema é direcionado à busca de um conjunto de parâmetros que caracterizam a dinâmica do sistema, tais como frequências naturais de vibração, amortecimento, etc. Em outros casos, procura-se as matrizes  $A, B, C$  e  $D$  (ou outras equivalentes via alguma transformação de semelhança) que satisfaçam as relações entrada-saída do sistema. O problema alcança relevância em áreas tais como análise modal de estruturas mecânicas e engenharia aeroespacial, entre outras. A principal dificuldade de se trabalhar neste tipo de problema é que a ordem do sistema é desconhecida e os sinais  $\mathbf{f}_k, \mathbf{y}_k$  são contaminados por *ruídos*, pois são resultados de medidas de laboratório ou de experimentos.

Como o interesse é analisar sistemas vibratórios, adotaremos um modelo de amortecimento viscoso, descrito por um modelo linear invariante no tempo da forma

$$M\ddot{u}(t) + C\dot{u}(t) + Ku(t) = f(t), \quad t \geq 0, \quad (2.31)$$

o qual é obtido freqüentemente quando sistemas vibratórios reais, tais como pontes, prédios, fuselagem de foguetes, asa de avião, etc, são discretizados espacialmente. No modelo acima,  $M, C$  e  $K$  são matrizes simétricas  $n \times n$ , de massa, amortecimento e rigidez, respectivamente, e  $u$  e  $f$  são funções vetoriais que correspondem a deslocamento e forças associadas a cada uma das coordenadas físicas (ou generalizadas) introduzidas na discretização do sistema. Um exemplo de discretização de uma estrutura real na qual são utilizadas 20 coordenadas físicas é mostrada na Figura 2.5, detalhes podem ser encontrados em [6].

Observe que o sistema (2.31) pode ser escrito na forma (2.29) com  $A_c, B_c, C_c, D_c$  e vetor de estados  $x$  definidos por

$$A_c = \begin{bmatrix} 0 & I_n \\ -M^{-1}K & -M^{-1}C \end{bmatrix}, \quad B_c = \begin{bmatrix} 0 & M^{-1} \\ -M^{-1} & -M^{-1}CM^{-1} \end{bmatrix},$$

$$C_c = [I_n \ 0]_{n \times 2n}, \quad D_c = 0_{n \times n}, \quad \text{e} \quad x = \begin{bmatrix} u \\ \dot{u} \end{bmatrix}.$$

A informação dinâmica do sistema é descrita pelos auto-pares  $\{\mu_j, x_j\}$  do polinômio

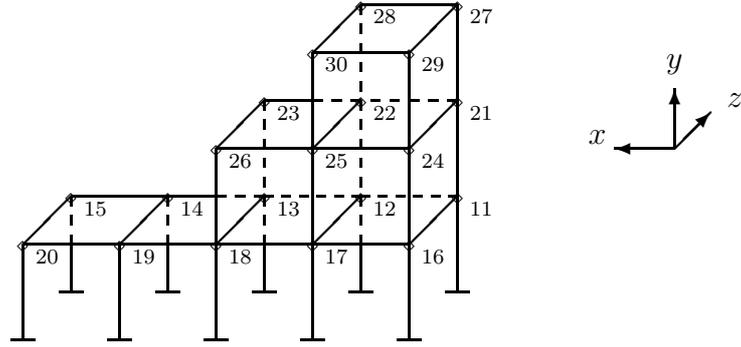


Figura 2.5: Protótipo de um prédio de 3 andares

quadrático associado:  $(M\mu^2 + C\mu + K)x = 0$ . Os modos de vibrar do sistema são descritos pelos autovalores  $x_j$  enquanto que as frequências de vibração  $\omega_j$  e os fatores de amortecimento  $\xi_j$  vêm dos autovalores  $\mu_j = -\alpha_j + i\omega_j$ ,  $\alpha_j > 0$ , através da relação

$$\xi_j = \frac{\alpha_j}{\sqrt{\alpha_j^2 + \omega_j^2}}.$$

### 2.3.2 Relações Entrada-Saída e o Problema de Identificação

Em vista de que na prática a análise é feita a partir de sinais de entrada e sinais de saída, uma primeira tarefa é descrever relações de entrada-saída envolvendo a informação dinâmica do sistema. Para tal, assumo o sistema em repouso. Tomando a transformada de Laplace em (2.31) segue

$$(Ms^2 + Cs + K)u(s) = f(s), \quad (2.32)$$

em que  $u(s) = \mathcal{L}(u(t))$ ,  $f(s) = \mathcal{L}(f(t))$ , com  $\mathcal{L}(\cdot)$  denotando o operador transformada de Laplace. A relação entrada-saída do sistema, no domínio de Laplace, é dada então por

$$u(s) = H(s)f(s), \quad (2.33)$$

em que

$$H(s) = (Ms^2 + Cs + K)^{-1}, \quad s \in \mathbb{C},$$

é a matriz *função de transferência do sistema*<sup>3</sup>. Quando a variável complexa  $s$  é restrita ao eixo imaginário, ou seja  $s = i\omega$ , a função de transferência é conhecida como *matriz de resposta em frequência* e denotada por  $H(i\omega)$ .

<sup>3</sup>Em aplicações práticas, a matriz de massa  $M$  é definida-positiva. Nesse caso, existem  $2n$  autovalores finitos em pares complexos conjugados (veja a Tabela 1.2), assim, a matriz  $H(s)$  é não singular para todo  $s$  diferente dos autovalores do polinômio quadrático.

A contraparte no domínio do tempo para  $\mathbf{H}(s)$  é a matriz resposta ao impulso unitário definida como

$$h(t) = \mathcal{L}^{-1}[\mathbf{H}(s)]. \quad (2.34)$$

Utilizando o teorema de convolução em (2.33) segue que a relação entrada-saída no domínio do tempo é dada então por [45, 73, 74]

$$u(t) = \int_0^t h(t - \tau) f(\tau) d\tau. \quad (2.35)$$

As funções de transferência  $\mathbf{H}(s)$  e a função de resposta ao impulso unitário  $h(t)$  concentram a informação dinâmica do sistema descrita pelos auto-pares  $\{\mu_j, x_j\}$  através das relações [45, 73]:

$$\mathbf{H}(s) = \sum_{j=1}^n \frac{x_j x_j^T}{s - \mu_j} + \frac{\bar{x}_j \bar{x}_j^T}{s - \bar{\mu}_j}, \quad (2.36)$$

$$h(t) = \sum_{j=1}^n x_j x_j^T e^{\mu_j t} + x_j \bar{x}_j^T e^{\bar{\mu}_j t}, \quad (2.37)$$

com a barra denotando conjugação complexa.

Como uma ilustração da relevância das relações entrada-saída, suponha que o sistema é excitado por uma força harmônica com frequência  $\omega_0$ :  $f(t) = e^{i\omega_0 t}$ . Substituindo a função força escolhida em (2.35), é fácil ver que uma solução particular para o sistema é dada por

$$u_p(t) = e^{i\omega_0 t} \left( \sum_{j=1}^n \frac{x_j^T v_0}{i\omega_0 - \mu_j} x_j + \sum_{j=1}^n \frac{\bar{x}_j^T v_0}{i\omega_0 - \bar{\mu}_j} \bar{x}_j \right). \quad (2.38)$$

Desta equação segue que se  $x_j^T v_0 \neq 0$  e  $\omega_0$  fica próxima de um autovalor  $\mu_j$ , o  $j$ -ésimo coeficiente em (2.38) cresce sem controle, fazendo com que a resposta do sistema se comporte com grandes oscilações, ou em termos técnicos, fazendo com que o sistema fique próximo do fenômeno conhecido como *ressonância*.

No caso de estruturas mecânicas tal como pontes, prédios, etc., o fenômeno de ressonância implica em vibrações descontroladas que podem danificar ou destruir completamente a estrutura. Um fato recente, envolvendo vibrações descontroladas numa estrutura real, ocorreu na cerimônia de abertura de uma ponte sobre o rio Thames (Londres, junho-2000), de 320 metros de comprimento, conhecida como a *ponte do milênio*. Uma

explicação do fenômeno, dada pelos construtores da ponte, foi que a estrutura foi excitada por uma força com uma frequência próxima de uma das frequências naturais de vibração do sistema (a ponte), gerada provavelmente pelo vento daquele dia e o caminhar quase sincronizado de um número grande de pessoas [37]. Curiosamente, a ponte foi fechada dois dias depois de sua abertura.<sup>4</sup>

Na prática, dispõe-se apenas de  $q$  sinais de entrada e  $p$  sinais de saída. Neste caso, a relação entrada-saída do sistema no domínio do tempo é dada através de uma função resposta ao impulso unitário  $p \times q$  do tipo

$$h(t) = \phi \operatorname{diag}(e^{\mu_1 t}, \dots, e^{\mu_n t}, e^{\bar{\mu}_1 t}, \dots, e^{\bar{\mu}_n t}) L, \quad (2.39)$$

em que

$$\phi = [\phi_1, \dots, \phi_n, \bar{\phi}_1, \dots, \bar{\phi}_n]_{p \times 2n}, \quad \phi_i \in \mathbb{C}^p, \quad i = 1 : n, \quad (2.40)$$

$$L^T = [l_1, \dots, l_n, \bar{l}_1, \dots, \bar{l}_n]_{q \times 2n}, \quad l_i \in \mathbb{C}^{1 \times q}, \quad i = 1 : n. \quad (2.41)$$

$\phi_j$  é o  $j$ -ésimo modo do sistema,  $l_j$  o  $j$ -ésimo vetor de fator de participação modal e  $\{\phi_j, \mu_j, l_j, \bar{\phi}_j, \bar{\mu}_j, \bar{l}_j\}$   $j = 1 : n$ , os parâmetros modais do sistema. Usualmente, dispõe-se de uma versão discreta da função  $h(t)$ , digamos  $h_k$ ,  $k = 0, 1, \dots, t_f$ , em intervalos igualmente espaçados do tempo, obtida a partir da inversa da função de resposta em frequência  $H(\omega_k)$  calculada através da inversa da transformada rápida de Fourier (FFT) [45, 73].

Dada uma seqüência finita de amostras  $\tilde{h}_k = h_k + \epsilon_k$ ,  $k = 0, 1, \dots, t_f$ , com  $\epsilon_k$  denotando incertezas na matriz  $h_k$ , o problema de identificação de parâmetros modais consiste em determinar estimativas dos parâmetros modais  $\{\phi_j, \mu_j, l_j, \bar{\phi}_j, \bar{\mu}_j, \bar{l}_j\}$ .

**Observação 2.3.1** Em vista de que a versão discreta de  $h(t)$  dada em (2.39) pode ser escrita como

$$h_k = \sum_{j=1}^n (\phi_j l_j^T \lambda_j^k + \bar{\phi}_j \bar{l}_j^T \bar{\lambda}_j^k) = \sum_{j=1}^n R_j e^{\mu_j \Delta t k} + \bar{R}_j e^{\bar{\mu}_j \Delta t k}, \quad (2.42)$$

com  $\lambda_j = e^{\mu_j \Delta t}$ ,  $\Delta t$  sendo a taxa de amostragem e  $R_j = \phi_j l_j^T$ ,<sup>5</sup> o problema de identificação de parâmetros modais é, em certo sentido, uma generalização do problema de recuperação

<sup>4</sup>Detalhes sobre a ponte do milênio e o fenômeno ocorrido são encontrados em

<http://www.arup.com/MillenniumBridge>.

<sup>5</sup>As matrizes  $R_j$  são matrizes de posto 1 conhecidas como *matriz de resíduos*.

de harmônicos descritos na Seção 2.1. Por causa disso, daqui em diante ambos os problemas serão ditos problemas HR (harmonical retrieval).

### 2.3.3 Abordagem Polinomial Matricial e de Subespaço

Seja  $H_{rs}(\ell)$  ( $\ell \geq 0$ ) uma matriz Hankel em blocos de ordem  $M \times N$  com  $M = r \times p$ ,  $N = s \times q$ , cuja entrada na posição bloco  $i, j$  é  $h_{\ell+i+j-2}$ :

$$H_{rs}(\ell) = \begin{bmatrix} h_{\ell} & h_{\ell+1} & \cdots & h_{\ell+s-1} \\ h_{\ell+1} & h_{\ell+2} & \cdots & h_{\ell+s} \\ \vdots & \vdots & \cdots & \vdots \\ h_{\ell+r-1} & h_{\ell+r} & \cdots & h_{\ell+r+s-2} \end{bmatrix}. \quad (2.43)$$

Usando o modelo (2.42), esta matriz pode ser fatorada como

$$H_{rs}(\ell) = \mathcal{O}_r \Lambda^{\ell} \mathcal{C}_s, \quad (2.44)$$

em que

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n, \bar{\lambda}_1, \dots, \bar{\lambda}_n), \quad (2.45)$$

e  $\mathcal{O}_r, \mathcal{C}_s$ , conhecidas respectivamente como matrizes de *observabilidade* e *controlabilidade*, são definidas como

$$\mathcal{O}_r = \begin{bmatrix} \phi \\ \phi\Lambda \\ \vdots \\ \phi\Lambda^{r-1} \end{bmatrix}, \quad \mathcal{C}_s = [L \ \Lambda L \ \cdots \ \Lambda^{s-1} L]. \quad (2.46)$$

O termo observabilidade é a propriedade do sistema (2.29) de permitir recuperar o estado inicial  $x(0)$  a partir de informações do sinal de saída  $y(t)$ . Quando isto não é possível, o sistema é dito *não observável*. Também, o sistema é dito *controlável* se para qualquer estado desejado do sistema, digamos  $w \in \mathbb{R}^n$ , sempre pode ser encontrada uma entrada (controle)  $f(t)$  que permita conduzir o sistema do estado inicial  $x(0)$  até o estado  $w$ . Em caso contrário, o sistema é dito *não controlável*.

**Teorema 2.3.2** *Assuma que o sistema (2.31) é observável e controlável. Então, sempre que  $\min\{r, s\} \geq 2n$ , tem-se*

$$\text{rank}(H_{r,s}(\ell)) = 2n, \quad \ell \geq 0. \quad (2.47)$$

**Demonstração:** A afirmação do teorema é uma consequência da fatoração (2.44) e do fato de que para sistemas controláveis e observáveis, tem-se

$$\text{rank}(\mathcal{O}_r) = \text{rank}(\mathcal{C}_s) = 2n,$$

veja, por exemplo, [34, Cap. 6].

□

Do teorema acima, segue que o espaço coluna de  $H_{r,s}(\ell)$  é gerado pelas colunas da matriz de observabilidade  $\mathcal{O}_r$  e que seu subespaço linha é gerado pelas colunas de  $\mathcal{C}_s^*$ . Assim, a  $j$ -ésima coluna de  $H_{r,s}(\ell)$ , digamos  $\check{h}_j(\ell)$ ,  $j = 1, \dots, s$ ; pode ser escrita como

$$\check{h}_j(\ell) = \mathcal{O}_r \Lambda^{\ell+j} L. \quad (2.48)$$

Como o vetor coluna em blocos  $\check{h}_s(\ell+1)$  também pertence ao espaço coluna de  $H_{r,s}(\ell)$ , pois o espaço coluna de  $H_{r,s}(\ell+1)$  também é gerado pelas colunas de  $\mathcal{O}_{r,s}$ , devem existir matrizes  $A_0, A_1, \dots, A_{s-1}$  todas de ordem  $q \times q$  tal que

$$\check{h}_1(\ell)A_0 + \check{h}_2(\ell)A_1 + \dots + \check{h}_{s-1}(\ell)A_{s-2} + \check{h}_s(\ell)A_{s-1} = \check{h}_s(\ell+1). \quad (2.49)$$

Usando (2.48), o fato de que  $\mathcal{O}_r$  é de posto completo, e propriedades de pseudo-inversas, segue que

$$\mathcal{O}_r \Lambda^\ell \mathcal{C}_s \begin{bmatrix} A_0 \\ A_1 \\ \vdots \\ A_{s-1} \end{bmatrix} = \mathcal{O}_r \Lambda^{\ell+s} L \iff \mathcal{C}_s \begin{bmatrix} A_0 \\ A_1 \\ \vdots \\ A_{s-1} \end{bmatrix} = \Lambda^s L.$$

A última igualdade pode ser rescrita como

$$LA_0 + \Lambda LA_1 + \dots + \Lambda^{s-2} LA_{s-1} + \Lambda^{s-1} LA_{s-1} = \Lambda^s L.$$

Multiplicando pela esquerda ambos os membros da equação anterior por  $e_j^T$  segue que

$$l_j^T I \lambda_j^s - l_j^T A_{s-1} \lambda_j^{s-1} + \dots - l_j A_1 \lambda_j - l_j A_0 = 0. \quad (2.50)$$

Esta equação assegura que  $\{\lambda_j, l_j\}$  são auto-pares (com  $l_j$  sendo autovetor à esquerda) do polinômio matricial

$$P_s(\lambda) = I \lambda^s - A_{s-1} \lambda^{s-1} + \dots - A_1 \lambda - A_0, \quad (2.51)$$

com os coeficientes  $A_j$ ,  $j = 0, \dots, s-1$  satisfazendo a equação matricial

$$H_{r,s}(\ell)X_A = \check{h}_s(\ell+1), \quad (2.52)$$

com

$$X_A = \begin{bmatrix} A_0 \\ A_1 \\ \vdots \\ A_{s-1} \end{bmatrix}. \quad (2.53)$$

Os  $q \cdot s - 2d$  autovalores restantes não têm significado físico e dependem da solução  $X_A$  escolhida entre as infinitas soluções que o sistema (2.52) possui.

O resultados teóricos acima são a base de vários métodos de identificação paramétrica, dois dos quais são descritos a seguir.

## Método das Referências Múltiplas

O método foi proposto por H. Vold e equipe [1, 127], e baseia-se no fato de que os fatores de participação modal  $l_j$  e os autovalores do sistema são auto-pares do polinômio  $P_s(\lambda)$  definido em (2.51). O método pode ser implementado como segue:

1. Para  $\ell$  fixo, digamos  $\ell = 0$ , encontrar coeficientes  $A_j$ ,  $j = 1: s-1$  a partir do sistema linear (2.52).
2. Encontrar os auto-pares  $\{\lambda_j, l_j\}$  (com  $l_j$  sendo autovetor à esquerda) do polinômio  $P_s(\lambda) = I\lambda^s + A_{s-1}\lambda^{s-1} + \dots + A_1\lambda + A_0$ , a partir da matriz companheira

$$C_A = \begin{bmatrix} 0 & 0 & \dots & 0 & -A_0 \\ I_q & 0 & \dots & 0 & -A_1 \\ 0 & I_q & \dots & 0 & -A_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_q & -A_{s-1} \end{bmatrix}. \quad (2.54)$$

3. Separar os auto-pares de interesse e encontrar a matriz de modos  $\phi$  usando a relação  $\phi\Lambda^k L = h_k$ .

A dificuldade do método é que além de precisar da solução de um problema de autovalor matricial de ordem potencialmente grande, a existência de  $s \cdot q - 2n$  autovalores sem significado físico dificulta a separação dos autovalores de interesse. Uma maneira de contornar as dificuldades acima é através de métodos de subespaço, como ilustramos a seguir.

## Método OPIA

O método OPIA (Optimized Pseudo Inverse Algorithm) foi desenvolvido por Bazán e Bavastri [6]. A ideia básica é que as matrizes Hankel em blocos  $H_{r,s}(\ell + 1)$  e  $H_{r,s}(\ell)$ , podem ser relacionadas a través da equação matricial

$$H_{r,s}(\ell + 1) = H_{r,s}(\ell)C_A, \quad \ell \geq 0, \quad (2.55)$$

com  $C_A$  sendo uma matriz companheira do tipo (2.54). Utilizando a fatoração (2.44) e propriedades de pseudo-inversas segue que

$$C_s C_A = \Lambda C_s. \quad (2.56)$$

Considere a decomposição SVD da matriz Hankel em bloco

$$H_{r,s}(\ell) = [U_1 \ U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} = U_1 \Sigma_1 V_1^T$$

em que  $\Sigma_1$  contém os valores singulares não nulos de  $H_{r,s}(\ell)$  e  $U_1$  e  $V_1$ , os vetores singulares associados.

Como as colunas de  $V_1$  geram o subespaço gerado pelas colunas de  $C_s$ , da relação (2.56) segue que

$$C_s C_A = \Lambda C_s \iff A_{BB} \doteq V_1^T C_A V_1 = (C_s V_1)^{-1} \Lambda (C_s V_1). \quad (2.57)$$

A relação acima mostra que o espectro da matriz  $A_{BB} \in \mathbb{C}^{2n \times 2n}$  é formado pelos autovalores do sistema, e que, se  $X$  é uma matriz de autovetores à esquerda da matriz  $A_{BB}$ , então os fatores de participação modal podem ser calculados como

$$L^T = [I_q \ 0]_{q \times q.s} V_1 X. \quad (2.58)$$

A proposta do método OPIA pode ser resumida como segue:

1. Para  $\ell$  fixo, digamos  $\ell = 0$ , calcule a SVD da matriz Hankel em bloco  $H_{r,s}(\ell)$  e estime a ordem do sistema através do número de valores singulares não nulos dessa matriz.
2. Calcule os autovalores do sistema através da matriz  $A_{BB}$  descrita em (2.57) e os fatores de participação modal usando (2.58).
3. Calcule os modos  $\phi$  analogamente ao método das referências múltiplas.

No caso de dados com ruído, a matriz de Hankel é de posto completo e as mesmas dificuldades do problema de recuperação de harmônicos são encontradas aqui. Porém, se o nível de ruído nos dados não for muito alto, a ordem do sistema pode ser descoberta através do número de valores singulares dominantes de matriz de Hankel,<sup>6</sup> desde que ela seja suficientemente sobredeterminada [6]. Muitos outros métodos do tipo subespaço em ambos os domínios, o domínio do tempo e o domínio da frequência, podem ser utilizados para resolver o problema de identificação. Uma boa referência com uma coletânea de métodos é Allemang [1].

A precisão do método de subespaço descrito aqui e de outros tal como ERA (Eigen-system Realization algorithm), no caso de dados com ruídos, dependem dos coeficientes do polinômio  $P_s(\lambda)$  utilizado pelo método das referências múltiplas.

## 2.4 Problema de Realização em Sistemas Dinâmicos

Considere um sistema dinâmico com  $q$  entradas  $\mathbf{f}_k \in \mathbb{R}^q$  e  $p$  saídas  $\mathbf{y}_k \in \mathbb{R}^p$ , descrito pelas equações de estado em (2.30). Suponha ainda que o sistema encontra-se em repouso. Então a resposta  $\mathbf{y}_k$  do sistema ao vetor de entrada  $\mathbf{f}_k$  é governada pela operação de convolução

$$\mathbf{y}_k = \sum_{i=1}^k G_{k-i} \mathbf{f}_i, \quad (2.59)$$

em que  $G_k \in \mathbb{R}^{p \times q}$ , conhecido como Parâmetro de Markov (ou resposta ao impulso unitário), é definido por

$$G_k = \begin{cases} D, & k = 0 \\ C A^{k-1} B, & k > 0. \end{cases} \quad (2.60)$$

Dada uma seqüência de sinais de entrada-saída  $\{\mathbf{f}_k, \mathbf{y}_k\}$ , o problema de realização consiste em determinar matrizes  $A, B, C$  e  $D$ , com  $A$  da menor ordem possível, tal que as relações de entrada-saída descritas pelas equações de estado (2.30) são satisfeitas. O problema é difícil porque podem ser encontradas muitas quádruplas  $\{A, B, C, D\}$ , cada uma delas com matrizes  $A$  de diferentes ordens, satisfazendo as relações entrada-saída.

A solução do problema envolve a computação de certos subespaços dominantes de matrizes estruturadas. Consideraremos aqui os dois casos apresentados a seguir.

---

<sup>6</sup>Os valores singulares  $\sigma_1, \dots, \sigma_d$  da matriz  $A \in \mathbb{R}^{m \times n}$  são dominantes se

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_d \gg \sigma_{d+1} \geq \sigma_{d+2} \geq \dots \sigma_n.$$

### 2.4.1 Realização a partir da Resposta ao Impulso Unitário

Assuma conhecida a seqüência finita do sinal  $G_k$ ,  $k = 0, 1, \dots, K$ . A solução do problema começa com a construção de uma matriz Hankel em bloco  $H_{r,s}(\ell)$ , com entradas na posição bloco  $i, j$  igual a  $G_{\ell+i+j-1}$  (veja (2.43)). Analogamente ao caso da matriz Hankel em bloco utilizada no problema de identificação de parâmetros modais, a matriz  $H_{r,s}(\ell)$  pode ser fatorada como

$$H_{r,s}(\ell) = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{r-1} \end{bmatrix} A^\ell [B \ AB \ \dots \ A^{s-1}B] = \mathcal{O}A^\ell\mathcal{C}, \quad \ell \geq 0, \quad (2.61)$$

em que  $\mathcal{O}_r$  e  $\mathcal{C}_s$  são as matrizes de observabilidade e controlabilidade, respectivamente. A propriedade do sistema ser observável e controlável implica então que

$$\text{rank}(H_{r,s}(\ell)) = n, \quad (2.62)$$

para  $\ell \geq 0$  e  $r, s \geq n$ . Esta propriedade, bem como a decomposição (2.61), são a base para uma série de métodos de realização [6, 12, 34, 54, 79, 88, 120, 125, 135]. Daqui em diante, assumiremos sempre que o sistema é controlável e observável.

### Método de Zeiger-Mac Ewen

Observando que  $\mathcal{O}$  e  $\mathcal{C}$  são de posto completo, igual a  $n$ , temos que para  $\ell = 1$

$$A = \mathcal{O}^\dagger H_{r,s}(1) \mathcal{C}^\dagger, \quad (2.63)$$

em que o símbolo  $\dagger$  denota a pseudo-inversa de uma matriz. Agora considere a decomposição SVD da matriz Hankel em bloco

$$H_{r,s}(0) = [U_1 \ U_2] \text{diag}(\Sigma_1, \Sigma_2) [V_1 \ V_2]^T = U_1 \Sigma_1 V_1^T,$$

em que  $\Sigma_1$  contém os valores singulares não nulos de  $H_{r,s}(\ell)$  e  $U_1$  e  $V_1$ , os vetores singulares associados. Observe ainda que esta decomposição pode ser rescrita similarmente como em (2.61):

$$H_{rs}(0) = (U_1 \Sigma_1^{1/2}) (\Sigma_1^{1/2} V_1^T).$$

Assim, fica claro que podemos tomar

$$\hat{\mathcal{O}} = U_1 \Sigma_1^{1/2}, \quad \text{e} \quad \hat{\mathcal{C}} = \Sigma_1^{1/2} V_1^T, \quad (2.64)$$

como estimativas para as matrizes  $\mathcal{O}$  e  $\mathcal{C}$ , respectivamente <sup>7</sup>. O método de Zeiger-Mac Ewen utiliza essas estimativas em (2.63) [135]. Daí vem que

$$\hat{A} = \Sigma_1^{-1/2} U_1^T H_{r,s}(1) V_1 \Sigma_1^{-1/2} \quad (2.65)$$

é uma matriz semelhante a  $A$ . As matrizes restantes para resolver o problema da realização são obtidas da seguinte forma

$$\hat{B} = \Sigma_1^{1/2} V_1^T(1 : n, 1 : p) \quad (2.66)$$

$$\hat{C} = U_1(1 : q, 1 : n) \Sigma_1^{1/2} \quad (2.67)$$

$$\hat{D} = G_0. \quad (2.68)$$

## Método de Kung

Este método está baseado na invariância ao deslocamento (shift-invariant) das matrizes de observabilidade e controlabilidade [120, 125, 135]. Especificamente, considere a estimativa para a matriz de observabilidade definida em (2.64) e defina  $\bar{U}$  como sendo a submatriz de  $\hat{O}$  formada pelas primeiras  $r - 1$  linhas bloco. Então a invariância ao deslocamento da matriz de observabilidade nos permite deduzir que

$$\bar{U} \Sigma_1^{1/2} \hat{A} = \underline{U} \Sigma_1^{1/2}. \quad (2.69)$$

Agora, como  $\bar{U}$  é de posto completo (igual a  $n$ ), a equação acima pode ser resolvida via pseudo-inversão. Assim, obtemos que

$$\hat{A} = \Sigma_1^{-1/2} \bar{U}^\dagger \underline{U} \Sigma_1^{1/2}. \quad (2.70)$$

Um fato interessante relacionado com a fórmula acima é que a pseudo-inversa envolvida pode ser calculada sem a necessidade de inverter matrizes. Informação adicional sobre realização de sistemas pode ser encontrada em [79].

A precisão dos métodos de subespaços descritos aqui e de outros tal como OPIA e ERA (Eigensystem Realization algorithm), entre outros, no caso de dados com ruídos, dependem dos coeficientes do polinômio matricial  $P_s(\lambda)$  utilizado pelo método das referências múltiplas (veja, (2.51), detalhes podem ser vistos em Bazán [15]).

---

<sup>7</sup>Pode-se provar que existe uma matriz não singular  $T$  tal que  $\mathcal{O} = \hat{O}T$  e  $\mathcal{C} = T^{-1}\hat{C}$ .

### 2.4.2 Realização a partir de Informação Input-Output

O problema de construir uma realização do sistema, utilizando apenas os sinais de entrada e saída é mais complicado do que o caso anterior. Assim, apresentamos apenas uma discussão introdutória do assunto. Começamos com a observação que os sinais  $\mathbf{f}_k, \mathbf{y}_k, k = 0, 1, \dots$ , satisfazem a equação [120, 88]

$$\mathbf{Y} = \mathbf{\Gamma}\mathbf{X} + \mathbf{H}\mathbf{F}, \quad (2.71)$$

nde

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_k & \mathbf{y}_{k+1} & \cdots & \mathbf{y}_{k+j-1} \\ \mathbf{y}_{k+1} & \mathbf{y}_{k+2} & \cdots & \mathbf{y}_{k+j} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{y}_{k+i-2} & \mathbf{y}_{k+i-1} & \cdots & \mathbf{y}_{k+i+j-3} \\ \mathbf{y}_{k+i-1} & \mathbf{y}_{k+i} & \cdots & \mathbf{y}_{k+i+j-2} \end{bmatrix}, \quad \mathbf{\Gamma} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{i-1} \end{bmatrix}, \quad (2.72)$$

$$\mathbf{H} = \begin{bmatrix} G_0 & 0 & 0 & \cdots & 0 \\ G_1 & G_0 & 0 & \cdots & 0 \\ G_2 & G_1 & G_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ G_{i-2} & G_{i-3} & G_{i-4} & \cdots & G_0 \end{bmatrix}, \quad (2.73)$$

$$\mathbf{X} = [\mathbf{x}_k \ \mathbf{x}_{k+1} \ \cdots \ \mathbf{x}_{k+j-1}],$$

e  $\mathbf{F}$  é uma matriz Hankel em bloco construída similarmente a  $\mathbf{Y}$ .

A idéia básica do método apresentado a seguir consiste inicialmente em encontrar uma matriz  $\check{\mathbf{F}}$  com a propriedade  $\mathbf{F}\check{\mathbf{F}} = 0$ . Se isto for possível, de (2.71) vem que

$$\mathbf{Y}\check{\mathbf{F}} = \mathbf{\Gamma}\mathbf{X}\check{\mathbf{F}}. \quad (2.74)$$

Desta relação deduz-se que se

$$\mathbf{A-1} \quad \text{rank}(\mathbf{\Gamma}) = n,$$

$$\mathbf{A-2} \quad \text{rank}(\mathbf{X}\check{\mathbf{F}}) = n,$$

então

$$\text{rank}(\mathbf{Y}\check{\mathbf{F}}) = n. \quad (2.75)$$

A condição **A-1** é verdadeira quando o sistema é observável. Embora **A-2** seja restritiva, pois depende muito do sinal de entrada, ela é frequentemente satisfeita quando o sinal

de entrada é um ruído Gaussiano com média zero [120]. Também, é importante observar que a existência da matriz  $\check{\mathbf{F}}$  exige que  $\mathbf{F}$  seja uma matriz com muito mais colunas do que linhas ( $j \gg l \times i$ ).

Resumindo, se **A-1** e **A-2** forem satisfeitas, então a matriz  $\mathbf{Y}\check{\mathbf{F}}$  é de posto incompleto e seu subespaço coluna é gerado pelas colunas da matriz de observabilidade  $\mathbf{\Gamma}$ . Esta informação pode ser utilizada para detectar a ordem  $n$  do sistema, bem como para calcular estimativas de  $A$  (em verdade, o que é estimado é uma matriz  $\tilde{A}$  semelhante a  $A$ ). De fato, seja  $\mathcal{U}$  uma matriz  $(i \times l) \times n$  cujas colunas formam uma base ortonormal para  $\mathbf{\Gamma}$ , calculada usando a SVD da matriz  $\mathbf{Y}\check{\mathbf{F}}$ , por exemplo. Defina  $\mathcal{U}_1$  como sendo a matriz formada pelas primeiras  $j - 1$  linhas-bloco de  $\mathcal{U}$  e  $\mathcal{U}_2$ , a matriz formada pelas últimas  $j - 1$  linhas-bloco de  $\mathcal{U}$ . Então, prova-se que existe uma matriz  $\tilde{A}$ ,  $n \times n$ , semelhante a  $A$ , satisfazendo a equação

$$\mathcal{U}_1 \tilde{A} = \mathcal{U}_2. \quad (2.76)$$

Daí é imediato que  $\tilde{A} = \mathcal{U}_1^\dagger \mathcal{U}_2$ . Note também que a matriz  $\tilde{C}$  pode ser estimada de imediato. O cálculo das matrizes  $\tilde{B}$  e  $\tilde{D}$  é mais complexo e fica para uma outra oportunidade.

Um comentário deve ser colocado sobre a solução do problema em situações reais. O fato é que como os sinais de entrada e saída são sempre contaminados por ruídos, mesmo com o sinal de entrada satisfazendo **A-2**, a propriedade (2.75) dificilmente é satisfeita. Assim, na prática, a ordem do sistema é estimada a partir do número de valores singulares dominantes da matriz  $(\mathbf{Y}\check{\mathbf{F}})$  e a matriz  $\mathcal{U}$  é construída então usando-se os vetores singulares à direita, associados a esses valores singulares.

Finalmente, assim como o problema de realização a partir de parâmetros de Markov está fortemente relacionada a polinômios matriciais preditores, pode ser demonstrado que realização a partir de sinais de entrada e saída está relacionada a funções racionais envolvendo polinômios matriciais; porém, o tópico foge do escopo deste texto; veja [34, Cap. 7].

## 2.5 Miscelânea

Talvez a área mais imediata na qual aparecem polinômios matriciais é aquela que trata de resolver sistemas de equações diferenciais ordinárias de ordem superior ou igual a dois (veja, por exemplo, Gohberg [57]). Além disso porém, existe uma grande variedade de aplicações em outras áreas envolvendo problemas de autovalor polinomial matricial, e um dos objetivos desta seção é fornecer alguns exemplos sobre algumas aplicações envolvendo o tópico em estudo, bem como referências bibliográficas de aplicações que se encontram espalhadas em diferentes ramos da ciência.

### 2.5.1 Solução de EDP's

Problemas envolvendo equações diferenciais parciais (EDP's) aparecem em áreas como vibro-acústica, mecânica dos fluidos, análise modal de estruturas mecânicas, etc. A solução desses problemas passa por um processo de discretização, o qual, em geral, conduz a problemas de álgebra linear tal como a solução de sistemas lineares e/ou problemas de autovalor. A título de exemplo, considere a vibração livre de uma corda presa em ambos os extremos descrita por

$$\begin{cases} u_{tt} + \epsilon a(x)u_t = \Delta u, & x \in [0, \pi], \quad \epsilon > 0, \\ u(t, 0) = u(t, \pi) = 0. \\ u(0, x) = h(x), \quad u_t(0, x) = 0, \end{cases} \quad (2.77)$$

Conhecemos, através do Método de Fourier que a solução da equação acima é

$$u(x, t) = \sum_{k=1}^{\infty} q_k(t) \text{sen}(kx), \quad (2.78)$$

para determinadas funções  $q_k(t)$ . Aproximando agora a solução até o seu  $n$ -ésimo termo,

$$u(x, t) = \sum_{k=1}^n q_k(t) \sin(kx) \quad (2.79)$$

e aplicando o Método de Galerkin,<sup>8</sup> obtemos um problema de valor inicial (PVI) envolvendo um sistema de equações diferenciais de segunda ordem do tipo:

$$\begin{cases} M\ddot{q}(t) + \epsilon C\dot{q}(t) + Kq(t) = 0, \\ q(0) = \hat{h} \in \mathbb{R}^n, \quad \dot{q}(0) = 0 \in \mathbb{R}^n. \end{cases} \quad (2.80)$$

em que  $\hat{h}$  e o vetor nulo  $0$  vêm da condição inicial (2.77),

$$q(t) = [q_1(t), \dots, q_n(t)]^T, \quad M = \frac{\pi}{2} I_n, \quad K = \frac{\pi}{2} \text{diag}(1, 2^2, \dots, i^2, \dots, n^2),$$

---

<sup>8</sup>A ideia básica do método de Galerkin para resolver aproximadamente um problema  $Ax = y$ , com  $A : X \rightarrow Y$  um operador linear e  $X, Y$  espaços com produto interno, é construir soluções aproximadas

$$x_N = \sum_{j=1}^N c_j \psi_j,$$

com  $\psi_j$ ,  $j = 1 :: N$ , uma base de um subespaço  $X_N \subset X$ , de tal modo que o resíduo  $r = y - Ax_N$  é ortogonal a  $X_N$ ; detalhes do método são encontrados em [99, 110].

e  $C$  com entradas  $c_{k,j}$  definidas por

$$c_{k,j} = \int_0^\pi a(x) \sin(kx) \sin(jx) dx.$$

Como todas as matrizes envolvidas são simétricas, e  $M$ ,  $K$  definidas-positivas, o polinômio matricial associado:

$$P(\lambda) = M\lambda^2 + \epsilon C\lambda + K,$$

têm auto-pares que vêm em pares complexos conjugados (veja, Tabela 1.2). Assim, se  $X \in \mathbb{C}^{n \times 2n}$  é a matriz de autovetores à direita de  $P(\lambda)$  e  $\Lambda \in \mathbb{C}^{2n \times 2n}$  a matriz de autovalores correspondentes, então a solução  $q(t)$  procurada é da forma

$$q(t) = X e^{\Lambda t} \alpha, \quad \alpha \in \mathbb{C}^{2n},$$

com  $\alpha$  sendo a solução do sistema linear

$$\begin{bmatrix} X \\ X\Lambda \end{bmatrix} \alpha = \begin{bmatrix} \hat{h} \\ 0 \end{bmatrix}.$$

## 2.5.2 Processamento de Sinais

Várias aplicações tal como processamento digital de voz, processamento de sinais em bio-medicina, localização de objetos por radar, etc, utilizam modelos auto-regressivo da forma ([36, 76])

$$x_t = - \sum_{k=1}^p \alpha_k x_{t-k} + \epsilon_t, \quad t = 1: n, \quad (2.81)$$

para prever informações futuras a partir de um conjunto de informações conhecidas. No modelo acima,  $x_t \in \mathbb{R}^p$  é uma série temporal de dados,  $n$  o número de dados,  $p$  a ordem do modelo,  $\epsilon_t$  é um processo Gaussiano com média zero e variância  $\sigma_\epsilon$ , e os  $\alpha$ 's, a serem determinados, são parâmetros que concentram a dinâmica do processo de predição.

Os parâmetros  $\alpha$ 's podem ser determinados através das equações de Yule-Walker

$$R\alpha = -r^x,$$

em que  $r^x = [r_1^x, \dots, r_p^x]^T$  com  $r_k^x$  sendo funções auto-correlação da série temporal  $x_t$  :

$r_k^x = \mathbf{E}(x_t, x_{t+k})$ , e  $R \in \mathbb{R}^{p \times p}$  uma matriz de Toeplitz (a matriz de auto-correlação) <sup>9</sup> cujas entradas são auto-correlações do tipo  $r_{|j-i|}^x$ , veja [36, 76].

Na prática, a série temporal disponível é da forma

$$y_t = x_t + \omega_t, \quad t = 1:n,$$

em que  $\omega_t$  é um ruído branco não correlacionado com variância desconhecida  $\sigma_w$ . Neste caso, a solução obtida através das equações de Yule-Walker pode ser incorreta e levar a conclusões errôneas. Uma maneira de contornar esta dificuldade é através de um problema de autovalor polinomial matricial quadrático simétrico, o qual pode ser descrito como segue:

Considere as equações de Yule-Walker compensadas definidas por [36]

$$(S - \lambda B)v = 0, \tag{2.82}$$

em que  $S \in \mathbb{R}^{(p+n) \times (p+n)}$ ,  $n \geq p$ , com entradas  $S_{i,j} = r_{i-j+1}^y$ , e

$$B = \begin{bmatrix} 0 & I_p \\ 0 & 0 \end{bmatrix}.$$

Neste caso, as incógnitas são  $v = [1, \alpha_1, \dots, \alpha_p]^T$  e  $\alpha$  é uma estimativa da variância desconhecida  $\sigma_y^2$ . Multiplicando pela esquerda ambos os membros de (2.82) por  $(S - \lambda B)^T$  obtém-se um problema de autovalor polinomial matricial quadrático do tipo

$$P(\lambda)v \doteq (A_2\lambda^2 + A_1\lambda + A_0)v = 0,$$

em que

$$A_2 = B^T B, \quad A_1 = -(S^T B + B^T S), \quad A_0 = S^T S.$$

Então, prova-se que as estimativas dos parâmetros  $\alpha$ 's são obtidas das componentes do autovetor  $v$  de  $P(\lambda)$  associado ao autovalor de menor valor absoluto [36].

Uma referência mais recente sobre o assunto, mas envolvendo polinômios matriciais de ordem superior a 2, é [75].

---

<sup>9</sup>E denota o operador *valor esperado* (expectation operator). A autocorrelação  $r_k^x$  é usualmente estimada como

$$r_k^x = \frac{1}{n} \sum_{t=1}^n x_t x_{t-k}.$$

### 2.5.3 Aplicações Diversas

Conforme comentado acima, uma grande diversidade de aplicações em ramos tecnológicos de interesse contemporâneo exige a solução ou análise de problemas de autovalor. Em áreas como análise modal experimental, engenharia aero-espacial, engenharia elétrica, acústica, etc, entre outras, o envolvimento com problemas de autovalor generalizado e/ou polinomial matricial quadrático é muito freqüente; aplicações relevantes podem ser encontradas em [25, 29, 37, 45, 60, 72, 73, 82, 90].

Outro problema relacionado com polinômios matriciais quadráticos aparece no cálculo de funções de resposta em freqüência  $H(\omega)$  de sistemas vibratórios, como em problemas de propagação eletromagnética, por exemplo [45, 47, 72, 73, 132]. O cálculo de  $H(\omega)$  está baseado na relação (veja (2.32))

$$(M\omega^2 + C\omega + K)H(\omega) = b(\omega),$$

com  $M, C, K$  e  $b(\omega) \in \mathbb{C}^n$  como dados de entrada, e  $\omega$  numa faixa de freqüências usualmente grande. Se a dimensão  $n$  das matrizes envolvidas não é muito grande, o problema pode ser resolvido através de uma linearização  $A - \omega B$  do polinômio matricial  $P(\lambda) = M\lambda^2 + C\lambda + K$ . Para  $n$  grande, outras abordagens são preferíveis, veja [80], por exemplo.

Problemas de autovalor polinomial matricial de grau superior também são freqüentes em várias áreas. A título de exemplo, um problema em aero-acústica que precisa da solução de um problema de autovalor polinomial matricial é reportado em [3]. Já em mecânica dos fluídos, em conexão com equações de Orr-Sommerfeld, o polinômio matricial utilizado é de grau 4 [29]. Muitas outras aplicações envolvendo problemas de autovalor polinomial matricial podem ser encontrados em [4] e [122].

# Capítulo 3

## Sensibilidade de Autovalores

Apresentamos uma discussão da sensibilidade do problema de autovalor polinomial matricial, baseada no conceito de condicionamento como medida de sensibilidade. Duas abordagens são apresentadas: uma que explora o conceito de número de condição do autovalor, derivado a partir do próprio polinômio, e outra baseada na teoria de perturbação de autovalores de matrizes não simétricas, introduzida por Wilkinson [131] em 1963, aplicada a matrizes companheiras em blocos associadas aos polinômios em questão.

### 3.1 Noções de Condicionamento

Uma característica da solução de problemas através de métodos computacionais é que o processo é desenvolvido utilizando aritmética de ponto flutuante. Como consequência disso, informações do problema são representadas apenas aproximadamente e erros nos dados de entrada são inevitáveis. Portanto, dependendo do problema em análise, as soluções computadas podem ser pouco acuradas ou talvez inaceitáveis.

Uma tarefa importante no processo de solução de cada problema portanto, é discutir ou analisar o comportamento da solução do problema devido a pequenas variações nos dados de entrada. Ou seja, em situações práticas, uma análise do grau de sensibilidade do problema a pequenas variações nos dados de entrada é de fundamental importância.

Do ponto de vista teórico, um problema pode ser idealizado como uma função  $f$ ,  $f : X \rightarrow Y$ , com  $X, Y$  sendo espaços normados ( $\mathbb{R}^n$ ,  $\mathbb{R}^m$ , ou  $\mathbb{C}^n$ , por exemplo), chamados espaço de entrada e de saída (ou de dados e de resposta), respectivamente, de modo que, se  $x \in X$  é considerado como dados de entrada e  $y = f(x) \in Y$  a resposta do problema, então o que deve-se analisar é a sensibilidade da função  $f$ , em  $x$ , a pequenas variações nas informações de entrada  $x$ .

Um problema é dito *bem condicionado* se pequenas perturbações em  $x$  produzem pequenas perturbações em  $y$ . Um problema é dito *mal condicionado* se pequenas per-

turbações em  $x$  produzem grandes perturbações em  $y$ . Obviamente, o significado de *grande* e *pequeno* depende muito do problema e/ou da aplicação sob estudo. O assunto é melhor compreendido analisando-se um número associado ao problema conhecido como *número de condição*, o resto do capítulo é dedicado exclusivamente a seu estudo.

Consideremos primeiro o caso  $X = Y = \mathbb{R}$ , ou seja  $f$  é uma função real de uma variável real. Vamos assumir que  $f$  é duas vezes continuamente diferenciável em  $x \neq 0$  e  $y = f(x) \neq 0$ . Se  $\delta x$  denota a perturbação de  $x$  e  $\delta y = f(x + \delta x) - f(x)$  é a perturbação correspondente em  $y$ , então o teorema de Taylor garante que

$$\delta y = f(x + \delta x) - f(x) = f'(x)\delta x + \mathcal{O}(\delta x)^2.$$

Esta relação mostra que o erro absoluto em  $y$  devido a variações de tamanho  $|\delta x|$  em  $x$  pode ser grande se  $|f'(x)|$  é muito grande. Ou seja, a relação acima mostra que o erro absoluto em  $y$  depende de  $|f'(x)|$ . Agora, se o interesse é medir variações relativas, a expressão acima pode ser reescrita como

$$\frac{\delta y}{y} = \frac{xf'(x)}{f(x)} \cdot \frac{\delta x}{x} + \mathcal{O}(\delta x),$$

a qual mostra que a variação relativa em  $y$  devido a erros relativos em  $x$  depende da quantidade

$$\left| \frac{xf'(x)}{f(x)} \right|.$$

A conclusão que pode ser tirada desta discussão é que se o problema for descrito por um função  $f : \mathbb{R} \rightarrow \mathbb{R}$ , então o número de condição de  $f$  no ponto  $x$ , pode ser definido como

$$\kappa(f)(x) = \begin{cases} \lim_{\delta \rightarrow 0} \sup_{|\delta x| \leq \delta} \frac{|\delta y|}{|\delta x|}, & \text{para erros absolutos} \\ \lim_{\delta \rightarrow 0} \sup_{|\delta x| \leq \delta} \frac{|\delta y|}{|y|} / \frac{|\delta x|}{|x|}, & \text{para erros relativos,} \end{cases} \quad (3.1)$$

e que este número de condição pode ser calculado como

$$\kappa(f)(x) = \begin{cases} |f'(x)|, & \text{para erros absolutos} \\ \left| \frac{xf'(x)}{f(x)} \right|, & \text{para erros relativos.} \end{cases} \quad (3.2)$$

Para o caso de serem  $X = \mathbb{R}^m$ ,  $Y = \mathbb{R}^n$ ,  $m, n$  arbitrários, a função  $f$  é descrita por

um conjunto de funções  $f_j : \mathbb{R}^m \rightarrow \mathbb{R}$ ,  $j = 1 : n$ , tal que

$$y = f(x) \Leftrightarrow y_j = f_j(x_1, \dots, x_m), \quad j = 1 : n,$$

e as variações em  $x$  e  $y$ , sejam absolutas ou relativas, são medidas utilizando normas vetoriais. Por exemplo, a variação relativa em  $x$  devido à variação  $\delta x$  é

$$\frac{\|\delta x\|}{\|x\|}, \quad \delta x = [\delta x_1, \dots, \delta x_m]^T.$$

Seja  $\delta y_j = f_j(x + \delta x) - f_j(x)$  a variação da  $j$ -ésima componente devido a variações em todas as variáveis  $x_i$ ,  $i = 1 : m$ . Então o teorema de Taylor garante que

$$\delta y_j = f_j(x + \delta x) - f_j(x) \approx \sum_{i=1}^n \frac{\partial f_j}{\partial x_i} \delta x_i.$$

Portanto, ao menos aproximadamente, segue que

$$\begin{aligned} |\delta y_j| &\leq \sum_{i=1}^n \left| \frac{\partial f_j}{\partial x_i} \right| |\delta x_i| \leq \max_i |\delta x_i| \sum_{i=1}^n \left| \frac{\partial f_j}{\partial x_i} \right| \\ &\leq \max_i |\delta x_i| \cdot \max_j \sum_{i=1}^n \left| \frac{\partial f_j}{\partial x_i} \right|, \end{aligned}$$

e, como esta desigualdade vale para  $j = 1 : n$ , em particular vale para  $\max_j |\delta y_j|$ . Utilizando a norma matricial  $\|\cdot\|_\infty$ , esta desigualdade pode ser escrita como

$$\|\delta y\|_\infty \leq \|\delta x\|_\infty \|Jf\|_\infty,$$

com  $Jf$  sendo a matriz Jacobiana de  $f$ .

A desigualdade acima sugere que o número de condição  $\kappa(f)(x)$  pode ser definido de maneira análoga como em (3.1), substituindo as barras de valor absoluto por normas vetoriais apropriadas, e que para o caso de considerarmos erros absolutos, o número de condição  $\kappa(f)(x)$  é dado pela norma matricial da matriz Jacobiana  $\|Jf\|$ . Um resultado análogo a aquele dado em (3.2), é imediato neste caso e dado por  $\|x\| \|Jf\| / \|f\|$ .

**Observação 3.1.1** Alertamos porém, que em certos casos, o número de condição envolvendo normas da matriz Jacobiana pode sobestimar ou subestimar a sensibilidade do problema. Uma outra abordagem é considerar números de condição individuais associados a cada variável  $y_j$  e provenientes de analisar variações em apenas uma variável  $x_i$ . Por

exemplo, utilizando o resultado descrito em (3.2), o número de condição do problema, na variável (ou resposta)  $y_j$ , devido apenas a perturbações em  $x_i$ , é

$$\kappa(f_j)(x_i) = \frac{\left| x_i \frac{\partial f_j}{\partial x_i} \right|}{|f_j|}. \quad (3.3)$$

Isto gera uma matriz  $n \times m$  de números de condição individuais, digamos  $\Upsilon$ , e um número de condição global pode ser definido utilizando alguma norma matricial de  $\Upsilon$ . Um exemplo que mostra o uso desta abordagem será apresentado depois. Agora vamos formalizar a definição para o número de condição de um problema  $f$  geral com dados de entrada  $x$  e resposta  $y$ .

**Definição 3.1.2** O número de condição do problema  $f$ , no ponto  $x$ , no caso de erros absolutos, é definido como

$$\kappa(f)(x) = \lim_{\delta \rightarrow 0} \sup_{\|\delta x\| \leq \delta} \frac{\|\delta f\|}{\|\delta x\|}, \quad (3.4)$$

enquanto que para o caso de erros relativos, ele é definido por

$$\kappa(f)(x) = \lim_{\delta \rightarrow 0} \sup_{\|\delta x\| \leq \delta} \frac{\|\delta f\|}{\|f(x)\|} / \frac{\|\delta x\|}{\|x\|}. \quad (3.5)$$

### 3.1.1 Raízes de um Polinômio

O problema aqui é investigar a sensibilidade do problema de calcular as raízes de um polinômio de grau  $n$ ,

$$p(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_1t + a_0, \quad a_0 \neq 0.$$

Por simplicidade, vamos assumir que a raiz  $\nu$  é uma raiz simples, ou seja,

$$p(\nu) = 0, \quad p'(\nu) \neq 0.$$

Neste caso, o problema é encontrar  $\nu$  utilizando os coeficientes  $a$ 's como dados de entrada. Formalmente isto pode ser descrito por uma função

$$\nu : \mathbb{R}^n \rightarrow \mathbb{C}, \quad \nu = \nu(a), \quad a = [a_0, a_1, \dots, a_{n-1}]^T,$$

definida implicitamente através de

$$[\nu(a)]^n + a_{n-1}[\nu(a)]^{n-1} + \cdots + a_j[\nu(a)]^j + \cdots + a_1[\nu(a)] + a_0 = 0.$$

Calculando a derivada parcial de  $\nu$  em relação a  $a_j$  vem a identidade

$$n[\nu(a)]^{n-1} \frac{\partial \nu}{\partial a_j} + (n-1)a_{n-1}[\nu(a)]^{n-2} \frac{\partial \nu}{\partial a_j} + \cdots + ja_j[\nu(a)]^{j-1} \frac{\partial \nu}{\partial a_j} + [\nu(a)]^j + \cdots + a_1 \frac{\partial \nu}{\partial a_j} = 0,$$

da qual segue,

$$p'(\nu) \frac{\partial \nu}{\partial a_j} + \nu^j = 0,$$

e assim (já que  $p'(\nu) \neq 0$ ),

$$\frac{\partial \nu}{\partial a_j} = \frac{-\nu^j}{p'(\nu)}.$$

Utilizando este resultado junto com aquele descrito em (3.2), vemos que o número de condição da raiz  $\nu$ , no caso de erros relativos, considerando apenas variações no coeficiente  $a_j$ , é

$$\kappa(\nu)(a_j) = \frac{\left| a_j \frac{\partial \nu}{\partial a_j} \right|}{|\nu|} = \frac{|a_j| |\nu|^j}{|\nu p'(\nu)|}. \quad (3.6)$$

Este número é geralmente grande, indicando assim que o problema de calcular as raízes de um polinômio pode ser muito sensível a pequenas variações nos coeficientes  $a$ 's. Existe um exemplo famoso devido a Wilkinson [131] que ilustra este fenômeno. Considere o polinômio

$$p(t) = \prod_{j=1}^{20} (t-j) = (t-1)(t-2) \cdots (t-20).$$

A Jacobiana neste caso é

$$\begin{aligned} J(\nu) &= \left[ \frac{\partial \nu}{\partial a_0}, \frac{\partial \nu}{\partial a_1}, \dots, \frac{\partial \nu}{\partial a_{n-1}} \right] \\ &= \frac{1}{p'(\nu)} [1, \nu, \nu^2, \dots, \nu^{n-1}]. \end{aligned}$$

O número de condição da raiz  $\nu$ , na norma-2, no caso de erros absolutos, é,

$$\kappa(\nu)(a) = \frac{1}{|p'(\nu)|} \sqrt{1 + |\nu|^2 + |\nu|^4, \dots, + |\nu|^{2(n-1)}}, \quad (3.7)$$

enquanto no caso de erros relativos, também na norma-2, este número é

$$\kappa(\nu)(a) = \frac{\sqrt{|a_0|^2 + |a_1|^2 + \dots + |a_{n-1}|^2} \sqrt{1 + |\nu|^2 + \dots + |\nu|^{2(n-1)}}}{|\nu p'(\nu)|}. \quad (3.8)$$

Ambos os números são potencialmente grandes, indicando portanto que o problema de calcular as raízes do polinômio de Wilkinson é muito sensível a pequenas variações nos coeficientes. Uma ilustração deste fenômeno é apresentado na Figura 3.1, na qual mostra-se o comportamento das raízes de um polinômio com coeficientes  $\tilde{c}_j$ , obtidos perturbando-se os coeficientes  $c_j$  do polinômio de Wilkinson através de

$$\tilde{c}_j = c_j(1 + \epsilon * r_j),$$

em que os  $r_j$  são números aleatórios normalmente distribuídos, com média zero e variância 1. O gráfico mostra o resultado de 50 realizações com  $\epsilon = 10^{-10}$ .

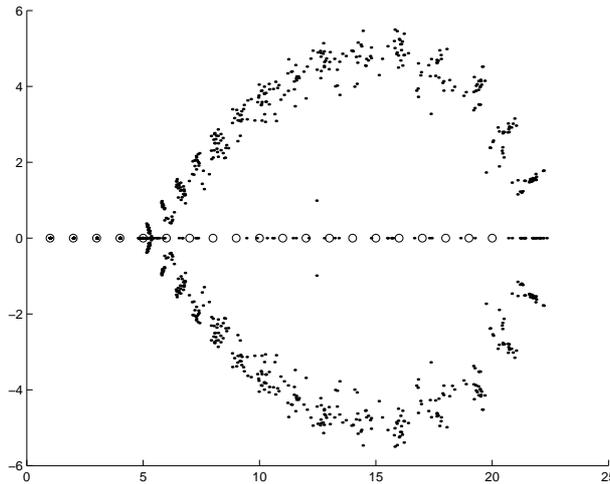


Figura 3.1: Raízes do polinômio de Wilkinson. o: raízes exatas, •: raízes do polinômio com coeficientes  $\tilde{c}_j$ , obtidos perturbando os coeficientes  $c_j$  do polinômio de Wilkinson.

Prova-se que a raiz mais sensível é  $\nu = 15$ . Para esta raiz, o número de condição (3.6) em relação a variações apenas em  $a_{15}$  é

$$\kappa(\nu)(a_{15}) \approx \frac{1.67 \times 10^9 \times 15^{14}}{5!14!} \approx 5.1 \times 10^{13}.$$

Finalmente, observamos que seguindo a abordagem de calcular números de condição individuais, conforme sugerido pela observação acima, um número de condição global para o caso de erros relativos pode ser estimado como

$$\bar{\kappa}(\nu)(a) = \|\kappa(\nu)(a_0), \dots, \kappa(\nu)(a_{n-1})\|.$$

Utilizando a norma-1 seque que

$$\bar{\kappa}(\nu)(a) = \frac{1}{|\nu p'(\nu)|} \sum_{i=1}^n |a_i| \nu_i^i,$$

a qual é menor do que aquela dada em (3.8). Para o caso das raízes do polinômio de Wilkinson de grau  $n$ , para  $n$  grande, prova-se que o número de condição do primeiro e última raiz (ordenadas crescentemente), são [53, Cap. 1, Sec. 3]

$$\bar{\kappa}(\nu_1)(a) \approx n^2, \quad \text{e}$$

$$\bar{\kappa}(\nu_n)(a) \approx \frac{1}{2 - \sqrt{2}\pi n} \left( \frac{\sqrt{2} + 1}{\sqrt{2} - 1} \right)^n.$$

### 3.1.2 Solução de um Sistema de Equações Lineares

Dados uma matriz  $A \in \mathbb{R}^{n \times n}$  e um vetor  $b \in \mathbb{R}^n$ , o problema é calcular a solução do sistema

$$Ax = b.$$

Por simplicidade, vamos assumir que apenas  $b$  é sujeito a erros. Neste caso, o problema pode ser formulado através da função  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  definida por

$$x = f(b) = A^{-1}b.$$

Ou seja, as componentes de  $b$  são os dados de entrada, e as componentes  $x$  os de saída.

É imediato que a Jacobiana neste caso é  $Jf(b) = A^{-1}$ . Portanto, utilizando os resultados da discussão acima, o número de condição, no caso de erros relativos, no vetor  $b$ , é

$$\kappa(f)(b) = \frac{\|b\| \|A^{-1}\|}{\|x\|} = \frac{\|b\| \|A^{-1}\|}{\|A^{-1}b\|}.$$

Isto mostra que o problema de resolver o sistema linear  $Ax = b$  pode ser muito sensível a pequenas variações em  $b$  quando  $\|A^{-1}\|$  é muito grande.

Utilizando propriedades das normas matriciais, segue facilmente que

$$\kappa(f)(b) \leq \|A\| \|A^{-1}\|.$$

Esta cota superior é atingida quando  $b$  é um múltiplo do vetor singular à esquerda, associado ao maior valor singular da matriz  $A$ ,  $\sigma_1(A)$ . Propriedades e definição de valores singulares de uma matriz são encontradas no apêndice A.3.

O número  $\|A\|\|A^{-1}\|$  é conhecido como *número de condição* da matriz  $A$  e denotado por  $\kappa(A)$ . Ele aparece como um indicador da sensibilidade de vários problemas em Álgebra linear numérica. Para o caso  $A \in \mathbb{R}^{m \times n}$ , o número de condição de  $A$  é definido como

$$\kappa(A) = \|A\|\|A^\dagger\|, \quad (3.9)$$

No caso da norma-2, com  $A$  de posto  $r$ , ele é dado como

$$\kappa_2(A) = \|A\|_2\|A^\dagger\|_2 = \frac{\sigma_1(A)}{\sigma_r(A)}. \quad (3.10)$$

### 3.1.3 Autovalores de uma Matriz

Neste caso, o problema pode ser descrito por uma função

$$\begin{aligned} f : \mathbb{C}^{n \times n} &\rightarrow \mathbb{C} \\ A &\mapsto f(A) = \lambda, \end{aligned}$$

sujeito a  $Ax = \lambda x$ ,  $x \neq 0$ . Esta é uma função que depende não-linearmente das entradas da matriz.

Para matrizes não simétricas (não Hermitianas), a análise do condicionamento de autovalores simples pode ser desenvolvido na maneira usual, mas em geral o problema é mal condicionado, ou seja, pequenas variações nas entradas da matriz podem produzir grandes perturbações nos autovalores. Para ilustrar este fato, considere as matrizes

$$A = \begin{bmatrix} 1.000 & 1.000 \\ 0 & 1.000 \end{bmatrix}, \quad \text{e} \quad \tilde{A} = \begin{bmatrix} 1.000 & 1.000 \\ 0.001 & 1.000 \end{bmatrix}.$$

Os autovalores destas matrizes são, respectivamente,  $\{1, 1\}$ , e  $\{0, 2\}$ . Isto mostra que uma variação de 0.01% em  $A$  produziu uma variação de 100 % nos autovalores, mostrando que o problema de autovalor matricial para matrizes não simétricas pode de fato ser muito mal condicionado.

Teoricamente, o número de condição do problema de autovalor matricial pode ser descrito como segue. Seja  $\lambda$  um autovalor simples de  $A$  com autovetores associados à direita e esquerda  $x$  e  $y$ , respectivamente. Se  $\delta\lambda$  é a variação em  $\lambda$  devido a uma variação  $\delta A$  em  $A$ , e se assumimos que o autovetor associado ao autovalore  $\lambda + \delta\lambda$  é  $x + \delta x$ , então

$$(A + \delta A)(x + \delta x) = (\lambda + \delta\lambda)(x + \delta x).$$

Desenvolvendo os produtos indicados, observando que  $Ax = \lambda x$ , e desprezando termos de segunda ordem, vem

$$A\delta + \delta Ax = \lambda\delta x + \delta\lambda x.$$

Agora, como  $y^*A\delta x = y^*\lambda\delta x$ , multiplicando ambos os membros da equação acima por  $y^*$ , segue que

$$\delta\lambda = \frac{y^*\delta Ax}{y^*x}.$$

A desigualdade Cauchy-Schwarz garante agora

$$\frac{|\delta\lambda|}{\|\delta A\|_2} \leq \frac{\|x\|_2\|y\|_2}{|y^*x|}.$$

Tomando limite quando  $\|\delta A\|_2 \rightarrow 0$ , segue que o número de condição do autovalor simples  $\lambda$  satisfaz

$$\kappa(\lambda, A) \leq \lim_{\|\delta A\|_2 \rightarrow 0} \frac{|\delta\lambda|}{\|\delta A\|_2} \leq \frac{\|x\|_2\|y\|_2}{|y^*x|}. \quad (3.11)$$

A cota superior em (3.11) foi introduzida por Wilkinson[131]. Se  $\lambda$  é simples, o número  $(\|x\|_2\|y\|_2)/|y^*x|$  é determinado univocamente, e portanto está bem definido. Este número mede de fato a sensibilidade do autovalor  $\lambda$  a pequenas variações em  $A$ .

A análise de condicionamento para o caso de autovalores múltiplos é mais difícil porque a função  $f$  é contínua mas não diferenciável. Assim, a abordagem descrita acima não é aplicável, detalhes podem ser encontrados em [33].

A situação muda drasticamente quando a matriz é simétrica (ou Hermitiana). Neste caso, independentemente da multiplicidade do autovalor, o problema é sempre bem condicionado [59].

## 3.2 Condicionamento de Autovalores de Polinômios Matriciais

A análise descrita nesta seção é restrita a polinômios matriciais regulares. Seja

$$\tilde{P}_m(\lambda) = \lambda^m + \tilde{A}_{m-1}\lambda^{m-1} + \dots + \tilde{A}_0$$

um polinômio matricial cujos coeficientes  $\tilde{A}_l$  são aproximações dos coeficientes  $A_l$  do polinômio  $P_m(\lambda)$  tal que

$$\|\delta A_0, \delta A_1, \dots, \delta A_{m-1}\|_2 \leq \delta, \quad \delta A_l = \tilde{A}_l - A_l, \quad l = 0 : m - 1. \quad (3.12)$$

Seja  $\lambda + \delta\lambda$  um autovalor de  $\tilde{P}_m(\lambda)$  e  $x + \delta x$  o autovetor associado. Então, seguindo a definição geral dada em (3.4), o número de condição do autovalor  $\lambda$ , no caso de erros absolutos, pode ser dado como segue.

**Definição 3.2.1** Assuma que as perturbações  $\delta A_l$  satisfazem (3.12), então o número de condição de um autovalor simples de  $P_m(\lambda)$ , no caso de erros absolutos, pode ser definido como [90, 121]

$$\kappa(\lambda, P) = \lim_{\delta \rightarrow 0} \sup \left\{ \frac{|\delta \lambda|}{\delta} \right\} \quad (3.13)$$

$$\text{Sujeito a } \tilde{P}_m(\lambda + \delta \lambda)(x + \delta x) = 0.$$

Claramente, o número  $\kappa(\lambda, P)$  dificilmente pode ser calculado através de (3.13). Esta dificuldade pode ser contornada utilizando o teorema seguinte.

**Teorema 3.2.2** *Seja  $\lambda$  um autovalor do polinômio matricial  $P_m(\lambda)$  e  $x, y \in \mathbb{C}^n$  os autovetores associados, à direita e esquerda, respectivamente. Então o número de condição  $\kappa(\lambda, P)$  é dado por*

$$\kappa(\lambda, P) = \eta \frac{\|y\|_2 \|x\|_2}{|y^* P'_m(\lambda)x|}, \quad (3.14)$$

em que,

$$\eta = \sqrt{1 + |\lambda|^2 + |\lambda|^4 + \dots + |\lambda|^{2(m-1)}}.$$

**Demonstração:** Primeiro observe que a restrição em (3.13) é equivalente a

$$(P_m(\lambda + \delta \lambda) + \delta P_m(\lambda + \delta \lambda))(x + \delta x) = 0,$$

em que  $\delta P_m(\lambda) = \delta A_m \lambda^m + \delta A_{m-1} \lambda^{m-1} + \dots + \delta A_1 \lambda + \delta A_0$ .

Expandindo  $P_m(\lambda + \delta \lambda)$  em series de Taylor e conservando termos de até primeira ordem, segue que

$$\delta \lambda P'_m(\lambda)x + P_m(\lambda)\delta x + \delta P_m(\lambda)x = \mathcal{O}(\delta^2).$$

Multiplicando pela esquerda pelo autovetor a esquerda  $y^*$ , segue que

$$\delta \lambda y^* P'_m(\lambda)x + y^* \delta P_m(\lambda)x = \mathcal{O}(\delta^2).$$

Daí, é imediato que

$$\delta \lambda = -\frac{y^* \delta P_m(\lambda)x}{y^* P'_m(\lambda)x} + \mathcal{O}(\delta^2). \quad (3.15)$$

Agora, em vista de que  $\delta A_m = 0$ , e de que

$$\delta P_m(\lambda)x = [\delta A_{m-1}, \delta A_{m-2}, \dots, \delta A_0] \begin{bmatrix} \lambda^{m-1}x \\ \lambda^{m-2}x \\ \vdots \\ \lambda x \\ x \end{bmatrix},$$

utilizando a desigualdade de Cauchy-Schwarz, segue que

$$y^* \delta P_m(\lambda)x \leq \|y\|_2 \|\delta A_0, \delta A_1, \dots, \delta A_{m-1}\|_2 \sqrt{1 + \|\lambda\|^2 + \|\lambda\|^4 + \dots + \|\lambda\|^{2(m-1)}} \|x\|_2.$$

Substituindo esta desigualdade em (3.15) e a seguir, utilizando a restrição (3.12), segue que

$$\frac{|\delta\lambda|}{\delta} \leq \eta \frac{\|x\|_2 \|y\|_2}{|y^* P'_m(\lambda)x|}. \quad (3.16)$$

Agora, se  $E = yx^*/(\|y\|_2 \|x\|_2)$  e as perturbações  $\delta A_j$  são definidas como

$$\delta A_{j-1} = \frac{\delta}{\nu} \lambda^{j-1} E, \quad j = 1 : m,$$

então, em vista de que  $\|E\|_2 = 1$ , fica claro que

$$\|[\delta A_{m-1}, \delta A_{m-2}, \dots, \delta A_0]\|_2 = \delta.$$

Este resultado garante que a cota superior em (3.16) pode ser atingida, ou seja, as perturbações acima, tornam (3.16) uma igualdade. A prova do teorema termina então, tomando limite o nessa igualdade. □

Uma outra maneira de se medir a sensibilidade de um autovalor simples é através do número de condição introduzido por Wilkinson[131].

**Definição 3.2.3** Seja  $\lambda$  um autovalor simples de  $P_m(\lambda)$  (e portanto da matriz companheira em blocos  $C_A$  associada a  $P_m(\lambda)$ ), e sejam  $r, \ell$ , autovetores à direita e esquerda de  $C_A$ , associados ao autovalor  $\lambda$ . Então o número de condição do autovalor  $\lambda$  segundo Wilkinson [131], é definido por

$$\kappa(\lambda, C_A) = \frac{\|\ell\|_2 \|r\|_2}{|\ell^* r|}. \quad (3.17)$$

Para polinômios matriciais gerais, ambos os números de condição  $\kappa(\lambda, P)$  e  $\kappa(\lambda, C_A)$  são de tamanhos comparáveis, mas potencialmente grandes. Além disso ambos precisam de informações contidas nos autovetores (em geral não disponíveis). Entretanto, conforme veremos mais adiante, estimativas interessantes podem ser deduzidas no caso do polinômio estar associado a problemas HR.

Os números de condição descritos acima descrevem a sensibilidade de cada autovalor individualmente, podendo existir autovalores com números de condição muito diferentes. Ou seja, alguns autovalores podem ser muito mais sensíveis a perturbações do que outros.

### Exemplo 3.2.4 Sensibilidade de um Problema de Autovalor Quadrático

Este exemplo considera o polinômio

$$P_2(\lambda) = \lambda^2 I + A_1 \lambda + A_0$$

associado ao sistema de  $n$  equações diferenciais de segunda ordem

$$M\ddot{q}(t) + \epsilon C\dot{q}(t) + Kq(t) = 0. \quad (3.18)$$

Ou seja, aqui:  $A_1 = \epsilon M^{-1}C$ , e  $A_0 = M^{-1}K$ .

Neste caso, devido ao fato do sistema ser proveniente da discretização de uma EDP (ver Cap. 2), a ordem  $n$  deve ser geralmente elevada para garantir boas soluções aproximadas do problema. As matrizes  $M$  e  $K$  são definidas por

$$M = \frac{\pi}{2}I, \quad K = \frac{\pi}{2}\text{diag}(1^2, 2^2, \dots, n^2),$$

enquanto a matriz  $C$  tem entradas tal que

$$c_{kj} = \begin{cases} 0 & \text{se } k + j \text{ for ímpar,} \\ 12\pi \left[ \frac{1}{(k+j)^4} - \frac{1}{(k-j)^4} \right] & \text{se } k + j \text{ for par e } k \neq j, \\ \frac{\pi^5}{60} - \frac{2.7\pi}{2} + \frac{3\pi}{4k^2} & \text{se } k = j. \end{cases} \quad (3.19)$$

O número  $\epsilon$  determina a quantidade de amortecimento presente no sistema e governa a sensibilidade do problema de autovalor: sabe-se que problemas mais amortecidos são em geral mais sensíveis a pequenas modificações no sistema. Apresentamos resultados numéricos para  $n = 80$ , e dois valores diferentes de  $\epsilon$ .

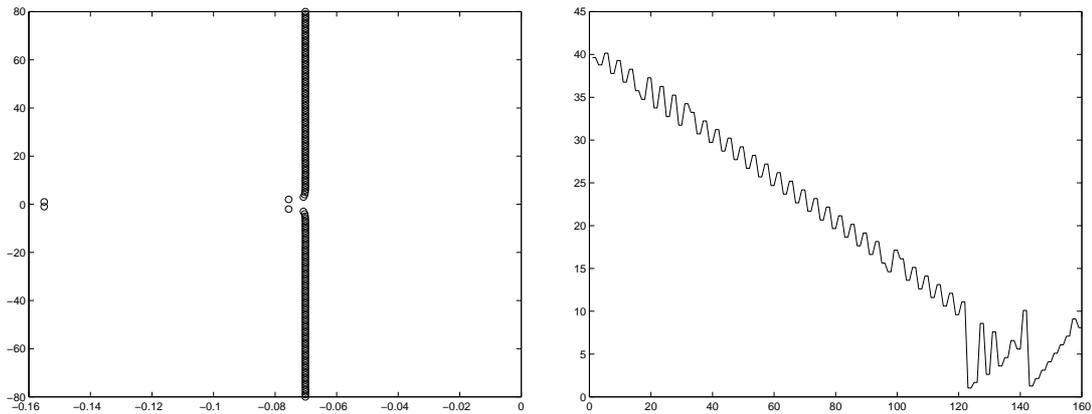


Figura 3.2: Esquerda: Autovalores de  $P_2(\lambda)$ . Direita:  $\kappa(\lambda, C_A)$ ,  $\epsilon = 0.1125$

Quando  $\epsilon$  é pequeno (neste caso  $\epsilon = 0.1125$ ), comprova-se de fato que os autovalores do polinômio são bem condicionados (ver Figura 3.2).

Já para  $\epsilon$  grande ( $\epsilon = 11.25$ ), verifica-se que a sensibilidade dos autovalores aumenta. Isto pode ser vista na Figura 3.3, a qual ilustra números de condição muito grandes (da ordem de  $10^7$ ).

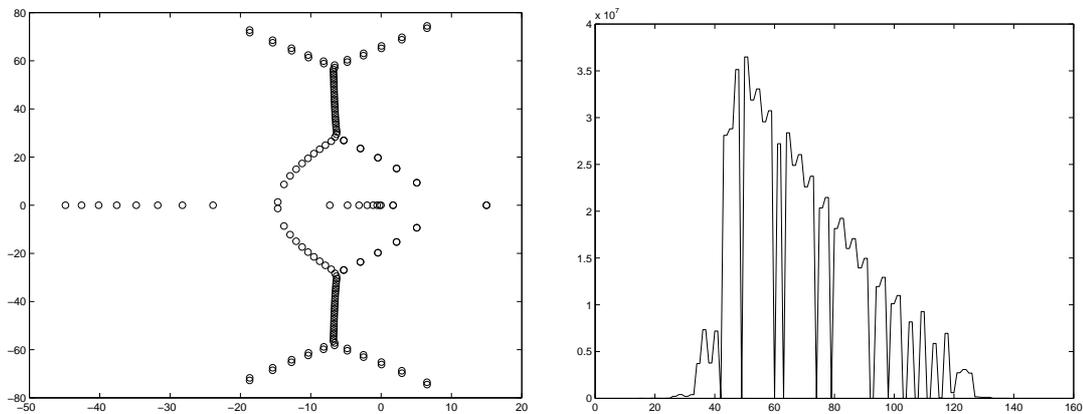


Figura 3.3: Esquerda: Autovalores de  $P_2(\lambda)$ . Direita:  $\kappa(\lambda, C_A)$ ,  $\epsilon = 11.25$

Finalmente, é importante considerar um número de condição que descreva a sensibilidade do problema de autovalor matricial de uma maneira global conforme dado na próxima definição.

**Definição 3.2.5** Seja  $A = X\Lambda X^{-1} \in C^{m \times m}$  uma decomposição de Jordan de  $A$ . O

número de condição de Jordan nas normas espectral e de Frobenius é definido por

$$\kappa_2(X) = \inf\{\|X\|_2\|X^{-1}\|_2\} \quad (3.20)$$

$$\kappa_F(X) = \inf\{\|X\|_F\|X^{-1}\|_F\} \quad (3.21)$$

Um resultado importante que relaciona o número de condição global com os números de condição individuais, cuja prova pode ser encontrada em [131], é descrito no próximo teorema.

**Teorema 3.2.6** *Se  $A = X\Lambda X^{-1}$  é diagonalizável, então*

$$\kappa_F(X) = \sum_{j=1}^m \kappa(\lambda_j, A).$$

## Capítulo 4

# Análise de Sensibilidade e de Erro em problemas HR

A característica principal dos problemas HR é que tanto os polinômios como as matrizes envolvidas são resultado de computações preliminares utilizando informações aproximadas (informações contaminadas por ruídos). O objetivo então é estimar o afastamento máximo dos autovalores computados em relação àqueles em que as informações são livres de ruídos.

A análise de sensibilidade e de estimativa de erro é desenvolvida em duas partes. Primeiro, são apresentadas duas abordagens para o caso escalar, e a seguir, as mesmas abordagens são estendidas para o caso polinomial matricial, enfatizando as vantagens da abordagem polinomial matricial sobre a escalar.

### 4.1 Sensibilidade de Autovalores: Caso Polinomial Escalar

Os resultados teóricos apresentados dependem fortemente do condicionamento de uma matriz retangular de Vandermonde, cujos elementos são potências dos autovalores de interesse. Começaremos portanto, com uma análise do condicionamento dessa matriz.

#### 4.1.1 Condicionamento de Matrizes de Vandermonde

Seja  $W_M$  a matriz de Vandermonde  $d \times M$  definida por

$$W_M = \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{M-1} \\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{M-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \lambda_d & \lambda_d^2 & \cdots & \lambda_d^{M-1} \end{bmatrix}. \quad (4.1)$$

O objetivo então é deduzir limitantes para  $\kappa(W_M) = \|W_M\|_2 \|W_M^\dagger\|_2$ . Algumas estimativas preliminares e simples de serem deduzidas são descritas a seguir.

**Teorema 4.1.1** *Seja  $W_M$  uma matriz de Vandermonde como em (4.1), com  $\lambda_j \neq \lambda_i$ , para  $j \neq i$ ,  $|\lambda_j| < 1$ . Então, para  $M > d$ ,  $\|W_M^\dagger\|_2$  decresce monotonicamente com  $M$ , e este decréscimo é estrito ao longo da subsequência de inteiros  $\{d, 2d, \dots\}$ . Além disso, se  $M = p \cdot d$ ,  $p \geq 1$ , então*

$$\|W_M^\dagger\|_2 \leq \frac{\|W_d^{-1}\|_2}{\sqrt{1 + \beta^{2d} + \beta^{4d} + \dots + \beta^{2(p-1)d}}}, \quad (4.2)$$

em que  $\beta = \min|\lambda_j|$ ,  $j = 1, \dots, d$ .

**Demonstração:** A demonstração pode ser encontrada em [11].

□

Observamos que apesar desta estimativa poder ser muito pessimista se  $W_d$  for muito mal condicionada (ou seja,  $\|W_M^\dagger\|_2$  muito grande), o resultado é interessante porque mostra que  $\|W_M^\dagger\|_2$  decresce monotonicamente com  $M$ . Ou seja, o resultado leva a esperar que  $\kappa_2(W_M)$  melhore a medida que  $M$  cresce. Por outro lado, ele pode ser muito útil para fins teóricos, como pode ser visto a seguir.

**Teorema 4.1.2** *Seja  $\hat{f}_M$  a solução de norma mínima do sistema subdeterminado*

$$W_M f = \Lambda^M e,$$

onde  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_d)$ , e é o vetor  $e = [1, 1, \dots, 1]^T \in \mathbb{R}^d$ . Suponha que todos os  $\lambda_j$  satisfazem  $|\lambda_j| = 1$ , ou  $|\lambda_j| < 1$ , então sempre que  $M \rightarrow \infty$  tem-se

$$\|\hat{f}_M\|_2 \rightarrow 0.$$

**Demonstração:** Observe que

$$\|\hat{f}_M\|_2 = \|W_M^\dagger \Lambda^M e\|_2 \leq \|W_M^\dagger\|_2 \|\Lambda^M e\| \leq \|W_M^\dagger\|_2 \sqrt{d} \alpha^M,$$

onde  $\alpha = \max_j |\lambda_j|$ ,  $j = 1, 2, \dots, d$ .

Utilizando o Teorema 4.1.1 segue que

$$\|\hat{f}_M\|_2 \leq \frac{\|W_d^{-1}\|_2}{\sqrt{1 + \beta^{2d} + \beta^{4d} + \dots + \beta^{2(N/d-1)d}}} \sqrt{d} \alpha^M.$$

Tomando o limite nesta desigualdade quando  $M \rightarrow \infty$ , vem que  $\|\hat{f}_M\|_2 \rightarrow 0$ .

□

O vetor  $\hat{f}_M$  descrito acima, é conhecido em problemas de predição linear como *vetor de parâmetros preditores*, em [7] podem ser encontrados mais detalhes. Neste contexto, ele prediz a última coluna de  $W_{M+1}$ . Este fato produz a seguinte relação:

$$W_M C_a = \Lambda W_M, \quad (4.3)$$

em que  $C_a$  é a matriz companheira definida por

$$C_a = [e_2 \ e_3 \ \dots \ e_M \ -\hat{f}_M], \quad (4.4)$$

sendo  $e_j$  o  $j$ -ésimo vetor canônico em  $\mathbb{R}^M$ , associada ao polinômio escalar

$$P_m(\lambda) = \lambda^M + a_{M-1}\lambda^{M-1} + \dots + a_1\lambda + a_0.$$

Muitos dos resultados aqui apresentados sobre  $\kappa(W_M)$  dependem dos autovalores e dos valores singulares de uma matriz  $d \times d$  fortemente relacionada com  $C_a$ , digamos  $F_M$ , definida por

$$F_M = W_M^* C_a W_M, \quad (4.5)$$

em quem  $W_M = W_M^* (W_M W_M^*)^{-1/2}$ . Observe que a definição acima é consistente já que  $W_M W_M^*$  é definida-positiva. O seguinte teorema descreve o espectro de  $F_M$ .

**Teorema 4.1.3** *A matriz  $F_M$  descrita em (4.5) possui uma decomposição espectral dada por:*

$$F_M = Q_M \Lambda Q_M^{-1}, \quad (4.6)$$

em que  $Q_M = (W_M W_M^*)^{-1/2}$ . Além disso, a matriz de autovetores satisfaz

$$\kappa_2(Q_M) = \kappa_2(W_M) = \|W_M\|_2 \|W_M^\dagger\|_2.$$

**Demonstração:** Utilizando (4.5)

$$\begin{aligned} F_M &= W_M^* C_a W_M = (W_M W_M^*)^{-1/2} W_M C_a W_M^* (W_M W_M^*)^{-1/2} \\ &= (W_M W_M^*)^{-1/2} W_M W_M^\dagger \Lambda W_M W_M^* (W_M W_M^*)^{-1/2} \\ &= (W_M W_M^*)^{-1/2} \Lambda (W_M W_M^*)^{1/2} \end{aligned}$$

Falta agora demonstrar que  $\kappa_2(Q_M) = \kappa_2(W_M)$ . De fato,

$$\begin{aligned} \kappa_2(Q_M) &= \|Q_M\|_2 \|Q_M^\dagger\|_2 \\ &= \|(W_M W_M^*)^{-1/2}\|_2 \|(W_M W_M^*)^{1/2}\|_2 \\ &= \|W_M^\dagger\|_2 \|W_M\|_2 = \kappa_2(W_M). \end{aligned}$$

□

O próximo teorema caracteriza o espectro singular de  $F_M$  (veja Bazán [11, Teor. 4]).

**Teorema 4.1.4** *Seja  $F_M$  a matriz definida em (4.5). Então, seu espectro singular é descrito por*

$$\begin{aligned}\sigma_1^2(F_M) &= \frac{2 + \|\hat{f}_M\|_2^2 - \|p_1\|_2^2 + \sqrt{(\|\hat{f}_M\|_2^2 + \|p_1\|_2^2)^2 - 4|f_1|}}{2}, \\ \sigma_j^2(F_M) &= 1, \quad j = 2, \dots, d-1, \\ \sigma_d^2(F_M) &= \frac{2 + \|\hat{f}_M\|_2^2 - \|p_1\|_2^2 - \sqrt{(\|\hat{f}_M\|_2^2 + \|p_1\|_2^2)^2 - 4|f_1|}}{2},\end{aligned}\tag{4.7}$$

onde  $p_1$  é a primeira coluna de  $\mathcal{P}_M$ , o operador de projeção ortogonal sobre  $\mathcal{R}(W_M^*)$ , e  $f_1$  é a primeira componente do vetor  $\hat{f}_M$ , introduzido no Teorema 4.1.2

O condicionamento dos autovalores em relação a perturbações na matriz pode ser estimado como uma função da proximidade entre os valores singulares e o valor absoluto dos autovalores, veja [104]. Assim, quanto mais próximos os valores singulares do valor absoluto dos autovalores, mais bem condicionado é o problema de autovalor.

Os resultados sobre limitantes para  $\kappa_2(W_M)$  são dados no teorema a seguir.

**Teorema 4.1.5 (Limitantes para  $\kappa(W_M)$ )** *Seja  $W_M$  uma matriz de Vandermonde definida como em (4.1), com os  $\lambda_j$  no disco unitário. Defina*

$$\begin{aligned}\alpha &= \max_j |\lambda_j|, \quad \beta = \min_j |\lambda_j|, \quad \delta = \min_{\substack{j \neq k \\ j, k}} |\lambda_j - \lambda_k|, \quad \text{e} \\ D_M^2 &\doteq D^2(F_M) = \|F_M\|_F^2 - (|\lambda_1|^2 + \dots + |\lambda_d|^2).\end{aligned}$$

Então, para  $M > d \geq 2$ , o número de condição  $\kappa_2(W_M)$  satisfaz

$$\frac{\sigma_1(F_M)}{\alpha} \leq \kappa_2(W_M) \leq \frac{1}{2} \left( \eta + \sqrt{\eta^2 - 4} \right),\tag{4.8}$$

onde  $\eta = \rho - d + 2$ ,

$$\rho = d \left[ 1 + \frac{D_M^2}{(d-1)\delta^2} \right]^{\frac{d-1}{2}} \phi_M(\alpha, \beta),\tag{4.9}$$

$$\phi_M(\alpha, \beta) = \sqrt{\frac{1 + \alpha^2 + \alpha^4 + \dots + \alpha^{2(N-1)}}{1 + \beta^2 + \beta^4 + \dots + \beta^{2(N-1)}}}.\tag{4.10}$$

**Demonstração:** Para mostrar a desigualdade à esquerda, basta tomar a norma-2 em ambos os lados de (4.6) e usar o Teorema 4.1.3.

Para provar a desigualdade a direita, observe que

$$\|e_j^* Q_M^{-1}\|_2 = \|e_j^* (W_M W_M^*)^{1/2}\|_2 = \|e_j^* W_M\|_2.$$

Agora, utilizando (A.7) e (A.6), segue que

$$\kappa_F(W_M) \leq \rho, \tag{4.11}$$

sendo  $\rho$  descrito em (4.9). Utilizando agora Teorema 1, de Smith [104], o qual afirma que, para uma matriz  $X \in \mathbb{C}^{d \times d}$  não singular

$$d - 2 + \kappa_2(X) + \kappa_2^{-1}(X) \leq \kappa_F(X),$$

considerando que

$$d - 2 + \kappa_2(X) + \kappa_2^{-1}(X) \leq \kappa_F(X) \Leftrightarrow [\kappa_2(X)]^2 - \nu[\kappa_2(X)] + 1 \leq 0,$$

em que  $\nu = \kappa_F(X) - d + 2$ ; resolvendo esta inequação do segundo grau, obtém-se:

$$\kappa_2(X) \leq \frac{1}{2} \left[ \nu + \sqrt{\nu^2 - 4} \right].$$

Adaptando este resultado ao problema em questão segue que

$$\begin{aligned} \kappa_2(W_M) &\leq \frac{1}{2} \left[ \kappa_F(W_M) - d + 2 + \sqrt{(\kappa_F(W_M) - d + 2)^2 - 4} \right] \\ &\leq \frac{1}{2} \left[ \rho - d + 2 + \sqrt{(\rho - d + 2)^2 - 4} \right] = \\ &= \frac{1}{2} \left[ \eta + \sqrt{\eta^2 - 4} \right]. \end{aligned}$$

□

O próximo passo é analisar o limitante superior em (4.8) como função de  $M$ . Observe que o limitante depende de três fatores: da separação entre os  $\lambda_j$  no disco unitário, do número  $D_M^1$ , e de  $\phi_M(\alpha, \beta)$ . Porém, a contribuição de  $\phi_M(\alpha, \beta)$  não é muito expressiva, uma vez que os  $\lambda_j$  estão no disco unitário, assim, a qualidade do limitante depende basicamente da razão  $D_M^2/(d-1)\delta^2$ . Portanto, quando  $D_M^2$  for da mesma magnitude de  $(d-1)\delta^2$ , e  $d$  não for muito grande, tem-se limitantes moderados. O próximo lema mostra que isto é possível.

---

<sup>1</sup>O número  $D_M$ , conhecido como *desvio da normalidade de A* (*departure from normality*), mede quanto uma matriz  $A$  difere de ser normal; se  $A$  é normal  $D(A) = 0$ .

**Lema 4.1.6** *Seja  $D_M^2$  como no teorema anterior. Então, para cada  $M \geq d$*

$$(d-1) + \frac{\prod_{j=1}^d |\lambda_j|^2}{1 + \|\hat{f}_M\|^2} - \sum_{j=1}^d |\lambda_j|^2 \leq D_M^2 \leq (d-1) + \|\hat{f}_M\|_2^2 + \prod_{j=1}^d |\lambda_j|^2 - \sum_{j=1}^d |\lambda_j|^2, \quad (4.12)$$

e portanto,

$$D_\infty^2 = \lim_{M \rightarrow \infty} D_M^2 = (d-1) + \prod_{j=1}^d |\lambda_j|^2 - \sum_{j=1}^d |\lambda_j|^2. \quad (4.13)$$

**Demonstração:** Usando o fato que o produto dos valores singulares de uma matriz quadrada é igual ao produto do valor absoluto dos correspondentes autovalores, e o Teorema 4.1.4, segue que

$$\sigma_1^2(F_M) \sigma_d^2(F_M) = \prod_{j=1}^d |\lambda_j|^2.$$

Como  $\|F_M\|_F^2 = \sigma_1^2(F_M) + \dots + \sigma_d^2(F_M) = \sigma_1^2(F_M) + \sigma_d^2(F_M) + (d-2)$ , da definição de  $D_M^2$ , obtém-se:

$$D_M^2 = (d-2) + \sigma_1^2(F_M) + \frac{\prod_{j=1}^d |\lambda_j|^2}{\sigma_1^2(F_M)} - \sum_{j=1}^d |\lambda_j|^2. \quad (4.14)$$

Mas, como

$$1 \leq \sigma_1^2(F_M) \leq 1 + \|\hat{f}_M\|_2^2, \quad (4.15)$$

pelo Teorema 4.1.4, aplicando (4.15) em (4.14), obtém-se a desigualdade desejada.

Para provar a igualdade envolvendo  $D_\infty^2$  em (4.13), basta tomar limite quando  $M \rightarrow \infty$  em ambos os lados de (4.12), e usar o Teorema 4.1.2.

□

O Lema 4.1.6 mostra que o comportamento de  $D_M^2$  como função de  $M$ , depende da velocidade com que  $\|\hat{f}_M\|^2$  converge para zero. Quando  $M$  cresce, para  $M$  grande, o tamanho de  $D_M^2$  irá depender do tamanho dos  $\lambda_j$ . Desta forma, pode-se concluir que, sempre que  $M$  é suficientemente grande e  $|\lambda_j| \approx 1$ , o número  $D_M^2$  é pequeno. Conseqüentemente, exceto para o caso no qual os autovalores são muito próximos uns dos outros, a condição  $D_M^2 < (d-1)\delta^2$  deve ser satisfeita, assegurando assim, limitantes pequenos para  $\kappa_2(W_M)$ .

$l$	$r_l$	$\lambda_l$	$ \lambda_l $	$\delta_l^2$
1	$5.8921 + 1.5788i$	$0.6342 - 0.7463i$	0.9794	0.1786
2	$9.5627 + 2.5623i$	$0.8858 - 0.4067i$	0.9747	0.0644
3	$5.7956 + 1.5529i$	$0.9663 - 0.1661i$	0.9805	0.0644
4	$2.7046 + 0.7247i$	$0.9642 + 0.2174i$	0.9884	0.0100
5	$16.4207 + 4.3999i$	$0.8811 + 0.2729i$	0.9224	0.0100

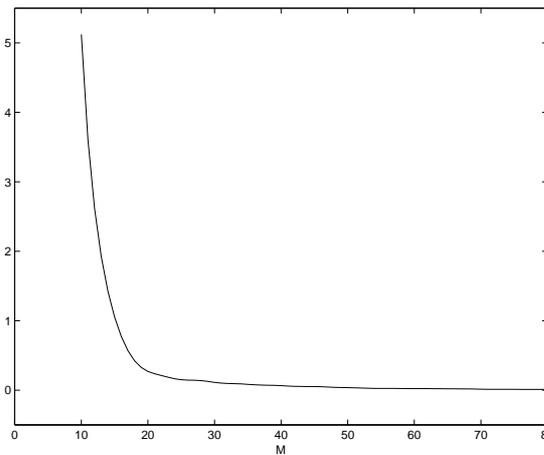
Tabela 4.1: Parâmetros do sinal usado nos testes

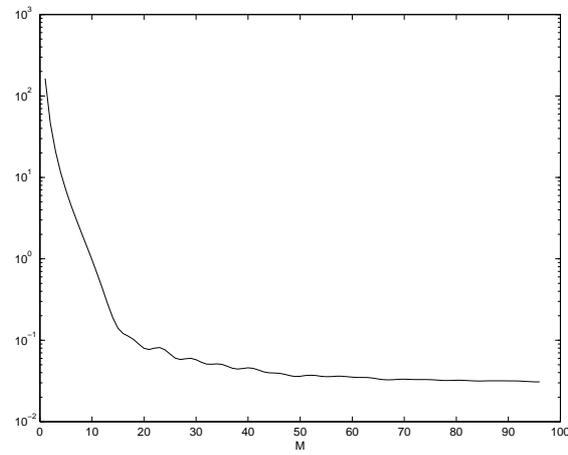
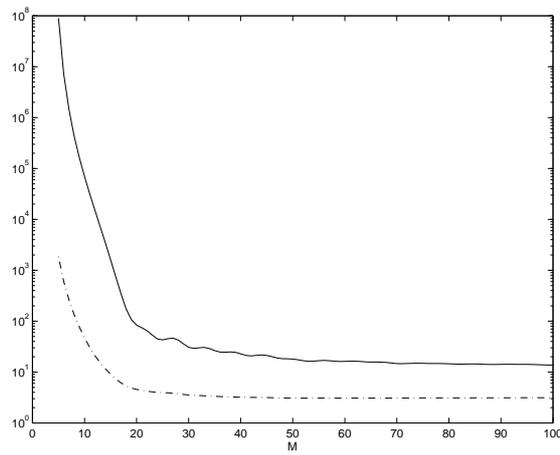
**Exemplo 4.1.7 (Análise de  $\kappa(W_M)$ ): Um problema proveniente de NMR)** O exemplo objetiva ilustrar o comportamento do número de condição  $\kappa(W_M)$  e seus limitantes teóricos como funções da dimensão  $M$ . Para tanto, utiliza-se uma matriz de vandermonde  $5 \times M$  em que os  $\lambda$ 's, descritos na Tabela 4.1, provêm de um sinal em NMR, veja [71].

O objetivo é mostrar que os limitantes para  $\|W_M^\dagger\|$  e o próprio  $\kappa_2(W_M)$  como funções de  $M$ , tornam-se pequenos a medida que  $M$  torna-se suficientemente grande.

O número  $D_M$  é calculado através de (4.14) e ilustrado na Figura 4.2. Na Figura 4.1 ilustra-se o rápido decréscimo de  $\|\hat{f}_M\|^2$  para 0, a medida que  $M$  cresce.

O comportamento do limitante (4.8), descrito na Figura 4.3, mostra que  $\kappa_2(W_M)$  torna-se cada vez menor a medida que  $M$  torna-se suficientemente grande. A desigualdade que garante valores moderados para este limitante,  $D_M^2 < (d-1)\delta^2$ , é rapidamente satisfeita já que, se por um lado, o valor de  $(d-1)\delta^2$  é 0.0399, por outro lado tem-se que  $D_{50}^2 = 0.0396$ ,  $D_{70}^2 = 0.0328$   $D_{100}^2 = 0.0308$ .

Figura 4.1: Comportamento de  $\|\hat{f}_M\|^2$ .

Figura 4.2: Comportamento de  $D_M^2$ .Figura 4.3: Limitante superior para  $\kappa_2(W_M)$ , obtida no teorema 4.8: linha sólida, e  $\kappa_2(W_M)$ : linha pontilhada, em escala logarítmica.

### 4.1.2 Análise de Sensibilidade: Abordagem da Matriz Companheira

Aqui analisamos a sensibilidade do autovalor  $\lambda_j$  através do número de condição  $\kappa(\lambda_j, C_a)$ , ou seja, a análise é feita considerando  $\lambda_j$  como autovalor de  $C_a$ , e não como autovalor de  $P_m(\lambda)$ .

**Teorema 4.1.8** *Seja  $\lambda_j$  um autovalor simples do polinômio*

$$P_m(\lambda) = \lambda^m + \lambda^{m-1}a_{m-1} + \cdots + a_0.$$

*Seja  $C_a$  a matriz companheira associada ao polinômio  $P_m(\lambda)$ . Assuma que  $C_a$  é decomposta como  $C_a = C_{\mathcal{P}} + C_{\mathcal{Q}}$ , em que  $C_{\mathcal{P}} = \mathcal{P}C_a$ ,  $C_{\mathcal{Q}} = \mathcal{Q}C_a$ , com  $\mathcal{P}$  sendo a matriz de projeção ortogonal sobre  $\mathcal{R}(W_M^*)$  e  $\mathcal{Q} = I - \mathcal{P}$ . Então o número de condição  $\kappa(\lambda_j, C_a)$  pode ser calculado como*

$$\begin{aligned} \kappa(\lambda, C_a) &= \sqrt{1 + |\lambda_j|^2 + |\lambda_j|^4 + \cdots + |\lambda_j|^{2(m-1)}} \times \\ &\quad \times \sqrt{\|W_m^\dagger e_j\|_2^2 + \|(C_{\mathcal{Q}} - \lambda_j I)^{-1} C_{\mathcal{Q}} W_m^\dagger e_j\|_2^2} \end{aligned} \quad (4.16)$$

**Demonstração:** Se  $\lambda_j$  é um autovalor de  $C_a$ , é imediato que um autovetor à esquerda associado a  $\lambda_j$ , é  $\ell_j = W_m^* e_j$ . Por outro lado, se  $r_j$  é um autovetor à direita associado ao mesmo autovalor, então  $r_j$  pode ser escrito como

$$r_j = \phi_j + \psi_j, \quad (4.17)$$

sendo  $\phi_j = W_m^\dagger e_j \in \mathcal{N}(W_m)^\perp$ , e  $\psi_j$  um vetor no espaço  $\mathcal{N}(W_m)$ .

É fácil provar que  $\phi_j$  é autovetor de  $C_{\mathcal{P}}$  associado ao autovalor  $\lambda_j$ . Utilizando esta informação, a relação (4.17), e a equação  $\lambda_j r_j = C_a r_j = (C_{\mathcal{P}} + C_{\mathcal{Q}}) r_j$ , vemos que

$$\psi_j = -(C_{\mathcal{Q}} - \lambda_j I)^{-1} C_{\mathcal{Q}} W_m^\dagger e_j.$$

A existência da inversa acima é garantida porque  $\lambda_j$  não é autovalor de  $C_{\mathcal{Q}}$  [7, 21]. A igualdade (4.16) resulta de substituir os autovetores  $\ell_j$  e  $r_j$  em (3.17), utilizando o fato de que eles satisfazem a condição  $\ell_j^* r_j = 1$ , e  $\phi_j \perp \psi_j$ .

□

Introduzindo  $\alpha = \max_j \{|\lambda_j|\}$ , prova-se que [16]

$$\sqrt{\|W_m^\dagger e_j\|_2^2 + \|(C_{\mathcal{Q}} - \lambda_j I)^{-1} C_{\mathcal{Q}} W_m^\dagger e_j\|_2^2} \leq \|W_M^\dagger\|_2 \sqrt{1 + \frac{\sqrt{d} \alpha^m}{1 - |\lambda_j|}}.$$

Esta desigualdade mostra que a sensibilidade do autovalor  $\lambda_j$ , quando considerado como autovetor da matriz  $C_a$ , depende fortemente do tamanho da norma  $\|W_m^\dagger\|$  e que  $\kappa(\lambda_j, C_a)$  pode-se tornar grande quando  $|\lambda_j| \approx 1$ .

### 4.1.3 Análise de Sensibilidade: Abordagem da Matriz Companheira Projetada

Para contornar a dificuldade do número de condição poder ser grande quando  $|\lambda_j| \approx 1$ , vamos diminuir a sensibilidade dos autovalores de interesse suprimindo de  $\kappa(\lambda_j, C_a)$  a influência da componente  $\psi_j \in \mathcal{N}(W_m)$  na norma do autovetor à direita  $r_j$  da matriz companheira  $C_a$ . Este é o assunto do próximo teorema.

**Teorema 4.1.9** *Seja  $\mathcal{V}$  uma matriz  $m \times d$  cujas colunas formam uma base ortonormal para  $\mathcal{R}(W_m^*)$ . Defina*

$$C_a = \mathcal{V}^* C_a \mathcal{V}. \quad (4.18)$$

*Então o número de condição  $\kappa(\lambda_j, C_a)$  do autovalor  $\lambda_j$ , considerado como autovalor de  $C_a$ , satisfaz*

$$\kappa(\lambda_j, C_a) = \sqrt{1 + |\lambda_j|^2 + |\lambda_j|^4 + \dots + |\lambda_j|^{2(m-1)}} \|W_m^\dagger e_j\|_2 \quad (4.19)$$

**Demonstração:** A prova deste teorema pode ser encontrada em [14].

□

**Observação 4.1.10** A matriz  $C_a$  definida em (4.18) é a projeção de  $C_a$  no subespaço gerado pelas colunas de  $\mathcal{V}$  em termos dessa base, veja [123, Lecture 33, p. 254]. Essa matriz é semelhante a  $F_M$ , descrita em (4.5), portanto, ambas as matrizes tem os mesmos autovalores. Isto justifica o nome da abordagem apresentada aqui.

A relevância do resultado em (4.19) é que além da dimensão do problema ter sido reduzido a  $d \times d$  (o que reduz significativamente o custo computacional), a sensibilidade do autovalor melhora significativamente: agora, ela depende apenas do tamanho do próprio autovalor e do comportamento da norma  $\|W_m^\dagger e_j\|_2$ . Ainda mais, já que

$$\kappa(\mathcal{V}^* W_m^\dagger) = \|\mathcal{V}^* W_m^\dagger\|_2 \|(\mathcal{V}^* W_m^\dagger)^{-1}\|_2 = \|W_m\|_2 \|W_m^\dagger\|_2 = \kappa(W_m),$$

veja [14] novamente, o teorema mostra que a sensibilidade global do problema de autovalores associado a matriz  $C_a$  depende do condicionamento da matriz de Vandermonde  $W_m$ . A análise de  $\kappa(W_m)$  como uma função do grau do polinômio e dos próprios autovalores,

desenvolvida anteriormente, mostra então que a sensibilidade global pode ser pequena desde que o grau seja suficientemente elevado.

Uma estimativa para  $\kappa(\lambda_j, \mathcal{C}_a)$  pode ser dada em termos do desvio da normalidade da matriz  $\mathcal{C}_a$ . Isso é sempre possível já que o espectro singular dessa matriz e  $D_m(\mathcal{C}_a)$  são agora conhecidos (veja Teorema 4 e Lemma 7 em [11]). Utilizando Teoremas 3 e 5 por Smith [104], segue o seguinte corolário.

**Corolário 4.1.11** *Seja  $\mathcal{C}_a$  a matriz definida em (4.18). Então o número de condição  $\kappa(\lambda_j, \mathcal{C}_a)$  satisfaz*

$$\kappa(\lambda_j, \mathcal{C}_a) \leq \left[ 1 + \frac{d-1 + \|\hat{f}_m\|_2^2 + \prod_{j=1}^d |\lambda_j|^2 - \sum_{j=1}^d |\lambda_j|^2}{(d-1)\delta_j^2} \right]^{n(n-1)/2}, \quad (4.20)$$

em que  $\delta_j = \min_{\substack{1 \leq k \leq d \\ k \neq j}} |\lambda_k - \lambda_j|$ .

Este resultado é muito mais expressivo que aquele dado no teorema acima já que agora observa-se claramente que para  $m$  suficientemente grande, exceto para o caso em que  $\delta_j \approx 0$ , deve-se ter  $\kappa(\lambda_j, \mathcal{C}_a)$  de tamanho moderado e próximo de 1 se  $|\lambda_j| \approx 1$ .

## 4.2 Sensibilidade de Autovalores: Caso Polinomial Matricial

Conforme descrito no Capítulo 2, assumiremos que os coeficientes  $A_j$  do polinômio matricial  $P_m(\lambda)$  satisfazem uma relação de recorrência do tipo

$$h_k A_0 + h_{k+1} A_1 + \cdots + h_{k+m-1} A_{m-1} = h_{k+m}, \quad k = 0, 1, \dots; \quad (4.21)$$

com  $h_k = R\Lambda^k L \in \mathbb{C}^{p \times q}$ ,  $R \in \mathbb{C}^{p \times d}$ ,  $L \in \mathbb{C}^{d \times q}$ , e  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_d)$ , tal que todos os autovalores de  $P_m(\lambda)$  são simples e fixos, e  $m \geq d$ . Então, conforme visto anteriormente, se  $l_j^*$  é a  $j$ -ésima linha de  $L$ ,  $l_j$  é autovetor à esquerda de  $P_m(\lambda)$  associado ao autovalor  $\lambda_j$ . Os coeficientes  $A_k$  satisfazem a equação matricial

$$\mathcal{K}_m X_A = \Lambda^m L, \quad (4.22)$$

em que  $\mathcal{K}_m$  é a matriz de posto completo definida por

$$\mathcal{K}_m = [L \quad \Lambda \quad \Lambda^2 L \cdots \Lambda^{m-1} L], \quad (4.23)$$

and  $X_A = [A_0^T \ A_1^T; \dots \ A_{m-1}^T]^T$ . A matriz  $\mathcal{K}_m$  é conhecida como matriz de Krylov, ou na área de sistemas dinâmicos como *matriz de controlabilidade*. Em termos da matriz companheira em blocos, a relação acima pode ser escrita como

$$\mathcal{K}_m C_A = \Lambda \mathcal{K}_m. \quad (4.24)$$

Esta equação simplesmente assegura que os vetores

$$\ell_j^* = e_j^* \mathcal{K}_m = [e_j^* L, \lambda_j e_j^* L, \dots, \lambda_j^{m-1} e_j^* L]$$

são autovetores à esquerda de  $C_A$  associados ao autovalor  $\lambda_j$ .

Uma análise de sensibilidade como aquela do caso escalar através da abordagem da matriz companheira em blocos pode também ser dada aqui. Para isso, os autovetores à direita de  $C_A$  podem ser descritos analogamente como no Teorema 4.1.8, e então uma estimativa como aquela dada em (4.16) pode ser deduzida, com  $\mathcal{K}_m^\dagger e_j$  em vez de  $W_m^\dagger e_j$ . Analogamente como no caso escalar, as estimativas assim obtidas podem ser muito pessimistas devido à contribuição da componente em  $\mathcal{N}(\mathcal{K}_m)$  do autovetor à direita. Para contornar essa dificuldade, analisaremos apenas a abordagem da matriz companheira em blocos projetada.

Daqui em diante, a notação será levemente modificada com o propósito de analisar o efeito da dimensão dos coeficientes  $A_k$  sobre a sensibilidade do problema de autovalor. Para tal, a matriz companheira em blocos, associada ao polinômio matricial de grau  $m$  com coeficientes  $A_k \in \mathbb{C}^{q \times q}$ , será denotada por  $C_A(m, q)$ . Em todos os casos, assume-se que  $m \geq d$  e  $q \geq 1$ . O caso  $q = 1$  significa simplesmente o caso escalar.

Um resultado preliminar para nossa análise, cuja prova é análoga àquela do Teorema 4.1.3, é dado no proximo lemma.

**Lema 4.2.1** *Seja  $\mathcal{V} = \mathcal{K}_m^\dagger \mathcal{Q}_{q,m}^{-1/2}$  em que  $\mathcal{Q}_{m,q} = \mathcal{K}_m \mathcal{K}_m^*$ . Defina*

$$\mathcal{C}_A(m, q) = \mathcal{V}^* C_A(m, q) \mathcal{V}.$$

Então

$$\mathcal{C}_A(m, q) = (\mathcal{Q}_{m,q})^{-1/2} \Lambda (\mathcal{Q}_{m,q})^{1/2}. \quad (4.25)$$

O lemma garante que a sensibilidade global do problema de autovalor associado à matriz companheira em blocos projetada é essencialmente governada por  $\sqrt{\kappa(\mathcal{Q}_{m,q})}$ .

**Teorema 4.2.2** *Se  $q \geq 1$  and  $m \geq d$  são números naturais fixos, então*

(a) O número de condição do autovalor  $\lambda_j$ , considerado como autovalor de  $\mathcal{C}_A(m, q)$ , pode ser calculado como

$$\kappa(\lambda_j, \mathcal{C}_A(m, q)) = \sqrt{1 + |\lambda_j|^2 + |\lambda_j|^4 + \cdots + |\lambda_j|^{2(m-1)}} \|\mathcal{K}_m^\dagger e_j\|_2, \quad (4.26)$$

em que  $e_j$  denota o  $j$ -ésimo vetor canônico em  $\mathbb{R}^d$ .

(b) O número de condição  $\kappa(\lambda_j, \mathcal{C}_A(m, 1))$  depende apenas da matriz de Vandermonde  $W_m$  descrita em (4.1) e não da matriz  $L$ .

(c) Assuma que  $m \geq d$  e que  $m$  seja fixo. Então, para cada  $q \geq 1$  vale

$$\kappa(\lambda_j, \mathcal{C}_A(m, q)) \leq \kappa(\lambda_j, \mathcal{C}_A(m, 1)).$$

(d) Seja  $\delta_j = \min_{\substack{k \\ k \neq j}} |\lambda_j - \lambda_k|$ ,  $1 \leq k \leq d$ . Então, para  $m \geq d$  e  $q \geq 1$  tem-se

$$\kappa(\lambda_j, \mathcal{C}_A(m, q)) \leq \left[ 1 + \frac{d-1 + \|\hat{f}^\dagger\|_2^2 + \prod_{j=1}^d |\lambda_j|^2 - \sum_{j=1}^d |\lambda_j|^2}{(d-1)\delta_j^2} \right]^{(d-1)/2}, \quad (4.27)$$

em que  $\hat{f}^\dagger$  é a solução de norma mínima do sistema linear  $W_m x = \Lambda^m e$ , onde  $e = [1, \dots, 1]^T \in \mathbb{R}^d$ .

**Demonstração:** Para provar (a) observe do Lemma 4.2.1 que

$$v_j = \mathcal{Q}_{m,q}^{-1/2} e_j, \quad e \quad u_j = \mathcal{Q}_{m,q}^{1/2} e_j$$

são autovetores à esquerda e à direita de  $\mathcal{C}_A(m, q)$ , respectivamente, associados ao autovalor  $\lambda_j$ , e que eles satisfazem a propriedade  $u_j^* v_j = 1$ . Além disso,

$$\|v_j\|_2^2 = v_j^* v_j = e_j^* \mathcal{Q}_{m,q}^{-1} e_j = \|\mathcal{K}_m^\dagger e_j\|_2^2,$$

e

$$\|u_j\|_2^2 = u_j^* u_j = e_j^* \mathcal{Q}_{m,q} e_j = \|e_j^* \mathcal{K}_m\|_2^2 = 1 + |\lambda_j|^2 + |\lambda_j|^4 + \cdots + |\lambda_j|^{2(m-1)}.$$

A última igualdade é porque foi assumido que as linhas de  $L$  têm norma-2 igual a 1. A igualdade (4.26) resulta da substituição das duas igualdades acima na definição de  $\kappa(\lambda_j, \mathcal{C}_A(m, q))$  conforme dada por Wilkinson, veja (3.17).

Para provar o item (b), observe que quando  $q = 1$  a submatriz  $L$  de  $\mathcal{K}_m$  reduz-se a um vetor coluna. Neste caso, pode-se escrever  $\mathcal{Q}_{m,1} = L^{(1)} W_m W_m^* L^{(1)*}$ , em que

$L^{(1)} = \text{diag}(L_{1,1}, \dots, L_{n,1})$ , e com  $W_m$  sendo a matriz de Vandermonde introduzida na seção prévia. Utilizando esta observação e a análise acima, segue que

$$\kappa_{j,1} = \|e_j^* W_m\|_2 \|W_m^\dagger e_j\|_2,$$

o qual prova o item (b).

A prova do item (c) está baseada no fato de que

$$\|\mathcal{K}_m^\dagger e_j\|_2 \leq \|W_m^\dagger e_j\|_2. \quad (4.28)$$

Isto pode ser visto da seguinte maneira. Seja  $\psi_j = \mathcal{K}_m^\dagger e_j$ . Então  $\psi_j$  é a solução de norma mínima do sistema subdeterminado

$$\mathcal{K}_m \psi = e_j. \quad (4.29)$$

Seja  $\widehat{\mathcal{K}}_m$  definida por

$$\widehat{\mathcal{K}} = [L^{(1)} W_m \ L^{(2)} W_m \ \dots \ L^{(q)} W_m],$$

em que

$$L^{(i)} = \text{diag}(L_{1,i}, \dots, L_{n,i}), \quad i = 1, \dots, q.$$

Obviamente,  $\widehat{\mathcal{K}}_m = \mathcal{K}_m \mathcal{J}$  com  $\mathcal{J}$  sendo um matriz de permutação apropriada. Defina ainda

$$\widehat{\mathcal{K}}_m^D = \begin{bmatrix} W_m^\dagger L^{(1)*} \\ W_m^\dagger L^{(2)*} \\ \vdots \\ W_m^\dagger L^{(q)*} \end{bmatrix}.$$

Então segue que

$$\widehat{\mathcal{K}}_m \widehat{\mathcal{K}}_m^D = L^{(1)} L^{(1)*} + L^{(2)} L^{(2)*} + \dots + L^{(q)} L^{(q)*} = I_n,$$

porque as linhas de  $L$  têm norma-2 igual a 1. Isso mostra que  $\widehat{\mathcal{K}}_m^D$  é uma inversa à direita de  $\widehat{\mathcal{K}}_m$ . Defina também  $\psi = \mathcal{J} \widehat{\mathcal{K}}_m^D e_j$ . Então  $\psi$  é uma de solução de (4.29) e

$$\begin{aligned} \|\psi\|_2^2 &= \|W_m^\dagger e_j\|_2^2 |L_{j,1}|^2 + \|W_m^\dagger e_j\|_2^2 |L_{j,2}|^2 + \dots + \|W_m^\dagger e_j\|_2^2 |L_{j,q}|^2 \\ &= \|W_m e_j\|_2^2 (|L_{j,1}|^2 + |L_{j,2}|^2 + \dots + |L_{j,q}|^2) \\ &= \|W_m^\dagger e_j\|_2^2. \end{aligned}$$

A última igualdade deve-se ao fato da  $j$ -éssima linha de  $L$  ter norm-2 igual a 1. Este resultado prova a desigualdade (4.28), porque  $\psi_j$  é a solução de norma mínima do sistema (4.29), e prova automaticamente o item (c).

Finalmente, o item (d) é uma consequência imediata do item (c) e Corolario 9 in Bazán [11].

□

As propriedades (a) e (b) garantem que a sensibilidade do autovalor depende essencialmente do condicionamento da matriz de Krylov  $\mathcal{K}_m$ . Já a propriedade (c) garante redução de sensibilidade do autovalor quando extraído de um polinômio matricial com coeficientes  $A_j$  de ordem  $q > 1$  (em teoria de sistemas, isto acontece quando as matrizes  $h_k$ , representam respostas ao impulso unitario correspondentes a um sistema com entradas múltiplas). Veremos posteriormente que a sensibilidade do autovalor usualmente diminui significativamente quando  $q = 2$  comparado com o caso  $q = 1$ , ou seja,  $q = 2$ , usualmente assegura que  $\kappa(\lambda, \mathcal{C}_A(m, 2)) \ll \kappa(\lambda, \mathcal{C}_A(m, 1))$ .

Finalmente, a desigualdade (4.27) sugere que autovalores tal que  $|\lambda_j| \approx 1$ , não extremamente próximos uns aos outros, (ou seja,  $\delta_j$  não muito pequeno), tornam-se quase perfeitamente bem condicionados toda vez que  $\|f^\dagger\|_2^2 \approx 0$ , porque neste caso  $\kappa(\lambda, \mathcal{C}_A(m, 1)) \approx 1$ . É interessante notar que as condições  $|\lambda_j| \approx 1$ ,  $\|f^\dagger\|_2^2 \approx 0$  são muito freqüentes em conexão com sistemas mecânicos levemente amortecidos. Alguns exemplos podem ser vistos em [10].

### 4.2.1 Análise de Sensibilidade: Um Caso Real

**Exemplo 4.2.3 (Mini-Mast Model)** Neste exemplo considera-se um polinômio matricial associado a um sistema dinâmico conhecida como Mini-Mast. O Mini-Mast é uma viga estrutural utilizada para pesquisa em dinâmica estrutural e controle ativo de vibrações na NASA Langley research Center [86]. O sistema é descrito por equações de estado do tipo

$$\begin{cases} \dot{x} = Ax + Bu \\ y = Cx, \end{cases} \quad (4.30)$$

em que  $A$ ,  $B$  e  $C$  são de ordem  $10 \times 10$ ,  $10 \times 2$ , e  $2 \times 10$ , respectivamente (ou seja, o modelo matemático considera duas entradas e duas saídas). As entradas das matrizes  $A$ ,  $B$  e  $C$  podem ser encontradas em [86]. As matrizes  $h_k$  da relação de recorrência (4.21) são portanto  $2 \times 2$  e descritas como

$$h_k = Ce^{A\Delta t k} B, \quad k = 0, 1, \dots$$

O sistema caracteriza-se por possuir 2 pares de autovalores muito próximos uns aos outros, sendo portanto um excelente teste para algoritmos de identificação na área de engenharia aero espacial. Os autovalores são dados por  $\lambda_j = e^{s_j \Delta t}$ ,  $j = 1 : 10$ , com  $\Delta t = 0.03s$ . As freqüências e amortecimentos (descritos pela parte imaginária e parte real dos autovalores  $\lambda_j$ ), os autovalores em módulo, e as separações  $\delta_j$  são descritos na Tabela 4.2.

Tabela 4.2: Autovalores e separações.

Modo $j$	Fator de Amort.	Frequência rad/s	$ \lambda_j $	$\delta_j$
1	0.32907	27.42011	0.99017	0.32299
2	0.38683	38.68230	0.98846	0.00982
3	0.38352	38.35103	0.98856	0.00982
4	0.09066	5.03555	0.99728	0.00011
5	0.09055	5.03176	0.99728	0.00011

Este exemplo visa ilustrar os resultados teóricos descritos no Teorema 4.2.2. Para tanto, calcula-se o número de condição do autovalor  $\lambda_j$  considerado como autovalor da matriz companheira projetada associada a polinômios de grau  $m = 10$  e  $m = 20$ . Considera-se  $q = 1$  (o caso escalar) e  $q = 2$  (o caso polinomial matricial). Os resultados da Tabela 4.3 mostram que o número de condição pode de fato, decrescer drasticamente com o aumento de  $m$  quando  $q > 1$ , tornando o autovalor muito bem condicionado, contrariamente ao que ocorre quando  $q = 1$ . Ou seja, redução de sensibilidade é efetivamente obtida quando o autovalor é associado a um polinômio matricial.

Tabela 4.3: Números de condição dos autovalores do Sistema.

Modo $j$	$\kappa(\lambda_j, \mathcal{C}_A(m, 1))$ $m = 10$	$\kappa(\lambda_j, \mathcal{C}_A(m, 1))$ $m = 20$	$\kappa(\lambda_j, \mathcal{C}_A(m, 2))$ $m = 10$	$\kappa(\lambda_j, \mathcal{C}_A(m, 2))$ $m = 20$
1	$0.00017 \times 10^7$	$0.00130 \times 10^3$	1.84786	1.00766
2	$0.00127 \times 10^7$	$0.02310 \times 10^3$	1.20076	1.00611
3	$0.00136 \times 10^7$	$0.02311 \times 10^3$	1.71432	1.00758
4	$3.10889 \times 10^7$	$4.75131 \times 10^3$	1.52448	1.00447
5	$3.11084 \times 10^7$	$4.75306 \times 10^3$	2.15234	1.00587

### 4.3 Análise de erro: Caso Escalar

Apresentamos estimativas para o erro  $|\lambda_l - \tilde{\lambda}_l|$ ,  $l = 1, \dots, d$ , considerando que os  $\lambda$ 's são extraídos a partir de uma matriz de transição  $T$  calculada pelo método de subespaço que utiliza a propriedade de deslocamento da matriz de Hankel associada ao sinal, ou seja, a matriz  $T$  satisfaz a equação (2.17) (veja Cap. 2), e que os  $\tilde{\lambda}$ 's são autovalores de  $\tilde{T} \approx T$ . Para tal, observe da prova do Teorema 2.1.4 (veja também (2.18)), que a matriz

de transição  $T$  tem uma decomposição espectral do tipo

$$T = (\mathcal{V}^* V_M) \Lambda (\mathcal{V}^* V_M)^{-1}. \quad (4.31)$$

Daí, segue que a matriz de autovetores satisfaz  $\kappa(\mathcal{V}^* V_M) = \kappa(V_M)$ , e assim, a sensibilidade dos autovalores  $\lambda$  depende de  $\kappa(V_M)$ , o número de condição da matriz de Vandermonde. Portanto, se  $\tilde{T}$  denota uma aproximação de  $T$ , a teoria clássica de perturbação de autovalores, veja a relação (A.2.6) e (4.31), garante que

$$|\lambda_l - \tilde{\lambda}_l| \leq \|T - \tilde{T}\| \kappa(V_M), \quad 1 \leq l \leq d. \quad (4.32)$$

Deseja-se mostrar que os autovalores  $\lambda_j$  podem-se tornar insensíveis a pequenas perturbações nos dados, sob certas condições.

Como o comportamento de  $\kappa_2(V_M)$  já foi estudado anteriormente, falta estudar agora o comportamento de  $\|T - \tilde{T}\|$ . O erro  $\|T - \tilde{T}\|$ , será analisado somente nos casos em que  $\tilde{T}$  é estimada através de técnicas de quadrados mínimos, como descrito em (2.19).

Sejam  $V_S$  e  $\tilde{V}_S$  matrizes  $M \times d$ , com colunas ortonormais, que geram o subespaço sinal exato  $\mathcal{S}_M$  e o subespaço sinal aproximado  $\tilde{\mathcal{S}}_M$  respectivamente. Decomponha  $\tilde{V}_S$  e  $V_S$  como

$$V_S = \begin{bmatrix} A \\ x^* \end{bmatrix} = \begin{bmatrix} y^* \\ B \end{bmatrix}, \quad \tilde{V}_S \begin{bmatrix} \tilde{A} \\ \tilde{x}^* \end{bmatrix} = \begin{bmatrix} \tilde{y}^* \\ \tilde{B} \end{bmatrix},$$

onde  $x^*$  e  $y^*$  representam a primeira e a última linha de  $V_S$  respectivamente, e  $\tilde{x}^*$  e  $\tilde{y}^*$  a primeira e a última linha de  $\tilde{V}_S$  respectivamente.

Observe que, para sinais livres de ruídos, a matriz  $T$  é a única solução do sistema (2.17). Sejam  $\tilde{A} = A + \Delta_A$ ,  $\tilde{B} = B + \Delta_B$  e assuma que  $\tilde{A}$  possui posto igual a  $d$ . Então o erro em  $T$  é:

$$\begin{aligned} \tilde{T} - T &= \tilde{A}^\dagger \tilde{B} - T \\ &= \tilde{A}^\dagger (B + \Delta_B) - T \\ &= \tilde{A}^\dagger (AT + \Delta_B) - T \\ &= \tilde{A}^\dagger (\tilde{A}T - \Delta_A T + \Delta_B) - T \\ &= \tilde{A}^\dagger (-\Delta_A T + \Delta_B). \end{aligned} \quad (4.33)$$

A última desigualdade segue do fato de que  $\tilde{A}^\dagger \tilde{A} = I$ , pois foi assumido que  $\text{rank}(A) = d$ . Tomando a norma-2 em ambos os lados de (4.33), tem-se:

$$\|\tilde{T} - T\| \leq \|\tilde{A}^\dagger\| \|T\| (\|\Delta_A\| + \|\Delta_B\|), \quad (4.34)$$

uma vez que  $\|T\| \geq 1$  (ver [10]). Observe que, para se obter qualquer estimativa de erro, favorável à análise,  $V_S$  deve ser escolhida o mais próximo possível de  $\tilde{V}_S$ . Isto é sempre possível, e pode ser feito escolhendo  $V_S = VX$ , onde  $V$  é qualquer matriz  $M \times d$ , com colunas ortonormais que geram o subespaço  $\mathcal{S}_M$ , e  $X \in \mathbb{C}^{d \times d}$  é uma matriz unitária que resolve o problema de Procrustes Ortogonal:

$$\min \|VX - \tilde{V}_S\|_F.$$

Defina  $G = \tilde{V}_S^*V$ , e considere a sua decomposição em valores singulares  $G = \bar{P}\bar{\Sigma}\bar{Q}^*$ . Então, a matriz unitária que resolve o problema de Procrustes é  $X = \bar{Q}\bar{P}^*$ , maiores detalhes sobre a solução deste problema podem ser encontrados em [59, p. 601]. Usando esta escolha de  $X$ , segue que  $V_S = V\bar{Q}\bar{P}^*$ , e assim

$$\begin{aligned} \|V_S - \tilde{V}_S\|^2 &= \|(V\bar{Q}\bar{P}^* - \tilde{V}_S)^*(V\bar{Q}\bar{P}^* - \tilde{V}_S)\| \\ &= \|2I - \bar{P}\bar{Q}^*V^*\tilde{V}_S - \tilde{V}_S^*V\bar{Q}\bar{P}^*\| \\ &= \|2I - \bar{P}\bar{Q}^*G^* - G\bar{Q}\bar{P}^*\| \\ &= 2\|I - \bar{P}\bar{\Sigma}\bar{P}^*\|. \end{aligned}$$

Da Seção A.6, no apêndice A, tem-se que os cossenos dos ângulos entre os subespaços  $\mathcal{S}_M$  e  $\tilde{\mathcal{S}}_M$  são os valores singulares de  $\tilde{V}_S^*V$ , os quais estão contidos em  $\bar{\Sigma}$ . Usando esta observação, a desigualdade acima implica em

$$\|\Delta V_S\| = \|V_S - \tilde{V}_S\| = 2 \operatorname{sen} \frac{\theta}{2}, \quad (4.35)$$

onde  $\theta$ , algumas vezes chamado *ângulo de subespaço*, denota o maior ângulo canônico entre  $\mathcal{S}_M$  e  $\tilde{\mathcal{S}}_M$ . Como  $\|\Delta_A\| \leq \|\Delta V_S\|$  e  $\|\Delta_B\| \leq \|\Delta V_S\|$ , usando estas desigualdades em (4.35), obtem-se o seguinte resultado:

**Lema 4.3.1**

$$\|T - \tilde{T}\| \leq 4 \|\tilde{A}^\dagger\| \|T\| \operatorname{sen} \frac{\theta}{2} \quad (4.36)$$

Observe que nesta estimativa  $\|\tilde{A}^\dagger\|$  e  $\|T\|$  dependem fortemente de  $M$ , o número de linhas da matriz  $\tilde{H}(l)$ . Isto sugere que devemos discutir o comportamento dessas quantidades como funções de  $M$ . De fato, seja  $C$  uma matriz companheira,  $M \times M$ , tal que

$$CV_M = V_M Z,$$

onde  $V_M$ , é uma matriz de Vandermonde como em (2.8). Então, como  $V_S V_S^* V_M = V_M$ , pois  $V_S V_S^*$  é o operador projeção ortogonal em  $\mathcal{S}_M$  e as colunas de  $V_M$  pertencem a este espaço, a equação acima pode ser reescrita como

$$CV_S V_S^* V_M = V_S V_S^* V_M Z.$$

Usando esta igualdade, o fato de que  $V_S^* V_S = I$  e (4.31), segue então que

$$V_S^* C V_S = (V_S^* V_M) Z (V_S^* V_M)^{-1} = T.$$

Usando agora o Teorema 2 de Bazán e Toint[10], tem-se que

$$\|T\|_2 = \|V_S^* C V_S\|_2 \leq \sqrt{1 + \|\hat{f}_M\|_2^2}.$$

Daí, por causa do decréscimo de  $\|\hat{f}_M\|$  para zero a medida que  $M$  cresce, segue que para  $M$  suficientemente grande, tem-se  $\|T\| \approx 1$ .

Por outro lado, observe também que  $\|\tilde{A}^\dagger\| = 1/(1 - \|\tilde{x}\|^2) = 1/(1 - \|\tilde{p}_M\|^2)$ , onde  $\tilde{p}_M$  é a última coluna do operador de projeção ortogonal no subespaço  $\tilde{\mathcal{S}}_M$ . Isto sugere que se  $\|\tilde{p}_M\|$  se comporta aproximadamente como  $\|p_M\|$ , a qual decresce com  $M$  e permanece próxima de 0, para  $M$  suficientemente grande (veja [10] novamente), então  $\|\tilde{A}^\dagger\|$  não deve ser muito maior do que 1. Desta forma, conclui-se que a estimativa dada no Lema 4.3.1, depende fortemente de  $M$ , apesar da proximidade de  $\mathcal{S}_M$  e  $\tilde{\mathcal{S}}_M$ , medida pelo ângulo entre os subespaços também ser importante. Isto significa que para valores de  $M$  grandes, o limitante (4.36) fica próximo de  $2 \sin \theta$ , a menos que  $\mathcal{S}_M$  esteja muito distante de  $\tilde{\mathcal{S}}_M$ . Apesar deste resultado não provar que os erros nas entradas da matriz de dados não são propagados no cálculo de  $\tilde{T}$ , para  $M$  suficientemente grande, ele assegura pequenos erros em  $T$  e ângulo entre subespaço suficientemente pequeno.

Substituindo (4.36) em (4.33) obtém-se o próximo teorema:

**Teorema 4.3.2** *O erro nos autovalores satisfaz*

$$|\lambda_i - \tilde{z}_i| \leq 2 \|\tilde{A}^\dagger\| \|T\| (\eta + \sqrt{\eta^2 - 4}) \sin \frac{\theta}{2}, \quad (4.37)$$

onde  $\eta$  é dado no Teorema 4.1.5.

Como já foi discutido na seção anterior, para valores de  $M$  suficientemente grandes,  $\eta$  torna-se uma constante moderada. Assim, o comportamento de  $\|\tilde{A}^\dagger\|$  e  $\|T\|$  como função de  $M$ , assegura que, se  $\tilde{V}_S$  é suficientemente próxima de  $V_S$ , no sentido da norma de

Frobenius, então o limitante para o erro nos autovalores do sinal pode ser considerado como o produto de  $\sin \theta$  por uma constante moderada, isto é

$$|\lambda_l - \tilde{z}_l| = \mathcal{O}(\sin \theta), \quad 1 \leq l \leq d. \quad (4.38)$$

Será assumido agora que  $\tilde{V}_S$  é estimada pela SVD da matriz de dados  $\tilde{H}(l) = H(l) + E$ . O próximo corolário fornece um limitante que depende do tamanho de  $\|E\|$  e do menor valor singular não nulo da matriz de dados exata  $\sigma_d(H(l))$ .

**Corolário 4.3.3** *Assuma que uma base para  $\tilde{\mathcal{S}}_M$  é calculada através da SVD de  $\tilde{H}(l)$ . Assumindo também  $\|E\| \ll \sigma_d(H(l))$ , a seguinte estimativa de primeira ordem para o limitante de  $\|T - \tilde{T}\|$  em (4.36) ocorre*

$$\|T - \tilde{T}\| \leq 2 \|\tilde{A}^\dagger\| \|T\| \frac{\|E\|}{\sigma_d(H(l))}. \quad (4.39)$$

Conseqüentemente, o limitante do erro nos autovalores do sinal (4.37) torna-se

$$|\lambda_l - \tilde{z}_l| \leq \|\tilde{A}^\dagger\| \|T\| (\eta + \sqrt{\eta^2 - 4}) \frac{\|E\|}{\sigma_d(H(l))} \quad 1 \leq l \leq d. \quad (4.40)$$

**Demonstração:** Sejam a SVD de  $H(l)$  e de  $\tilde{H}(l)$  respectivamente:

$$H(l) = [U_1 \ U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^* \\ V_2^* \end{bmatrix} \quad \text{e} \quad \tilde{H}(l) = [\tilde{U}_1 \ \tilde{U}_2] \begin{bmatrix} \tilde{\Sigma}_1 & 0 \\ 0 & \tilde{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \tilde{V}_1^* \\ \tilde{V}_2^* \end{bmatrix},$$

onde  $U_1, \tilde{U}_1, V_1$  e  $\tilde{V}_1$  possuem  $d$  colunas. Se  $\|E\| \ll \sigma_d(H(l))$ , então uma base ortonormal para o subspaço sinal aproximado (gerado pelas colunas de  $\tilde{U}_1$ ), é formada pelas colunas de uma matriz  $M \times d$ , digamos  $\hat{U}_1$ , tal que

$$\hat{U}_1 = (U_1 + U_2 P)(I + P^* P)^{-1/2}, \quad (4.41)$$

onde  $P$  é uma matriz de ordem  $(M - d) \times d$ , cuja norma é da ordem de  $\|E\|$  ( em [119] podem ser encontradas maiores explicações), e  $(I + P^* P)^{-1/2}$  é a inversa da raiz quadrada da matriz positiva definida  $(I + P^* P)$ . Embora não seja possível encontrar a matriz  $P$ , a sua norma  $\|P\|$  é interpretada como a tangente do maior ângulo canônico  $\Theta$ , veja Stewart [107]. Por outro lado, como

$$U_1^* \hat{U}_1 = \hat{U}_1^* U_1 = (I + P^* P)^{-1/2}, \quad (4.42)$$

e, como os valores singulares desta matriz Hermitiana, são os cossenos dos ângulos canônicos entre  $\mathcal{S}_M$  e  $\tilde{\mathcal{S}}_M$ , após algumas manipulações algébricas segue que,  $\|U_1 - \hat{U}_1\|_F = 2 \sin \frac{\theta}{2}$ . Isto assegura que  $U_1$  é a matriz mais próxima de  $\hat{U}_1$ , no sentido que  $\|U_1 - \hat{U}_1\|_F$  é minimizada. Finalmente, como

$$2 \sin \frac{\theta}{2} \leq \tan \Theta = \|P\| \leq \frac{\|E\|}{\sigma_d(H(l))},$$

onde a última desigualdade vale até a primeira ordem de aproximação, veja Vaccaro [119]. As desigualdades (4.39) e (4.40) seguem da substituição desta desigualdade em (4.36) e (4.37), respectivamente.

□

Observe que a desigualdade (4.39) mostra que irão ocorrer reduções nos erros de entrada, sempre que  $2 \|\tilde{A}^\dagger\| \|T\| < \sigma_d(H(l))$ , um fato que é freqüentemente observado se  $M$  for suficientemente grande. Por outro lado, esta desigualdade sugere escolher as dimensões da matriz de dados, de forma que o limitante seja minimizado para garantir um ângulo pequeno entre subespaços. Não existe nenhum resultado teórico sobre o comportamento de  $\sigma_d(H(l))$  como uma função de  $M$  e  $N$ , mas evidências empíricas e alguns resultados teóricos recentes [10, 16], sugerem escolher a matriz de dados o mais próximo possível de uma matriz quadrada. Desta forma  $\sigma_d(H(l))$  pode atingir um valor máximo, neste caso o limitante pode ser minimizado.

### 4.3.1 Resultados Numéricos

O objetivo desta subseção é ilustrar numericamente as estimativas teóricas de erro obtidas neste seção. Para tanto, utilizamos um sinal típico da área de Ressonância Magnética Nuclear extraído de [71]. O sinal é modelado por (2.1) e possui 5 exponenciais. Os parâmetros que descrevem o sinal, bem como a separação entre os  $\lambda$ 's, são dados na Tabela 4.1. Os parâmetros  $\lambda_l$ ,  $l = 1, \dots, 5$ , satisfazem  $|\lambda_l| \approx 1$ , e a separação entre eles é satisfatória. O intervalo de amostragem é  $\Delta t = 0.0001s$ .

Para efeito de ilustração do comportamento dos limitantes para o erro nos autovalores  $|\lambda_l - \tilde{\lambda}_l|$ , foram calculados o limitante (4.32), o erro  $|\lambda_l - \tilde{\lambda}_l|$ ,  $\sin \theta$ ,  $2 \sin \frac{\theta}{2}$ ,  $\|E\|/\sigma_d(H)$  e  $\|T - \tilde{T}\|$ . Os autovalores  $\tilde{\lambda}_l$  foram calculados usando (2.17), as matrizes  $\tilde{A}$  e  $\tilde{B}$  foram calculadas a partir da SVD da matriz  $\tilde{H}(l) = H(l) + E$ , cujas entradas são amostras do sinal perturbado  $\tilde{h} = h + \epsilon$ , onde o ruído  $\epsilon$  é da ordem de 3%, isto é,  $\|\epsilon\|/\|h\| \approx 0.03$ . Os resultados apresentados são a média dos valores obtidos para 100 experimentos com

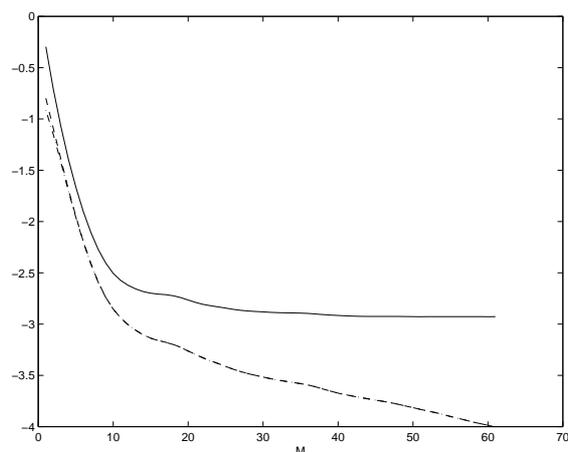


Figura 4.4:  $\|E\|/\sigma_d(H(l))$ : linha contínua,  $\sin \theta$ : linha tracejada,  $2 \sin \frac{\theta}{2}$ : linha tracejada e pontilhada, em escala logarítmica.

ruídos diferentes gerados no Matlab, porém da mesma ordem. A Figura 4.4 mostra os comportamentos de  $\sin \theta$ ,  $2 \sin \frac{\theta}{2}$ ,  $\|E\|/\sigma_d(H)$ . Percebe-se que estes valores decrescem a medida que a matriz  $\tilde{H}(l)$  torna-se uma matriz quadrada. O limitante para o erro nos autovalores (4.32), o erro  $\|T - \tilde{T}\|$  e o erro propriamente dito, nos autovalores  $\lambda_2$ ,  $\lambda_4$  e  $\lambda_5$  são apresentados na Figura 4.5, na qual, percebe-se novamente que o erro nos autovalores diminui a medida que a matriz torna-se quadrada.

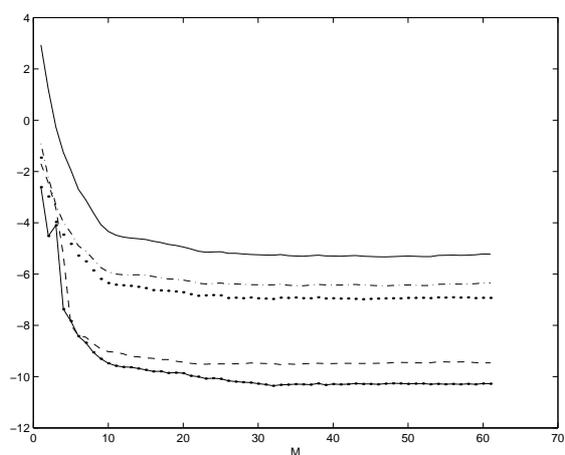


Figura 4.5: Limitante (4.32): linha contínua,  $\|T - \tilde{T}\|$ : linha tracejada e pontilhada, e  $|\lambda_l - \tilde{\lambda}_l|$  para  $l = 2, 4, 5$ : linhas pontilhada, tracejada e contínua-pontilhada, respectivamente.

## 4.4 Análise de Erro: Caso Polinomial Matricial

Analogamente ao caso escalar, o objetivo é deduzir estimativas para o erro  $|\tilde{\lambda}_j - \lambda_j|$ , em que  $\tilde{\lambda}_j$  é o autovalor extraído da matrix  $\tilde{\mathcal{C}}_A(m, q) = \tilde{\mathcal{V}}^* \tilde{\mathcal{C}}_A \tilde{\mathcal{V}}$ , na qual é assumido que  $\tilde{\mathcal{V}}$  é uma aproximação para  $\mathcal{V}$ , e  $\tilde{\mathcal{C}}_A$  é uma aproximação para a matriz companheira em blocos  $\mathcal{C}_A$ .

A ideia chave para desenvolver a análise de erro para o caso polinomial matricial é perceber que, desconsiderando autovalores nulos,

$$\lambda(\mathcal{C}_A(m, q)) = \lambda(\mathcal{V}^* \mathcal{C}_A \mathcal{V}) = \lambda(\mathcal{V} \mathcal{V}^* \mathcal{C}_A).$$

Ou seja, a ideia é perceber que o erro no autovalor  $\lambda_j$  também pode ser estimado a partir das matrizes  $\mathcal{V} \mathcal{V}^* \mathcal{C}_A$  e  $\tilde{\mathcal{V}} \tilde{\mathcal{V}}^* \tilde{\mathcal{C}}_A$ . Para simplificar a análise e notação, estas matrizes serão denotadas por  $A_{\mathcal{P}}$  e  $\tilde{A}_{\mathcal{P}}$  respectivamente. Ou seja

$$A_{\mathcal{P}} = \mathcal{V} \mathcal{V}^* \mathcal{C}_A, \quad \tilde{A}_{\mathcal{P}} = \tilde{\mathcal{V}} \tilde{\mathcal{V}}^* \tilde{\mathcal{C}}_A. \quad (4.43)$$

Portanto, as estimativas de erro que deduziremos são baseadas no resultado clássico da teoria de perturbação de autovalores:

$$|\lambda_j - \tilde{\lambda}_j| \leq \kappa_{j,q} \|\tilde{A}_{\mathcal{P}} - A_{\mathcal{P}}\|_2, \quad j = 1, \dots, n, \quad (4.44)$$

em que  $\kappa_{j,q}$  são os números de condição do autovalor  $\lambda_j$  considerado como autovalor da matriz  $A_{\mathcal{P}}$ . A principal vantagem desta abordagem é que os  $\kappa_{j,q}$ 's são exatamente os números de condição  $\kappa(\lambda_j, \mathcal{C}_A(m, q))$  estudados anteriormente (veja Bazán [14]), mais importante ainda, é que o erro em norma  $\|\tilde{A}_{\mathcal{P}} - A_{\mathcal{P}}\|_2$ , como veremos a seguir, é simples de se estimar.

### 4.4.1 Estimativas para $\|A_{\mathcal{P}} - \tilde{A}_{\mathcal{P}}\|_2$

Começaremos com uma definição preliminar.

**Definição 4.4.1** Sejam  $\mathcal{S}$ ,  $\tilde{\mathcal{S}}$  dois subespaços de  $\mathbb{R}^n$  com a mesma dimensão. Sejam  $P$  e  $\tilde{P}$  as matrizes de projeção ortogonal sobre  $\mathcal{S}$  e  $\tilde{\mathcal{S}}$ , respectivamente. Então a distância entre os subespaços  $\mathcal{S}$  and  $\tilde{\mathcal{S}}$  é definida como

$$d(\mathcal{S}, \tilde{\mathcal{S}}) = \|P - \tilde{P}\|_2. \quad (4.45)$$

Um resultado bem conhecido da área de Álgebra linear numérica é [59, p. 603]

$$d(\mathcal{S}, \tilde{\mathcal{S}}) = \sin(\theta), \quad (4.46)$$

em que  $\theta$  denota o maior ângulo canônico entre  $\mathcal{S}$  e  $\tilde{\mathcal{S}}$ . Detalhes sobre ângulos canônicos são encontrados em [59, p. 603], veja também Apêndice A.6.

**Lema 4.4.2** *Sejam  $A_{\mathcal{P}}$  e  $\tilde{A}_{\mathcal{P}}$  as matrizes introduzidas em (4.43) tal que o último vetor coluna em blocos de  $C_A$ ,  $X_A^+$ , é a solução de norma mínima do sistema (4.22) e  $\tilde{X}_A$  é o último vetor coluna em bloco da matriz  $\tilde{C}_A$ . Então*

$$\|A_{\mathcal{P}} - \tilde{A}_{\mathcal{P}}\|_2 \leq \sqrt{\sin(\theta)^2 + \|\eta\|_2^2}, \quad (4.47)$$

em que  $\sin(\theta)$  é a distância entre os subespaços gerados por  $\mathcal{V}$  e  $\tilde{\mathcal{V}}$ , e  $\eta = X_A^+ - \tilde{X}_A$ .

**Demonstração:** Particione os operadores de projeção ortogonal  $\mathcal{P} = \mathcal{V}\mathcal{V}^*$  e  $\tilde{\mathcal{P}} = \tilde{\mathcal{V}}\tilde{\mathcal{V}}^*$  como

$$\mathcal{P} = [p_1, p_2, \dots, p_N] \quad \text{e} \quad \tilde{\mathcal{P}} = [\tilde{p}_1, \tilde{p}_2, \dots, \tilde{p}_N], \quad (4.48)$$

em que  $p_i$  e  $\tilde{p}_i$ ,  $i = 1 : N$  são de ordem  $N \times q$ .

Utilizando (4.43) é imediato que  $A_{\mathcal{P}} - \tilde{A}_{\mathcal{P}} = [\epsilon_2, \dots, \epsilon_N, \eta]$ . Portanto,

$$\begin{aligned} (A_{\mathcal{P}} - \tilde{A}_{\mathcal{P}})(A_{\mathcal{P}} - \tilde{A}_{\mathcal{P}})^T &= \epsilon_1 \epsilon_1^T + \dots + \epsilon_N \epsilon_N^T + \eta \eta^T - \epsilon_1 \epsilon_1^T \\ &= (\mathcal{P} - \tilde{\mathcal{P}})(\mathcal{P} - \tilde{\mathcal{P}})^T + \eta \eta^T - \epsilon_1 \epsilon_1^T, \end{aligned}$$

Tomando a norma-2 em ambos os membros segue que

$$\|A_{\mathcal{P}} - \tilde{A}_{\mathcal{P}}\|_2^2 \leq \sin(\theta)^2 + \|\eta\|_2^2, \quad (4.49)$$

o qual prova (4.47). □

**Observação 4.4.3** Em problemas práticos, o subespaço  $\text{span}(\tilde{\mathcal{V}})$  e os coeficientes  $\tilde{X}_A$  são computados a partir de uma matriz de informações com estrutura Hankel em blocos:  $\tilde{H} = H + E$ , em que  $H$  e  $E$  são desconhecidas. Portanto, quantidades que aparecem no lado direito de (4.43) dependem da quantidade de ruído nos dados e do número de condição da matriz de informações  $\kappa(H)$ . É bem conhecido que se  $\|E\|_2 \ll \sigma_d(H)$ , então  $\text{span}(\tilde{\mathcal{V}}) \approx \text{span}(\mathcal{V})$  em cujo caso  $\sin(\theta) \approx 0$ ; e a mesma conclusão pode ser tirada em

relação a  $\|\eta\|_2$  desde que a matriz de Hankel não seja mal condicionada. (veja, Hansen [65, Teor. 3.21, p. 56]).

A condição  $\|E\|_2 \ll \sigma_d(H)$  depende da própria seqüência  $\{h_k\}$  e da dimensão da matriz Hankel em bloco  $H$ . Para o caso escalar ( $h_k \in \mathbb{C}$ ), foi provado por Bazán e Toint [16] que essa condição é passível de acontecer quando a matriz de Hankel é quadrada e a dimensão é suficientemente grande. Na prática, verifica-se freqüentemente que  $\sigma_d(H)$  cresce mais rapidamente que  $\|E\|_2$ . Porém, nada é conhecido em relação ao caso multivariável. Entretanto, existem evidências empíricas que tal condição é satisfeita quando a dimensão da matriz Hankel em blocos é suficientemente grande (desde que a seqüência de matrizes  $\{h_k\}$  não seja dominada pelo ruído); alguns exemplos podem ser vistos em [6].

Finalmente, deve-se observar que, se o erro relativo nos coeficientes  $X_A^+$  é pequeno, ou seja se  $\|\eta\|_2^2 \ll \|X^+\|_2^2$ , então, devido ao fato de que  $\|X^+\|_2$  aproxima de zero quando  $N$  é suficientemente grande<sup>2</sup>, conclui-se que uma estimativa razoável para o erro em  $A_{\mathcal{P}}$  é:

$$\|A_{\mathcal{P}} - \tilde{A}_{\mathcal{P}}\|_2 \approx \sin(\theta). \quad (4.50)$$

Substituindo (4.47) em (4.44) e utilizando a estimativa para  $\kappa(\mathcal{C}_A(m, 1))$  dada no Teorema 4.2.2, obtém-se o seguinte resultado.

**Teorema 4.4.4** *Utilizando as mesmas notações do Teorema 4.2.2, a seguinte estimativa vale*

$$|\tilde{\lambda}_j - \lambda_j| \leq \left[ 1 + \frac{d-1 + \|f^\dagger\|_2^2 + \prod_{j=1}^d |\lambda_j|^2 - \sum_{j=1}^d |\lambda_j|^2}{(d-1)\delta_j^2} \right]^{(d-1)/2} \times \quad (4.51)$$

$$\times (\sin(\theta)^2 + \|\eta\|_2^2)^{1/2}, \quad j = 1, \dots, d.$$

Já que os fatores do lado direito de (4.51) dependem fortemente da dimensão da matriz de Hankel, se  $H$  é suficientemente grande e se os autovalores não são muito próximos uns dos outros e próximos do círculo unitário, então, exceto para os casos em que o nível de ruído nos dados é extremamente alto, o erro nos autovalores pode ser estimado como

$$|\tilde{\lambda}_j - \lambda_j| \approx \sin(\theta), \quad j = 1, \dots, d. \quad (4.52)$$

Esta conclusão vem do fato de que o fator da esquerda se aproxima de 1 enquanto que o fator da direita aproxima de  $\sin(\theta)$ . Autovalores próximos do círculo unitário aparecem freqüentemente em conexão com estruturas dinâmicas muito flexíveis. Um exemplo que ilustra a estimativa (4.52) é dado a seguir.

---

<sup>2</sup>A prova de que  $\|X^+\| \rightarrow 0$  quando  $N \rightarrow \infty$  é analoga a do Teorema 4.1.2

### 4.4.2 Análise de Erro: Um Caso Real (Mini-Mast)

Ilustraremos a potencialidade das estimativas (4.51) como um instrumento de avaliação de erro nos autovalores da estrutura MIni-Mast. Para tanto, matrizes de Hankel em blocos de várias dimensões foram perturbadas por ruídos aleatórios com média zero e distribuição Gaussiana. A seguir, foram calculados as estimativas dadas por  $\sin(\theta)$ , o próprio erro  $|\tilde{\lambda}_j - \lambda_j|$ , a norma do erro na matriz de dados:  $\|E\|_2 = \|\tilde{H} - H\|_2$ , e  $\sigma_d(H)$ , o menor valor singular não nulo da matriz de Hankel ( $d = 10$ , neste caso).

São apresentados os valores médios de 100 realizações correspondentes a dois níveis de ruído, especificados pelo desvio padrão dos números aleatórios.

Os resultados correspondente ao desvio padrão  $8 \times 10^{-9}$  (nível baixo) são apresentados na Figura 4.6 (a) e (b). A parte (a) ilustra que  $\sigma_d(H)$  realmente cresce mais rapidamente que a norma da matriz de perturbações  $\|E\|_2$ , enquanto que a parte (b) mostra que  $\sin(\theta)$  é de fato uma aproximação razoável para o erro nos autovalores (veja (4.52)). Um fato bastante interessante a ser observado é que o erro tende a decrescer a medida que a dimensão da matriz de Hankel aumenta.

Os resultados correspondentes ao outro nível de ruído, com desvio padrão igual a  $4.75 \times 10^{-6}$  (nível alto), são apresentados na Figura 4.7. Esta figura repete a figura anterior mas com matrizes de maior dimensão. Neste caso o desvio padrão foi calculado

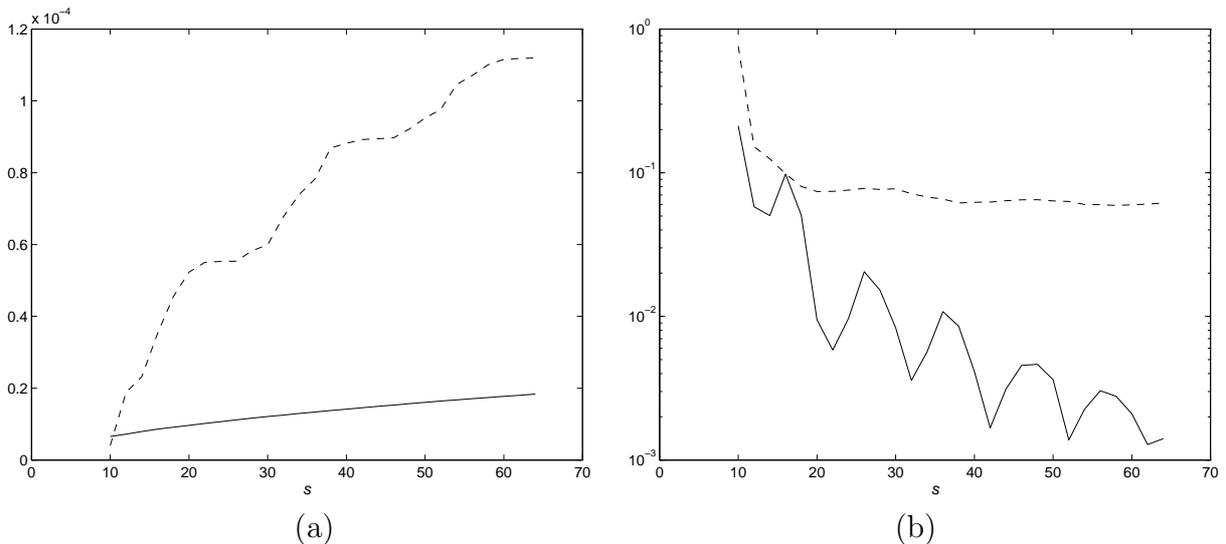


Figura 4.6: Resultados utilizando ruído baixo: (a)  $\sigma_d(H)$  (linha tracejada) e valor médio de  $\|E\|_2$  (linha sólida) como funções da dimensão da matriz  $H$ . (b) Valor médio do erro máximo em valor absoluto de  $\lambda_j$  (linha sólida) e valor médio da estimativa dada por  $\sin(\theta)$  (linha tracejada).

de modo que o valor singular  $\sigma_d(H)$  seja aproximadamente dominado pelo ruído (veja Figura 4.7-(a)). A qualidade dos autovalores estimados deteriora um pouco em relação aos do caso anterior, isto pode ser apreciado na Figura 4.7-(b)). Analogamente ao caso anterior, o erro ainda tende a decrescer com o aumento da dimensão.

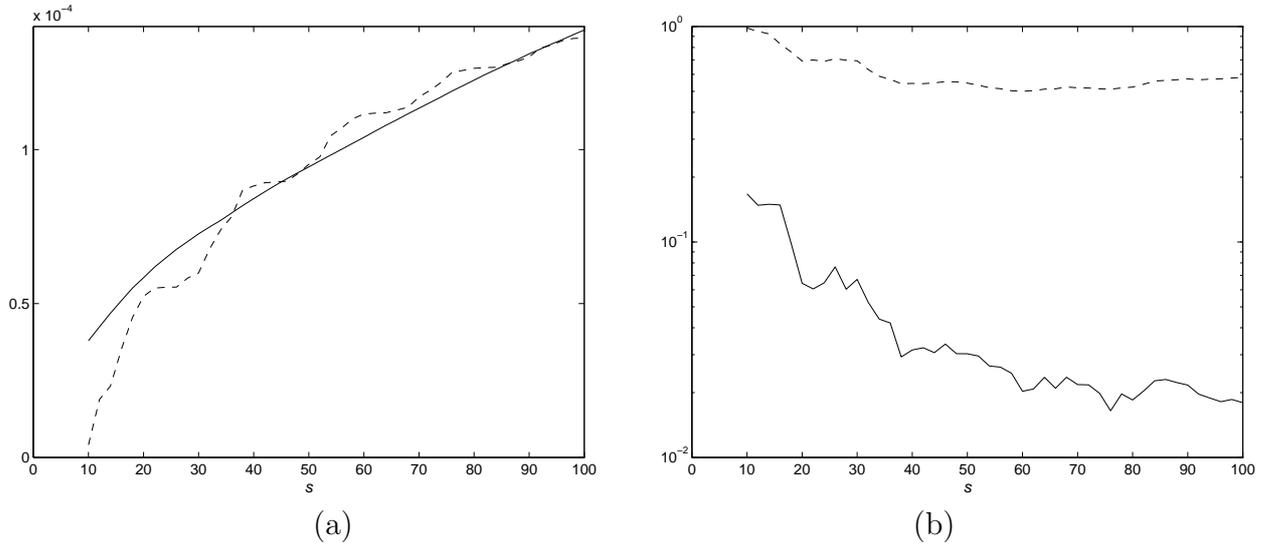


Figura 4.7: Resultados utilizando ruído alto: (a)  $\sigma_d(H)$  (linha tracejada) e valor médio de  $\|E\|_2$  (linha sólida) como funções da dimensão da matriz  $H$ . (b) Valor médio do erro máximo em valor absoluto de  $\lambda_j$  (linha sólida) e valor médio da estimativa dada por  $\sin(\theta)$  (linha tracejada).



# Capítulo 5

## Computação de Autovalores

### 5.1 Introdução

Vários problemas práticos requerem o cálculo de apenas alguns autovalores da matriz associada ao problema e respectivos autovetores. Por exemplo, a instabilidade local de um sistema dinâmico próximo a um ponto de singularidade é consequência da existência de autovalores, da matriz jacobiana do sistema, com parte real positiva. Uma técnica utilizada para verificar se uma matriz tem autovalores com parte real positiva é utilizar transformações de Möbius que leva o semiplano complexo  $Re z > 0$ , ao círculo unitário [24]. Outro exemplo é o problema de compressão de imagens digitalizadas via decomposição em valores singulares (*SVD*) (ver apêndice) da matriz de dados  $X$ : armazenam-se apenas os maiores valores singulares e respectivos vetores singulares associados a  $X$  (à esquerda e à direita). Uma ilustração simples (ver [42]) é a seguinte: no Matlab várias imagens estão disponíveis em arquivos *.mat*, por exemplo, *clown.mat*, *mandrill.mat*, *earth.mat*, *durer.mat* etc. Os comandos

```
load mandrill.mat; [U,S,V]=svd(X); colormap('gray');  
image(U(:,1:k)*S(1:k,1:k)*V(:,1:k)')
```

produzem uma imagem com resolução que depende da aproximação de  $X$  a partir da aproximação via decomposição parcial  $U_k S_k V_k^T$ , em que  $S_k$  é uma matriz diagonal formada pelos  $k$  valores singulares dominantes (os  $k$  maiores valores singulares). O comando *load mandrill.mat* cria uma variável  $X$  na qual é armazenada a matriz de dados cuja ordem, no caso, é  $480 \times 500$ . O comando *image(X)* produz a imagem original.

Os valores singulares de uma matriz  $A$  são as raízes quadradas dos autovalores positivos da matriz  $A^H A$ . Como calcular os autovalores de uma matriz? Computar o seu polinômio característico implica em muitas operações numéricas, o que pode ocasionar muitos erros de arredondamento. E, uma vez computado, o cálculo de suas raízes tem

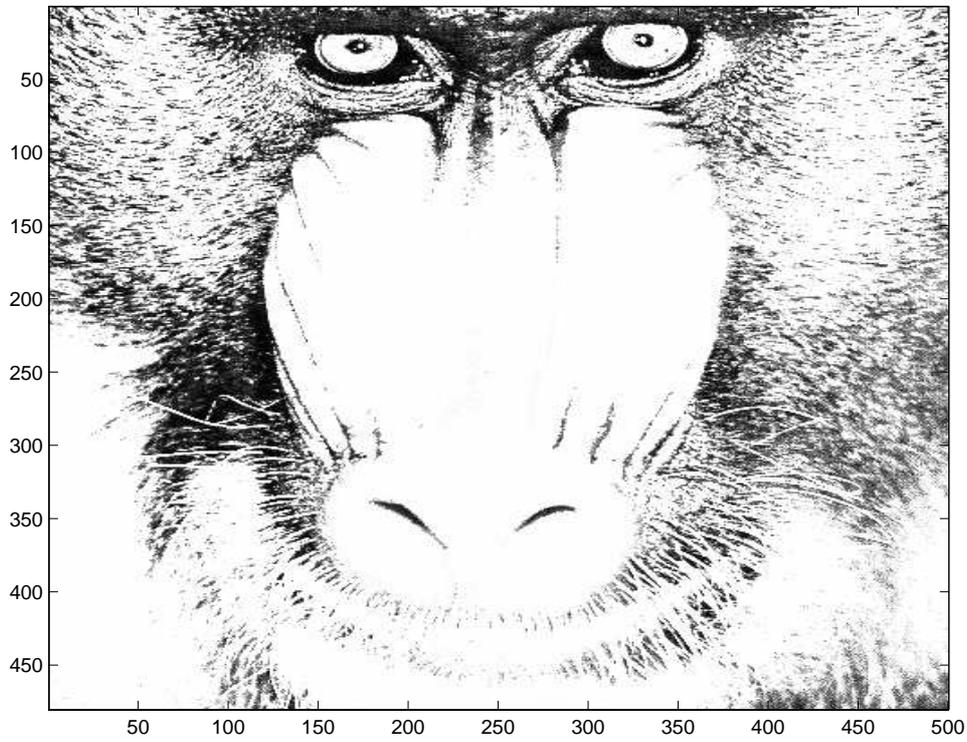
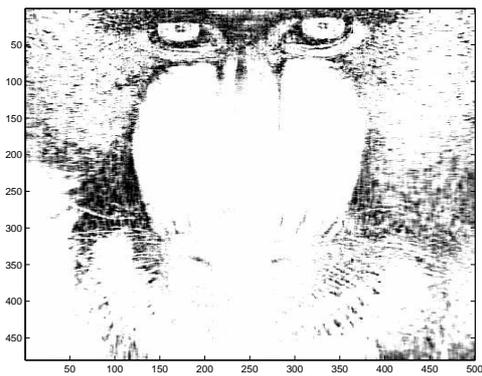
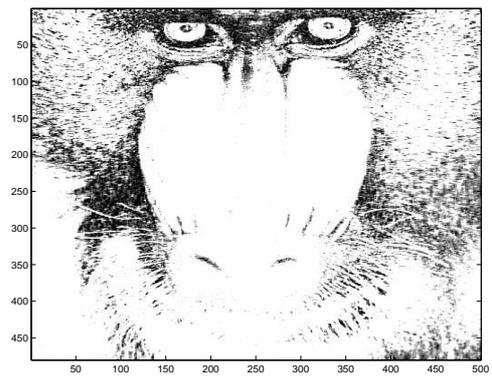


Figura 5.1: Imagem original

Figura 5.2:  $k=40$ Figura 5.3:  $k=80$

que ser feito por métodos iterativos, se a ordem da matriz for maior que 4 - não há formas fechadas de resolução de polinômios de grau  $\geq 5$ , conforme Teoria de Galois. Há métodos para a computação de autovalores, baseados em álgebra matricial. Em Matlab (ou no sistema iterativo Octave, distribuído com o Linux) calculam-se todos os autovalores de uma matriz pela função *eig*. Essa função é construída a partir de uma implementação otimizada do Método QR, que é um método de decomposição espectral muito eficiente. As implementações práticas desse método transformam o problema de calcular os autovalores da matriz  $A$  em outro: o de calcular os autovalores da sua forma Hessemberg (ou forma quasi-triangular). Então procura-se desacoplar o problema (diminuir a ordem da matriz) por iterações que tornam o elemento  $a_{n,n-1}$  cada vez mais próximo de zero. E, assim, por diante, até calcular todos os autovalores. Técnicas para zerar esses elementos abaixo da diagonal principal da matriz de Hessemberg fazem com que a convergência ocorra preferencialmente para os autovalores dominantes da matriz. Uma técnica de decomposição parcial do espectro é inserir critérios de parada quando se atinge um determinado número de autovalores convergidos.

Neste capítulo, discutiremos brevemente o método QR e apresentaremos outros métodos numéricos que calculam alguns autovalores de uma matriz (não necessariamente dominantes), e a seguir, discutiremos um conjunto de métodos para calcular os valores singulares dominantes de matrizes provenientes de problemas HR. O capítulo termina com uma breve apresentação sobre métodos numéricos para autovalores de polinômios matriciais.

## 5.2 Autovalores Matriciais

Seja  $A \in \mathbb{C}^{n \times n}$  uma matriz diagonalizável tal que os seus  $n$  autovalores,  $\lambda_1, \dots, \lambda_n$ , satisfazem  $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_n|$ . Seja  $\{v_1, \dots, v_n\}$  base de autovetores de  $\mathbb{C}^{n \times n}$ , associados respectivamente aos autovalores acima. Seja  $x_0 \in \mathbb{C}^{n \times n}$ ,  $v \neq 0$ , tal que  $v = a_1 v_1 + \dots + a_n v_n$ ,  $a_1 \neq 0$ .

O método de potência ( $z_k = Ax_{k-1}$ ,  $x_k = z_k/c_k$ , em que  $c_k$  é coordenada de  $z_k$  de máximo valor absoluto) é tal que  $x_k$  converge para um vetor do autoespaço gerado por  $v_1$ , o autoespaço dominante de dimensão 1 (porque associado ao maior autovalor em valor absoluto);  $c_k$  converge para  $\lambda_1$ . É só observar que o vetor  $x_k$  está no subespaço gerado por  $u_k = A^k x_0$ . Assim,

$$\begin{aligned} u_k &= a_1 \lambda_1^k v_1 + \dots + a_n \lambda_n^k v_n = \\ &= \lambda_1^k \left[ a_1 v_1 + a_2 \left( \frac{\lambda_2}{\lambda_1} \right)^k v_2 + \dots + a_n \left( \frac{\lambda_n}{\lambda_1} \right)^k v_n \right] \end{aligned}$$

e, logo, quanto maior for  $k$ , menor a influência das direções dos autovetores  $v_2, \dots, v_n$  em  $x_k$ .

Em resumo, dados  $A$  e  $x_0$  como acima, o algoritmo abaixo gera vetores  $x'_k$ s cada vez mais próximos de  $[v_1]$ , o subespaço gerado pelo vetor  $v_1$ . A rapidez com que esse método converge depende do quociente  $|\frac{\lambda_2}{\lambda_1}|$ .

$$\begin{aligned} z_k &= Ax_{k-1} \\ c_k &= e_k^T z_k, \quad |c_k| = \|z_k\|_\infty \\ x_k &= z_k/c_k \end{aligned}$$

Há outras versões do método de potência nas quais a normalização é realizada com outras normas [59].

Exemplo: seja  $A = \begin{pmatrix} 1 & 2 \\ -1 & 4 \end{pmatrix}$ . Seja  $x_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ . Assim,  $x_1 = \begin{pmatrix} 1/2 \\ 1 \end{pmatrix}$ ,  $c_1 = 4$ ;  $x_2 = \begin{pmatrix} 5/7 \\ 1 \end{pmatrix}$ ,  $c_2 = 4 - 1/2 = 7/2$ ;  $x_3 = \begin{pmatrix} 19/23 \\ 1 \end{pmatrix}$ ,  $c_3 = 4 - 5/7 = 23/7$ ;  $x_4 = \begin{pmatrix} 65/73 \\ 1 \end{pmatrix}$ ,  $c_4 = 4 - 19/23 = 73/23$ ; etc. É fácil ver que  $x_k$  tende ao vetor  $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ , que é autovetor associado a 3, para o qual  $c_k$  tende por sua vez.

Podemos pensar o método de potência como um caso particular de métodos do tipo

$$x_k = \frac{f(A)x_{k-1}}{c_k},$$

onde  $f(A)$  é em geral uma função analítica e  $c_k$ , um normalizador, para evitar números muito grandes (note que o conceito de número grande faz sentido em aritmética de ponto flutuante). Se tudo correr bem, o vetor de iteração converge para um autovetor associado ao maior autovalor de  $f(A)$  em valor absoluto, isto é,

$$\max_{1 \leq i \leq n} |f(\lambda_i)|.$$

Vimos acima o método de potência clássico, com  $f(A) = A$ . Observamos também que a taxa de convergência depende da razão dos dois maiores autovalores em valor absoluto. Observe que o quociente agora é entre os dois maiores autovalores em valor absoluto de  $f(A)$ ; logo, se eu estiver interessado em computar  $\lambda_{12}$ , por exemplo,  $f$  deve ser de tal modo que  $f(\lambda_{12})$  seja o maior dos autovalores, em valor absoluto. Interessante seria se tivéssemos uma função que fosse ao mesmo tempo fácil de ser computada e para a qual a

razão entre o segundo maior autovalor em valor absoluto de  $f(A)$  e  $f(\lambda_{12})$  fosse o menor possível. Uma boa escolha é a função

$$f(A) = (A - \mu I)^{-1},$$

onde  $\mu$  é uma estimativa do autovalor em que eu estou interessado. Nesse caso, a direção mais importante é aquela associada ao autovalor mais próximo de  $\mu$ , ou seja, a

$$\left| \frac{1}{\lambda - \mu} \right| = \max_{1 \leq i \leq n} \left| \frac{1}{\lambda_i - \mu} \right|.$$

Quanto mais próximo for  $\mu$  de algum autovalor, mais rápido o método convergirá. Assim, além da escolha do vetor inicial, é importante a escolha do deslocamento (*shift*) inicial  $\mu_0$ . O próprio método pode dar melhores estimativas para deslocamentos durante o processamento. Por exemplo, o algoritmo seguinte utiliza uma estratégia em que deslocamentos são atualizados a partir de iterações anteriores:

$$\begin{aligned} x_0 &\neq 0 \\ \text{Para } k &\geq 1 \\ y_k &= (A - \mu_{k-1}I)^{-1}x_{k-1} \\ c_k &= e_k^T y_k, \quad |c_k| = \|y_k\|_\infty \\ x_k &= y_k / c_k \\ \mu_k &= \mu_{k-1} + 1/c_k \end{aligned}$$

Quando  $A$  é real e simétrica, se  $x$  é um vetor real não nulo,

$$r(x) = \frac{x^T A x}{x^T x}$$

minimiza a forma quadrática  $\|Ax - \lambda x\|^2$ . Esse número é chamado de **quociente de Rayleigh** de  $x$ . O algoritmo de iteração inversa, nesse caso, pode ser reescrito da seguinte forma:

$$\begin{aligned} \|x_0\|_2 &= 1 \\ r_0 &= r(x_0) \\ \text{Para } k &\geq 1 \\ y_k &= (A - r_{k-1}I)^{-1}x_{k-1} \\ x_k &= y_k / \|y_k\|_2 \\ r_k &= r(x_k) \end{aligned}$$

Esse algoritmo, para matrizes simétricas, converge quase sempre (quando  $\|Ax_k - r_k x_k\|_2 < tol$ , existe um autovalor  $\lambda$  de  $A$  tal que  $|r_k - \lambda| < tol$  [42]). A convergência é em geral cúbica. Ou seja, o número de dígitos corretos triplica assintoticamente a cada passo.

### 5.2.1 Transformações de Möbius

Outras funções interessantes são as do tipo

$$f(x) = \frac{x + b}{x + d}.$$

Observe que  $f(x) = 1 + (b - d)/(x + d)$ . No caso de uma matriz  $A$ ,  $f(A) = I + (b - d)(A + dI)^{-1}$ , ou seja, potência com  $f$  não é muito diferente do método de iteração inversa com deslocamento [24], [23].

Essas funções são interessantes pois transformam semiplanos complexos em círculos, o que faz com que o problema de se calcular autovalores em semiplanos resume-se a calcular autovalores numa região limitada.

## 5.3 Métodos de Iteração Simultânea

Generalizar o método de potência para calcular alguns autovalores dominantes de uma matriz implica em realizar algumas operações com os vetores resultantes da multiplicação matriz-vetor a fim de evitar que convirjam todos para o autoespaço dominante gerado por  $v_1$ , por exemplo, no caso de  $|\lambda_1| > |\lambda_2|$ . Uma operação seria normalizar os vetores de algum modo, para que os vetores não fiquem muito grandes; outra, para evitar que os vetores se tornem linearmente dependentes, seria, por exemplo, ortogonalizá-los (Iteração Ortogonal), ou fazer uma decomposição  $LU$  da matriz de vetores etc.

Seja  $A \in \mathbb{C}^{n \times n}$ . Vamos supor que  $A$  é diagonalizável, que  $\lambda_1, \dots, \lambda_n$  são seus autovalores e que

$$|\lambda_1| \geq \dots \geq |\lambda_p| > |\lambda_{p+1}| \geq \dots \geq |\lambda_n|.$$

Um método interessante é uma generalização do método de Rayleigh-Ritz para vários vetores [35]:

$$\begin{aligned} X_0 &\in \mathbb{C}^{n \times n-p}, Y_0 \in \mathbb{C}^{n \times n-p}, Y_0^T X_0 = I \\ \text{Para } k &\geq 0 \\ B_k &= (Y_k^T X_k)^{-1} (Y_k^T A X_k) \\ P_k D_k P_k^{-1} &= B_k \\ X_{k+1} &= X_k P_k \\ Y_{k+1} &= Y_k P_k^{-T} \end{aligned}$$

$D_k$ , quando  $k$  cresce, tende a  $\lambda_1, \dots, \lambda_p$ .  $X_k$  tende aos respectivos autovetores à direita e  $Y_k$  tende aos respectivos autovetores à esquerda. Uma implementação de um método de iteração simultânea (também chamado de método de iteração em subespaço) pode ser visto em [5].

### 5.3.1 Métodos de Arnoldi

Os métodos de Arnoldi computam uma base ortonormal  $Q_p$  de um subespaço de Krylov -  $q, Aq, A^2q, \dots, A^{p-1}q$  - tal que  $Q_p^T A Q_p = H_p$  é uma matriz de Hessenberg cujos autovalores são próximos de autovalores de  $A$ . O comando *eigs* do Matlab utiliza a versão contida no pacote ARPACK [84].

#### Arnoldi Básico

Seja  $A \in \mathbb{R}^{n \times n}$ . Seja  $Q = (q_1 \cdots q_n)$  uma matriz ortogonal tal que  $Q^T A Q = H$ ,  $H$  hessenberg superior. Logo,  $AQ = QH$  e, comparando colunas, temos que

$$Aq_k = \sum_{i=1}^{k+1} h_{ik} q_i \quad 1 \leq k \leq n-1.$$

Assim,

$$h_{k+1,k} q_{k+1} = Aq_k - \sum_{i=1}^k h_{ik} q_i = r_k$$

e  $h_{ik} = q_i^T Aq_k$ ,  $i = 1, \dots, k$ . Logo, se  $r_k \neq 0$ ,

$$q_{k+1} = r_k / h_{k+1,k}, \quad h_{k+1,k} = \|r_k\|_2.$$

Dado então um vetor  $q_1$  unitário, o procedimento de Arnoldi é o seguinte:

```

r_0 = q_1
h_{1,0} = 1
k = 0
enquanto h_{k+1,k} ≠ 0
    q_{k+1} = r_k / h_{k+1,k}
    k = k + 1
    r_k = Aq_k
    para i = 1, ..., k
        h_{ik} = q_i^T w
        r_k = r_k - h_{ik} q_i
    fim
    h_{k+1,k} = \|r_k\|_2
fim

```

Se  $w = Aq_k$ , o *loop* externo realiza o processo de ortogonalização de Gram-Schmidt; se  $w = r_k$ , o processo de ortogonalização de Gram-Schmidt modificado. Podemos observar que os vetores  $q_1, \dots, q_k$  formam uma base para o espaço de Krylov  $[q_1, Aq_1, \dots, A^{k-1}q_1]$ . O

método gera a cada  $k$  passos uma matriz  $Q_k$  com  $k$  colunas ortonormais tal que  $AQ_k = Q_k H_k + r_k e_k^T$ , em que  $H_k$  é uma matriz hesseberg superior. Obviamente, se  $r_k = 0$ , chegamos a um espaço invariante. No caso em que o resíduo é diferente de zero, a pergunta é se as colunas de  $Q_k$  geram um espaço próximo a um espaço invariante por  $A$ . Nesse caso, os autovalores de  $H_k$  seriam aproximações de autovalores de  $A$ . Esses autovalores são chamados de **valores de Ritz**. Dados  $\lambda$ , um valor de Ritz, e  $z$ , um autovetor de  $H_k$  associado a  $\lambda$ , o vetor  $x = Q_k z$  é dito um vetor de Ritz associado a  $\lambda$ . Finalmente, observemos ainda que o método depende crucialmente da escolha do vetor inicial  $q_1$ .

### 5.3.2 Arnoldi com Recomeço

Uma estratégia para vitalizar esse método seria recomeçar o método após alguns passos, digamos  $k$ , a partir de um novo vetor  $q$  escolhido de algum modo no espaço coluna de  $Q_k$ . Como esse espaço é de Krylov,  $q$  é da forma  $p(A)q_1$ , em que  $p$  é um polinômio de grau  $(k - 1)$ . A construção desses polinômios é feita de acordo com a parte do espectro que se quer evitar calcular, isto é, como se diz no jargão de analista numérico, filtrar [114], [115], [106]. Implementações de métodos de Arnoldi aparecem também em [85], [84], [89], [116], [117], entre outros.

## 5.4 Método QR

Seja  $A \in C^{n \times n}$  tal que seus autovalores,  $\lambda_1, \dots, \lambda_n$ , são tais que

$$|\lambda_1| > \dots > |\lambda_n| > 0.$$

Seja  $A = U_0 R_0$ , a fatoração QR de  $A$ . Tomando  $U_0$  como ponto de partida para o método de iteração ortogonal, obteríamos a seguinte sequência:

$$\begin{aligned} \text{Para } k \geq 0 \\ A_{k+1} &:= AU_K \\ U_{k+1} R_{k+1} &:= A_{k+1} \text{ (fatoração QR de } A_{k+1}) \end{aligned}$$

Da sequência acima, concluímos que

$$A^2 = U_1 R_1 R_0, \dots, A^{k+1} = U_k R_k R_{k-1} \cdots R_1 R_0,$$

ou seja, a cada passo  $k$  a matriz  $U_k$  é a matriz unitária da decomposição QR de  $A^{k+1}$ . Assim, o primeiro vetor coluna de  $U_k$  tende ao espaço  $[x_1]$ , o segundo, ao espaço  $[x_1, x_2]$ , etc. Logo,  $U_k^H A U_k$  tende a uma matriz triangular superior - a forma de Schur de  $A$ , com

os autovalores na diagonal, em ordem decrescente por valor absoluto. Esse ainda não é o método QR, mas tem estreita ligação com ele. O método QR se baseia no seguinte algoritmo:

$$\begin{aligned} A_0 &:= A \\ \text{Para } k &\geq 0 \\ Q_k R_k &:= A_k \text{ (fatoração QR de } A_k) \\ A^{k+1} &:= R_k Q_k (= Q_k^H A_k Q_k) \end{aligned}$$

Apliquemos agora o algoritmo acima à matriz  $A = \begin{pmatrix} 6 & 2 \\ 2 & 3 \end{pmatrix}$ , cujos autovalores são 7 e 2:

$$\begin{aligned} Q_0 R_0 &= \frac{1}{\sqrt{40}} \begin{pmatrix} 6 & -2 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} \sqrt{40} & 9\sqrt{40}/20 \\ 0 & 7\sqrt{40}/20 \end{pmatrix} := A, \\ A_1 &:= R_0 Q_0 = \begin{pmatrix} 7 - 1/10 & 7/10 \\ 7/10 & 2 + 1/10 \end{pmatrix}, \\ Q_1 R_1 &= \frac{1}{\sqrt{4810}} \begin{pmatrix} 69 & -7 \\ 7 & 69 \end{pmatrix} \begin{pmatrix} \sqrt{4810}/10 & 63/\sqrt{4810} \\ 0 & 140/\sqrt{4810} \end{pmatrix} := A_1, \\ A_2 &:= R_1 Q_1 = \begin{pmatrix} 7 - 4/481 & 98/481 \\ 98/481 & 2 + 4/481 \end{pmatrix}. \end{aligned}$$

Observemos que a sequência, já nos dois primeiros termos, mostra tendência de convergir para a matriz diagonal formada por 7 e 2. Mas, para cada decomposição QR são necessárias  $O(n^3)$  operações se a matriz não é esparsa. Há um modo desse número de operações diminuir: operar com uma matriz conjugada à matriz original por uma matriz unitária, a sua forma de Hessemberg  $H$ . Uma matriz de Hessemberg superior é uma matriz da forma

$$\begin{pmatrix} \times & \times & \cdots & \cdots & \times \\ \times & \times & \cdots & \cdots & \times \\ 0 & \times & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \times & \times \end{pmatrix}.$$

Ou seja, é uma matriz  $H$  tal que  $h_{ij} = 0$ , se  $i \geq j + 2$ . A matriz de Hessemberg inferior é a transposta de uma Hessemberg superior.

**[Forma de Hessemberg]**  $\forall A \in C^{n \times n}$  existe uma matriz unitária  $Q$  tal que  $Q^H A Q = H$ , matriz de Hessemberg superior.

A prova da proposição acima se baseia num procedimento que gera a matriz unitária  $Q$  como um produto de transformações de Householder [59]. Notemos que uma matriz de Hessemberg conjugada a uma matriz hermitiana é uma matriz tridiagonal. A seguir algumas propriedades interessantes das matrizes de Hessemberg:

- a fatoração QR de uma matriz de Hessemberg envolve  $O(n^2)$  operações e, se ela for tridiagonal, apenas  $O(n)$ ;
- se  $H = QR$ ,  $Q$  e  $RQ = Q^H H Q$  são também de Hessemberg.

Agora, observemos o seguinte exemplo:

$$QR = \frac{1}{\sqrt{20}} \begin{pmatrix} 4 & 2 \\ 2 & -4 \end{pmatrix} \begin{pmatrix} \sqrt{20} & \sqrt{20}/2 \\ 0 & 0 \end{pmatrix} := \begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix}$$

$$A_1 := RQ = \begin{pmatrix} 5 & 0 \\ 0 & 0 \end{pmatrix}.$$

O método convergiu acima numa só iteração. De modo geral, se  $A$  é de Hessemberg, o posto de  $A$  é  $r < n$  e  $A = QR$ , então  $R$  tem  $r$  linhas nulas correspondentes às  $r$  colunas de  $Q$  que estão no espaço ortogonal ao espaço coluna de  $A$ . Assim, o produto  $RQ$  tem  $r$  linhas nulas e o problema de autovalores de  $A$  se desacopla em problemas menores de autovalores. A idéia é fazer então um deslocamento em  $A$  próximo a algum autovalor. Que deslocamento escolher? Já vimos que o método QR tem estreita ligação com o método de potência ( $f(A) = A$ ), que privilegia as direções associadas aos maiores autovalores de  $A$  em valor absoluto. Se a matriz  $A$  for real e seus autovalores  $\lambda_1, \dots, \lambda_n$  forem tais que  $|\lambda_1| > \dots > |\lambda_n|$  (logo, todos reais), a entrada  $(n, n-1)$  de  $A_k$  tende a ficar pequena,

$$A_k = \begin{pmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & 0 & \epsilon & \hat{\lambda}_n \end{pmatrix},$$

ou seja, a entrada  $(n, n)$  de  $A_k$  tende a  $\lambda_n$  e, assim por diante.

Vejamos então como fica o método QR com esta estratégia de deslocamento aplicado à mesma matriz  $A$  de um exemplo anterior:

$$Q_0 R_0 = \frac{1}{\sqrt{13}} \begin{pmatrix} 3 & 2 \\ 2 & -3 \end{pmatrix} \cdot \begin{pmatrix} \sqrt{13} & 6/\sqrt{13} \\ 0 & 4/\sqrt{13} \end{pmatrix} := A - 3I$$

$$A_1 := R_0 Q_0 + 3I = \begin{pmatrix} 7 - 1/13 & 8/13 \\ 8/13 & 2 + 1/13 \end{pmatrix}$$

$$\begin{aligned}
Q_1 R_1 &= \frac{1}{\sqrt{4033}} \begin{pmatrix} 63 & 8 \\ 8 & -63 \end{pmatrix} \begin{pmatrix} \sqrt{4033}/13 & 8/\sqrt{4033} \\ 0 & 64/13\sqrt{4033} \end{pmatrix} \\
&:= A_1 - \left(2 + \frac{1}{13}\right) I \\
\begin{pmatrix} 7 - 1/52429 & 512/52429 \\ 512/52429 & 2 + 1/52429 \end{pmatrix} &= R_1 Q_1 + \left(2 + \frac{1}{13}\right) I
\end{aligned}$$

Vemos acima que o método QR com deslocamento converge muito mais rapidamente que o método QR simples (pelo menos nas duas primeiras iterações).

**Método QR com deslocamento** Seja  $A \in C^{n \times n}$  uma matriz de Hessenberg superior. O algoritmo seguinte é dito o algoritmo QR com deslocamento simples:

$$\begin{aligned}
A_0 &:= A \\
\text{Para } k &\geq 0 \\
\mu_k &:= (A_k)_{nn} \\
Q_k R_k &:= A_k - \mu_k I \text{ (fatoração QR de } A_k - \mu_k I) \\
A^{k+1} &:= R_k Q_k + \mu_k I \text{ (= } Q_k^H A_k Q_k)
\end{aligned}$$

#### Observações:

1) Se  $A$  é uma matriz real não simétrica, há um modo de se fazer dois deslocamentos complexos conjugados em aritmética real [59].

2) Se  $A$  é uma matriz real, tridiagonal e simétrica, QR com deslocamento de Wilkinson (o autovalor da submatriz  $A(n-1:n, n-1:n)$  mais próximo de  $A(n, n)$ ) converge globalmente [95].

### 5.4.1 Matrizes Simétricas Reais

Os algoritmos de decomposição parcial do espectro de matrizes simétricas reais, em geral, realizam suas operações sobre a forma tridiagonal (Hessenberg simétrica) da matriz. Além dos métodos de iteração simultânea citados na seção anterior, o método de bisseção é muito utilizado (ver [42]). O método de biseção explora bem o teorema da Inércia de Sylvester e está implementado no LAPACK ([2]).

Os métodos descritos para matrizes genéricas quando transcritos para matrizes simétricas reais têm um comportamento mais controlado, com resultados mais precisos quanto à velocidade de convergência, por exemplo. O que queremos dizer é que os algoritmos são os mesmos, porém implementados em aritmética real e com convergência assegurada em

muitos casos (pelo menos, em aritmética exata). É o caso, por exemplo, do método QR aplicado a matrizes de Jacobi [41]. Métodos de Lanczøs (é como se chamam os métodos de Arnoldi no caso simétrico real) com recomeços estão bem explicados em [8]. Uma coletânea dos resultados mais interessantes dessa referência, serão descritos da próxima seção em diante, dando ênfase a métodos para o cálculo de alguns valores singulares da matriz de Hankel proveniente de problemas HR. Outras referências são [42], com uma análise do método de Lanczøs em aritmética exata e em aritmética de ponto flutuante.

## 5.5 Computação de Valores Singulares Dominantes

Daqui em diante, assumiremos que  $A \in \mathbb{C}^{n \times n}$  é uma matriz Hermitiana obtida de um produto do tipo  $H^*H$ , com  $H \in \mathbb{C}^{m \times n}$ , tal que provavelmente  $\sigma_d(H) \gg \sigma_{d+1}(H)$ . Os métodos a serem discutidos visam calcular estimativas para os  $d$  primeiros valores singulares e os subespaços singulares correspondentes, de matrizes de Hankel  $H$  provenientes de problemas HR.

### 5.5.1 O Procedimento de Rayleigh-Ritz

Seja  $A = H^*H$ ,  $A \in \mathbb{C}^{n \times n}$ , onde  $H$  denota a matriz  $H(l)$  definida em (2.6), e  $Q \in \mathbb{C}^{n \times d}$  uma matriz ortonormal, cujas  $d$  primeiras colunas são aproximações dos  $d$  autovetores associados aos  $d$  autovalores desejados. A idéia do Procedimento de Rayleigh-Ritz é aproximar os  $d$  autovalores de  $A$ , pelos autovalores de  $B = Q^*AQ$ , a qual possui dimensão  $d \times d$ . Os autovalores de  $B$  são chamados valores de Ritz. Esta aproximação é feita da seguinte maneira:

#### 5.1 Algoritmo Procedimento de Rayleigh Ritz

Aplicado a  $A = H^*H$ . Dada  $Q \in \mathbb{C}^{n \times d}$ ,  $Q^*Q = I$ .

Para  $k=1, 2, \dots$  faça

1.  $AQ = H^*(HQ)$

2.  $B = Q^*(AQ)$

3. Calcule os autovetores e autovalores de  $B$ ,  $Bg_i = \tau_i g_i$  e os vetores de Ritz

$$y_i = Qg_i, \quad i = 1, \dots, d$$

4. Calcule os resíduos  $S_i = Ay_i - \tau_i y_i$ ,  $i = 1, \dots, d$

Observe que se  $d = 1$ , então no passo 2 tem-se que  $B = Q^*AQ$  é exatamente o Quociente de Rayleigh Ritz,  $\rho(q, A)$ . Assim, para  $d > 1$  tem-se uma generalização do Quociente de Rayleigh Ritz. Além disso,  $B$  é Hermitiana e positiva definida, o passo 3 pode ser calculado de maneira rápida e com baixo tempo computacional. Para aproveitar a estrutura da matriz  $H$ , e a multiplicação rápida, a multiplicação do passo 1 é dividida em duas etapas. Primeiramente é calculado  $HQ$  e a seguir  $H^*(HQ)$ . As novas aproximações dos  $d$  autovalores e autovetores desejados, são dadas pelos valores de Ritz e pelos vetores de Ritz, respectivamente. Parlett [95, pag. 234] demonstra três maneiras nas quais estas aproximações são ótimas.

A primeira pode ser vista como um corolário do Teorema do Minimax (veja, por exemplo, [59]).

$$\lambda_j(A) = \min_{\mathcal{F}^j \subset C^n} \max_{f \in \mathcal{F}^j} \frac{f^* A f}{f^* f} \quad (f \neq 0), \quad j = 1, 2, \dots, n$$

onde  $\mathcal{F}^j$  é um subespaço de  $C^n$ , com dimensão  $j$ .

Seja  $\mathcal{Q}^d = \text{span}(Q)$ , e  $\mathcal{G}^j$  um subespaço de  $\mathcal{Q}^d$ , com dimensão  $j$ . Assim, uma definição natural da melhor aproximação para  $\lambda_j$ , no subespaço  $\mathcal{Q}^d$  é:

$$\sigma_j(A) = \min_{\mathcal{G}^j \subset \mathcal{Q}^d} \max_{g \in \mathcal{G}^j} \frac{g^* A g}{g^* g} \quad (g \neq 0), \quad j = 1, 2, \dots, d.$$

A segunda abordagem, define uma matriz de resíduos  $R(C) = AQ - AC$  e mostra que  $B = Q^*AQ$  minimiza o resíduo, isto é,  $\|R(B)\| < \|R(C)\|$ .

De um terceiro modo, Parlett verifica que os valores e vetores de Ritz são os autovalores e autovetores da projeção de  $A$  em  $\mathcal{Q}^d$ , ou seja, da matriz que gera em  $\mathcal{Q}^d$ , o subespaço mais próximo de  $\text{span}(A)$ .

O valor mínimo da norma da matriz de resíduos,  $R(C)$ , pode ainda ser visto como uma medida de quão perto  $\mathcal{Q}^d$  está de se tornar um subespaço invariante de  $A$ . Se  $\mathcal{Q}^d$  for um subespaço invariante de  $A$ , então o produto  $Aq_i$ , onde  $q_i$  é uma coluna de  $Q$ , será uma combinação linear das colunas de  $Q$ , isto é,  $Aq_i = Qc$ . Assim, existe uma matriz  $C$ , tal que  $R(C) = 0$ . Se  $\mathcal{Q}^d$  não é um subespaço invariante de  $A$ , então não existe nenhuma matriz  $C$ , tal que  $\|R(C)\| = 0$ , porém,  $B = Q^*AQ$  continua minimizando  $R(C)$ .

Seja agora  $X = (x_1, x_2, \dots, x_d)$  uma base ortonormal de  $\mathcal{Q}^d$  e  $\Delta$  uma matriz diagonal,  $\Delta = \text{diag}(\delta_1, \delta_2, \dots, \delta_d)$ , então  $\|AX - X\Delta\|$  é minimizada quando  $x_i = y_i$  e  $\delta_i = \tau_i$ ,  $i = 1, \dots, d$ . De fato: Seja  $Y = (y_1, \dots, y_d)$ ,  $\Phi = \text{diag}(\tau_1, \dots, \tau_d)$  e  $G = (g_1, \dots, g_d)$

$$\begin{aligned} \|AY - Y\Phi\| &= \|AQG - QG\Phi\| = \|AQ - QG\Phi G^*\| = \|AQ - QBGG^*\| = \\ &= \|AQ - QB\| = \|R(B)\|. \end{aligned}$$

Mesmo quando  $X \neq Y$  ou  $\Delta \neq \Phi$ , ainda é possível expressar  $X$  na base  $Q$ ,  $X = QD$ , onde  $D^*D = DD^* = I$ , pois elas geram o mesmo subespaço. Obtém-se então

$$\|AX - X\Delta\| = \|AQD - QD\Delta\| = \|AQ - QD\Delta D^*\| > \|R(B)\|. \quad (5.1)$$

Assim, quando  $\mathcal{Q}^d$  é um subespaço invariante de  $A$ , (5.1), juntamente com o fato de  $B = Q^*AQ$  satisfazer  $R(B) = 0$ , significam que, os valores e vetores de Ritz são exatamente os autovalores e autovetores de  $A$ .

Quando os autovalores de  $A$  são bem separados, uma vez calculados os resíduos  $s_i$ , pode-se facilmente obter limitantes para os valores de Ritz. Estes limitantes podem ser obtidos a partir do teorema abaixo, cuja demonstração pode ser encontrada em [87].

**Teorema 5.5.1** *Seja  $Q \in C^{n \times d}$ ,  $Q^*Q = I$ , e seja  $(\tau_i, y_i)$ ,  $i = 1, \dots, d$ , os valores e vetores de Ritz de  $A$ , com resíduos  $s_i = Ay_i - \tau_i y_i$ . Então, cada intervalo*

$$[\tau_i - \|s_i\|, \tau_i + \|s_i\|]$$

*possui um autovalor de  $A$ .*

Como em cada intervalo correspondente aos valores de Ritz, existe um autovalor de  $A$ , se todos estes intervalos forem disjuntos, teremos  $d$  autovalores de  $A$ . Porém, se alguns deles se interceptarem, podemos ter dois valores de Ritz convergindo para o mesmo autovalor de  $A$ . No teorema abaixo, tem-se um limitante adicional para valores de Ritz em intervalos que se interceptam, sua demonstração pode ser encontrada em [95, cap. 11].

**Teorema 5.5.2** *Seja  $Q \in C^{n \times d}$ ,  $Q^*Q = I$ , e seja  $(\tau_i, y_i)$ ,  $i = 1, \dots, d$ , os valores e vetores de Ritz de  $A$ , com resíduos  $s_i = Ay_i - \tau_i y_i$ . Então, existem  $d$  autovalores de  $A$ ,  $\lambda'_i$ ,  $i = 1, \dots, d$ , tais que*

$$|\tau_i - \lambda'_i| \leq \|S\|$$

*onde,  $S = (s_1, \dots, s_d) = AY - Y\Phi$ .*

A determinação de limitantes para os vetores de Ritz é um pouco mais complicada, pois, autovetores associados a autovalores com multiplicidade algébrica  $\neq 1$ , não são unicamente determinados. Qualquer combinação de autovetores correspondentes ao mesmo autovalor, é um autovetor. Portanto, não é possível determinar uma estimativa de erro para estes autovetores. Porém, este fato é irrelevante, visto que o mais interessante é o subespaço gerado por estes autovetores. Da mesma forma, vetores de Ritz correspondentes a autovalores muito próximos, tendem a ser sensíveis e darem estimativas ruins dos correspondentes autovetores, porém, o subespaço por eles gerado pode ser uma boa aproximação do subespaço associado a estes autovalores muito próximos.

### 5.5.2 Iteração em Subespaços

A primeira versão da Iteração em Subespaços surgiu em 1957, e foi introduzida por Bauer, com o nome de *Treppeniteration* (iteração escada). A Iteração em Subespaços pode ser vista como uma generalização do Método das Potências, só que, ao invés de iterar um vetor, a iteração em subespaços itera  $d$  vetores simultaneamente, e obtém aproximações para os  $d$  autovetores de  $A$ , associados aos  $d$  maiores autovalores. Infelizmente, esta generalização apresenta um problema. Suponha uma matriz inicial  $Q^{(0)} = (q_1^{(0)}, q_2^{(0)}, \dots, q_d^{(0)})$ ,  $Q^{(0)H}Q = I$  e multiplique  $A$  iterativamente. Então, após  $j$  iterações obtém-se

$$Q^j = A^j Q^{(0)} = (A^j q_1^{(0)}, A^j q_2^{(0)}, \dots, A^j q_d^{(0)}).$$

Pela teoria do Método das Potências, [129, cap. 4], pode-se deduzir que todas as colunas de  $Q^j$  irão convergir para o mesmo autovetor de  $A$ , o qual está associado ao maior autovalor. Este problema surge pois, as colunas de  $Q^j$  tornam-se linearmente dependentes. Para evitar isto, basta manter a ortogonalidade entre as colunas de  $Q^j$ , através de um passo de ortogonalização em cada iteração. A seguir, é apresentado o algoritmo da Iteração em Subespaços.

#### 5.2 Algoritmo Iteração em Subespaços

Seja  $A = H^*H$ . Dada  $Q^{(0)} \in C^{m \times d}$ ,  $(Q^{(0)})^*Q^{(0)} = I$ .

Para  $j = 1, 2, \dots$  faça:

1.  $Z^{(j)} = AQ^{(j-1)} = H^*(HQ^{(j-1)})$
2. Ortonormalize  $Z^{(j)} = X^{(j)}R^{(j)}$ , onde  $R^{(j)}$  é uma matriz triangular superior e  $X^{(j)}$  é unitária.
3. Faça  $Q^{(j)} = X^{(j)}$
4. Teste a convergência de cada coluna

Assim como no Método das Potências, a convergência da Iteração em Subespaços depende do *gap* entre os autovalores, neste caso no *gap* entre o primeiro autovalor não desejado e os  $d$  desejados. Assim, se este *gap* for pequeno, para obter uma convergência mais rápida, basta aumentar o número de autovalores desejados, de forma que este *gap* fique maior. O passo 1 é feito em duas etapas, para aproveitar a multiplicação rápida.

### 5.5.3 Iteração em Subespaços com Aceleração de Ritz

Uma maneira de acelerar a convergência da iteração em subespaços, é usar o Procedimento de Rayleigh Ritz em cada iteração. Após o passo 2 do algoritmo 5.2, os vetores de Ritz são calculados a partir da matriz  $X^j$ , e as colunas da nova  $Q^j$  serão estes vetores. Esta idéia foi criada por Rutishauser, na década de 60, e é chamada Iteração em Subespaços com Aceleração de Ritz (SIR). Como as aproximações ótimas dos autovetores são calculadas em cada iteração, as componentes nas direções destes vetores são ampliadas, o que aumenta a razão de convergência.

#### 5.3 Algoritmo Iteração em Subespaços com Aceleração de Ritz.

Seja  $A = H^*H$ . Dada uma matriz inicial  $Q^{(0)} \in C^{m \times d}$ .

Para  $j = 1, 2, \dots$  faça:

1.  $Z^{(j)} = HQ^{(j-1)}$
2. Calcule a svd reduzida de  $Z^{(j)}$ :  $[X_1, \Sigma_1, Y_1^*] = svd(Z^{(j)})$ , onde  $X_1 \in C^{m \times d}$ ,  $\Sigma_1 \in R^{d \times d}$ ,  $Y_1 \in C^{d \times d}$
3.  $W^{(j)} = H^*X_1$
4. Calcule a svd reduzida de  $W^{(j)}$ :  $[X_2, \Sigma_2, Y_2^*] = svd(W^{(j)})$ , onde  $X_2 \in C^{m \times d}$ ,  $\Sigma_2 \in R^{d \times d}$ ,  $Y_2 \in C^{d \times d}$
5.  $Q^{(j)} = X_2$

Observe que o algoritmo 5.3, não deixa explícito passos da Iteração em Subespaços, nem passos da Aceleração de Ritz. Contudo, se for considerado que a *SVD* é usada somente para calcular uma base ortogonal de uma matriz, então os passos 1,2 e 3 equivalem ao passo 2 do algoritmo 5.2. O passo 4 calcula uma base ortogonal para  $H^*HQ$ , e esta base é usada como a nova  $Q$ . Desta forma, o algoritmo 5.3 é equivalente ao algoritmo 5.2, porém, com um passo extra de ortogonalização, o passo 2.

Para ficarem explícitos os passos da Aceleração de Ritz, é necessário levar em conta os efeitos dos cálculos da *SVD*. Seja  $B := Q^*H^*HQ = Y_1\Sigma_1^2Y_1^*$  a correspondente decomposição autovalor para a *SVD* de  $HQ$ , calculada no passo 2, assim o quadrado dos elementos de  $\Sigma_1$  são os valores de Ritz de  $A = H^*H$ , e as colunas de  $QY_1$  seus vetores de Ritz. Como os vetores de Ritz são boas aproximações dos autovalores de  $A$ , eles também são boas aproximações dos vetores singulares à direita de  $H$ , os  $v_i$ 's. Através da relação  $u_i = \tau_i^{-1}Hv_i$ , é fácil verificar que as colunas de  $X_1$  são aproximações dos vetores singulares à esquerda de  $H$ . Analogamente, verifica-se que as colunas de  $X_2$  são aproximações dos

vetores singulares à direita de  $H$ . Ou seja, o algoritmo acima apresenta dois passos da Aceleração de Ritz.

## 5.6 Método de Lanczos

O Método de Lanczos foi criado em 1950 por C. Lanczos, era utilizado para encontrar uma matriz tridiagonal  $T$ , unitariamente similar a uma matriz simétrica  $A$ ,  $T = Q^*AQ$ , onde  $Q = (q_1, q_2, \dots, q_n)$  e  $Q^*Q = I$ . Em aritmética exata, isto pode ser feito em  $n$  passos, porém, em aritmética finita aparecem problemas numéricos, decorrentes da perda de ortogonalidade das colunas de  $Q$ . Por esta razão, durante alguns anos o Método de Lanczos foi substituído por métodos mais estáveis, como os métodos de Givens [56] e Householder [68].

O interesse pelo Método de Lanczos só retornou em 1971, quando Paige [93] fez uma análise de erro detalhada do método, mostrando sua eficiência no cálculo de somente uma parte do espectro de uma matriz simétrica. A partir daí, devido à sua simplicidade, pois somente utiliza multiplicações do tipo matriz-vetor, e também devido à qualidade das suas aproximações, o Método de Lanczos tem sido fonte de diversas pesquisas, especialmente no sentido de evitar a perda de ortogonalidade das colunas de  $Q$ . Inicialmente, o método é descrito na sua forma básica, a seguir são apresentadas outras versões, as quais evitam a perda de ortogonalidade de diferentes maneiras.

### 5.6.1 Método de Lanczos Básico

O Método de Lanczos está baseado no fato que de é possível, em um número finito de passos, transformar uma matriz Hermitiana  $A$ , em uma matriz tridiagonal  $T$ , por meio de transformações unitárias, [59].

$$T = Q^*AQ. \quad (5.2)$$

Ainda mais, se  $Q^*AQ = T$ , onde  $T$  é uma matriz tridiagonal real, com os elementos das diagonais secundárias todos positivos, então  $T$  e  $Q = (q_1, q_2, \dots, q_n)$  são unicamente determinados por  $A$  e  $q_1$ . A demonstração deste fato pode ser encontrada em [95, cap. 7]. A hipótese de os elementos das diagonais secundárias serem positivos, não é uma restrição essencial.

Equacionando as  $j$  primeiras colunas de (5.2) obtém-se

$$AQ_j = Q_jT_j + r_j e_j^T, \quad (5.3)$$

onde  $r_j = \beta_j q_{j+1}$  e  $Q_j = (q_1, \dots, q_j)$ ,  $Q_j^*Q_j = I$ ,  $T_j$  é o menor  $j \times j$  de  $T$

$$T_j = \begin{pmatrix} \alpha_1 & \beta_1 & & & \\ \beta_1 & \alpha_2 & \beta_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \beta_{j-1} \\ & & & \beta_{j-1} & \alpha_j \end{pmatrix},$$

com  $\alpha_i = q_i^* A q_i$  e  $\beta_i = \|r_i\|$ . Os vetores  $q_i$  são chamados vetores de Lanczos.

Similarmente, a equação (5.3) pode ser rescrita como

$$r_j = \beta_j q_{j+1} = A q_j - \alpha_j q_j - \beta_{j-1} q_{j-1}. \quad (5.4)$$

A partir desta equação é deduzida a recursão de Lanczos, a qual dado um vetor inicial  $q_1$ , gera a cada iteração, um novo vetor de Lanczos, e uma nova linha e coluna são adicionadas a matriz  $T$ . Os autovalores da matriz  $T$  convergem para os autovalores de  $A$ .

#### 5.4 Algoritmo Método de Lanczos Básico.

Seja  $A = H^* H$ . Dados  $r_0 \neq 0$ ,  $\beta_0 = \|r_0\|$ .

Para  $j = 1, 2, \dots$  faça:

1.  $q_j := r_{j-1} / \beta_{j-1}$
2.  $u_j := A q_j = H^*(H q_j)$
3.  $r_j := u_j - \beta_{j-1} q_{j-1}$
4.  $\alpha_j := q_j^* r_j$
5.  $r_j := r_j - \alpha_j q_j$
6.  $\beta_j := \|r_j\|$
7. Se desejar, calcule os valores de Ritz  $\tau_i$  e vetores de Ritz  $y_i$ . Se a quantidade desejada de valores e vetores de Ritz convergiu, então pare.

Quando é completado um ciclo, isto é, quando são completados os passos de 1 até 6, tem-se um passo de Lanczos. Este procedimento gera vetores de Lanczos até que  $\beta_j = 0$ , o que deve ocorrer para algum  $j \leq n$ . Em particular, múltiplos autovalores de  $A$ , forçam  $\beta_j = 0$ , para algum  $j < n$  (veja [131]). Os vetores de Lanczos formam uma base ortogonal no subespaço gerado por  $(q_1, A q_1, \dots, A^j q_1)$  e os autovetores aproximados estão

neste subespaço. Portanto, o método não é capaz de detectar nenhum autovetor ortogonal a  $q_1$ , o que significa que somente é possível determinar um autovetor, do subespaço correspondente a um autovalor múltiplo.

Da equação (5.4) tem-se que  $r_j$  é ortogonal a  $q_{j-1}$  e  $q_j$ . Pode-se mostrar, [95], que  $r_j$  é ortogonal a todos os vetores de Lanczos  $(q_1, \dots, q_j)$ . Os vetores  $\alpha_j q_j$  e  $\beta_{j-1} q_{j-1}$  são as projeções ortogonais de  $Aq_j$  em  $q_j$  e  $q_{j-1}$ . Isto significa que o próximo vetor de Lanczos é obtido da ortogonalização de  $Aq_j$  em relação a  $q_j$  e  $q_{j-1}$  e os vetores  $q_i$ ,  $1 \leq i \leq j$ , formam uma base ortogonal no subespaço de Krylov

$$\mathcal{K}_j(A, q_1) = \text{span}\{q_1, Aq_1, \dots, A^{j-1}q_1\}. \quad (5.5)$$

Usando a ortogonalidade dos vetores de Lanczos e a relação (5.3), verifica-se que  $T_j = Q_j^* A Q_j$ , isto é,  $T_j$  é a projeção de  $A$  em  $\text{span}(Q_j)$ . Assim, no passo 7 do algoritmo 5.4, os valores de Ritz são calculados como os autovalores de  $T_j$

$$T_j g_i^{(j)} = \tau_i^{(j)} g_i^{(j)}, \quad i = 1, 2, \dots, j \quad (5.6)$$

e os vetores de Ritz são definidos por

$$y_i^{(j)} = Q_j g_i^{(j)}, \quad i = 1, 2, \dots, j. \quad (5.7)$$

Os teoremas 5.5.1 e 5.5.2 mostram que a norma  $\|Ay_i^{(j)} - \tau_i^{(j)} y_{(i)}\|$  é uma boa estimativa da precisão dos valores de Ritz. A princípio, é possível calcular  $\tau_i$  e  $y_i$  a partir de  $T_j$  em cada iteração do algoritmo de Lanczos, porém, felizmente, é possível calcular  $\|Ay_i^{(j)} - \tau_i^{(j)} y_{(i)}\|$  sem calcular  $y_i$ . Usando (5.3), (5.6), (5.7) e o fato de  $\|q_{j+1}\| = 1$  tem-se:

$$\begin{aligned} \|s_i^{(j)}\| &= \|Ay_i^{(j)} - \tau_i^{(j)} y_{(i)}\| = \|AQ_j g_i^{(j)} - \tau_i^{(j)} Q_j g_i^{(j)}\| \\ &= \|AQ_j g_i^{(j)} - Q_j T_j g_i^{(j)}\| = \|\beta_j q_{j+1} e_j^{(T)} g_i^{(j)}\| \\ &= \beta_j |e_j^{(T)} g_i^{(j)}| =: \beta_{ij}. \end{aligned} \quad (5.8)$$

Assim, a convergência dos valores de Ritz para os autovalores de  $A$ , pode ser verificada pelos primeiros elementos dos autovetores normalizados de  $T_j$ , evitando o cálculo dos vetores de Ritz. Este resultado explica porque alguns valores de Ritz podem ser precisos, mesmo quando  $\beta_j$  não é muito pequeno.

Como já foi citado anteriormente, em aritmética finita muitas das propriedades do procedimento de Lanczos são perdidas. Nas primeiras iterações, os resultados não diferem muito do processo exato. Até que um novo vetor de Lanczos é calculado, e este numericamente deixa de ser ortogonal aos anteriores. Algumas iterações mais tarde,  $Q_j$  deixou de

ter posto completo, isto é, suas colunas são linearmente dependentes. E então começam os problemas, a partir daí, o método começa a gerar cópias de vetores de Ritz já calculados, pois os novos vetores de Lanczos possuem componentes nas direções dos vetores de Ritz que já convergiram. Paige ([93],[94]) fez uma análise detalhada do desempenho do método em aritmética finita, e concluiu que a ortogonalidade entre os  $q$ 's é perdida quando surge um  $\beta_{ij}$  muito pequeno. Na prática, o processo nunca terminará com  $\beta_j = 0$ , mas irá continuar calculando mais e mais cópias do mesmos vetores de Ritz. Os erros de arredondamento fazem contribuições em todas as direções, as quais crescem a medida que o procedimento é executado. Em resumo, de um lado, a presença de um  $\beta_{ij}$  pequeno indica que foi encontrada uma boa aproximação de pelo menos um autovalor de  $A$ , e de outro lado, isto também indica que as dificuldades numéricas começaram.

Uma maneira de resolver este problema, é fazer uma reortogonalização completa, isto é, manter cada novo vetor de Lanczos ortogonal aos anteriores. Só que neste caso, além do trabalho extra para a reortogonalização, é necessário manter todos os vetores de Lanczos armazenados. Esta idéia é chamada de Lanczos com reortogonalização completa. E no outro extremo, tem-se o Método de Lanczos sem reortogonalização, porém, com a exigência de uma análise mais detalhada das aproximações obtidas. Entre estes dois estão os métodos que fazem uma reortogonalização parcial, ou reortogonalização seletiva, isto é, reortogonalizam somente quando há necessidade. Uma maneira de não deixar o número de vetores de Lanczos ficar muito grande, é recomeçar o processo após um número fixo de iterações, com um vetor inicial escolhido de maneira a aproveitar a estrutura da matriz  $A$ . E neste caso a reortogonalização é possível, pois somente um número fixo de vetores de Lanczos é mantido na memória. Nas próximas subseções, vamos dar mais detalhes sobre alguns destes métodos.

### 5.6.2 Método de Lanczos com Reortogonalização Completa

A maneira mais direta de resolver o problema da perda de ortogonalidade, proposta pelo próprio Lanczos, é fazer a reortogonalização de cada novo vetor de Lanczos em relação aos anteriores, obtendo assim, o seguinte procedimento

#### 5.5 Algoritmo Método de Lanczos com Reortogonalização Completa (LANCR).

Seja  $A = H^*H$ . Dados  $r_0 \neq 0$ ,  $\beta_0 = \|r_0\|$ .

Para  $j = 1, 2, \dots$  faça:

1.  $q_j := r_{j-1}/\beta_{j-1}$
2.  $u_j := Aq_j = H^*(Hq_j)$
3.  $r_j := u_j - \beta_{j-1}q_{j-1}$

$$4. \alpha_j := q_j^* r_j$$

$$5. r_j := r_j - \alpha_j q_j$$

$$5.5 r_j := r_j - q_\nu(q_\nu^* r_j), \nu = j, j-1, \dots, 1$$

$$6. \beta_j := \|r_j\|$$

7. Se desejar, calcule os valores de Ritz  $\tau_i$  e vetores de Ritz  $y_i$ . Se a quantidade desejada de valores e vetores de Ritz convergiu, então pare.

Observe que  $r_j$  é explicitamente ortogonalizado em relação a  $q_j$  e  $q_{j-1}$ . Consequentemente, é necessário manter armazenado todos os vetores de Lanczos. Esta é a desvantagem da reortogonalização completa, a qual aumenta a cada iteração. Por outro lado, este algoritmo tem a vantagem de não calcular nenhum vetor de Ritz redundante, ou seja, não calcula mais de uma vez o mesmo vetor de Ritz. E como vetores de Ritz redundantes requerem um número maior de iterações para convergir, o algoritmo 5.5 requer um número mínimo de iterações para convergir.

Quando são desejados somente alguns dos maiores autovalores, as desvantagens deste algoritmo não são tão sérias. Normalmente os valores de Ritz correspondentes a estes autovalores convergem primeiro, e não são necessárias muitas iterações. Isto depende naturalmente do espectro em questão.

Como cita Parlett, [95, pag. 303], a reortogonalização não pode por si só garantir a ortogonalidade numérica dos vetores de Lanczos. Quando o corrente vetor de Lanczos tem uma componente considerável na direção de um dos outros vetores, então pode haver algum tipo de cancelamento no passo 5.5. Segundo Parlett, um passo extra de reortogonalização é suficiente para tornar estes dois vetores numericamente ortogonais. Na próxima subseção, é visto que a reortogonalização completa é mais do que é necessário, e para diminuir a quantidade de trabalho, a reortogonalização será feita somente quando for preciso.

### 5.6.3 Lanczos com Semiortogonalização

As desvantagens da reortogonalização completa com relação ao trabalho extra e armazenamento, podem diminuir a eficiência quando os vetores de Lanczos possuem muitas componentes, ou quando o problema requer um grande número de iterações. Isto levou muitas pessoas a estudarem uma forma de transformar o método de Lanczos Básico, em um método mais estável que o LANCR. Este estudo está baseado nos trabalhos de Paige ([93] [94]), o qual fez uma análise de erro, e procurou saber quando e onde começa a perda de ortogonalidade.

Uma idéia é reortogonalizar somente quando for necessário, e somente em relação àqueles autovetores que perderam a ortogonalidade. O primeiro algoritmo deste tipo,

foi feito em 1979 por Parlett e Scott [96]. Eles publicaram um artigo com um método chamado: Lanczos com Ortogonalização Seletiva (LANSO), no qual a ortogonalidade é mantida através da ortogonalização dos novos vetores de Lanczos em relação aos vetores de Ritz que já convergiram. Em 1984, Simon [103] escreveu um artigo no qual mostrou que, mantendo os vetores num nível de ortogonalidade de  $\sqrt{\mathbf{u}}$ , isto é,  $|q_i^* q_{j+1}| \leq \sqrt{\mathbf{u}}$ ,  $1 \leq i < j$ , é suficiente para obter boas aproximações dos autovalores e autovetores, onde  $\mathbf{u}$  denota a menor unidade de arredondamento. Ele usou a noção de semiortogonalidade, quando os vetores de Lanczos satisfazem este nível de ortogonalidade. Esta análise levou a um novo algoritmo chamado Lanczos com Reortogonalização Parcial (LANPR).

#### 5.6.4 Método de Lanczos com Recomeços Implícitos

Em problemas onde não existe um gap distinto entre os autovalores desejados e os não desejados, normalmente são necessários muitos passos de Lanczos para convergir. Assim a reortogonalização completa, apresenta uma desvantagem, pois ela exige o armazenamento de todos os vetores de Lanczos. Uma forma de contornar este problema, é recomeçar as iterações após um número finito de passos de Lanczos, mantendo assim, a quantidade de vetores a ser armazenada fixa, e a reortogonalização feita somente quando for necessário. Calvetti, Reichel e Sorensen desenvolveram um método de Lanczos com recomeços implícitos (IRL), [30], o qual é uma adaptação do método de Arnoldi com recomeços implícitos, [83], para o caso Hermitiano. Este método força o vetor inicial a estar em um subespaço invariante da matriz, por meio de repetidas filtragens do vetor inicial, usando um polinômio filtro, e recomeçando as iterações implicitamente.

Este estudo está baseado no trabalho de Calvetti, Reichel e Sorensen, [30], porém, está restrito ao cálculo dos  $d$  maiores autovalores. Inicialmente, é discutido como o vetor inicial deve ser escolhido, para que o método convirja em  $d$  iterações, e produza um subespaço invariante correspondente aos  $d$  autovetores desejados. A seguir, é descrito o algoritmo (IRL), como uma mistura do método de Lanczos, com o Algoritmo  $QR$  com deslocamento implícito.

#### O Vetor Inicial

Quando se deseja aproximar  $d$  autovalores e autovetores, usando o método de Lanczos, a situação ideal é encontrar um subespaço invariante após exatamente  $d$  iterações, isto é, quando  $r_d$  da equação (5.3), se anula. Como já foi visto no início da seção 5.6.1, a iteração de Lanczos fica completamente definida pelo seu vetor inicial, assim,  $q_1$  determina se a iteração irá convergir em  $d$  iterações. Surge então uma pergunta: Quais condições o vetor inicial deve satisfazer, para que se obtenha um subespaço invariante após  $d$  iterações? O teorema abaixo afirma que se  $q_1$  for uma combinação linear de  $d$  autovetores, então

a recursão terminará em  $d$  passos, com  $r_d = 0$ . Além disto, a equação (5.4), força os vetores de Lanczos a estarem no subespaço gerado por estes  $d$  autovetores. Assim, uma boa escolha de vetor inicial, é como tal combinação.

**Teorema 5.6.1** *Seja  $A_{n \times n}$  uma matriz simétrica, e seja  $AQ_d - Q_dT_d = r_d e_d^T$ , com  $T$  simétrica não reduzida, passo  $d$  da fatoração de Lanczos de  $A$ . Então  $r_d = 0$  se e somente se  $q_1 = V_d x$ , onde*

$$AV_d = V_d \Lambda_d \quad (5.9)$$

com  $V_d^* V_d = I_d$  e  $\Lambda_d$  uma matriz diagonal de ordem  $d$ .

**Demonstração:** ( $\Rightarrow$ ) Suponha que  $r_d = 0$ , então,  $AQ_d = Q_dT_d$ . Seja  $T_d G_d = G_d \Lambda_d$  a decomposição autovalor-autovetor de  $T_d$ ,  $G_d^* G_d = I_d$ , chamando  $V_d = Q_d G_d$ , então  $AQ_d G_d = Q_d G_d \Lambda_d$ , isto é,  $AV_d = V_d \Lambda_d$ , além disso,  $\text{posto}(V_d) = d$ , e  $q_1 = Q_d e_1 = Q_d G_d G_d^* e_1 = V_d x$ , com  $x = G_d^* e_1$ .

( $\Leftarrow$ ) Suponha agora que  $AV_d = V_d \Lambda_d$ ,  $\text{posto}(V_d) = d$ , e  $q_1 = V_d x$ . Então, para qualquer inteiro  $m > 0$ ,  $A^m V_d = V_d \Lambda_d^m$  e assim, para todo  $m$ ,

$$A^m q_1 = A^m V_d x = V_d \Lambda_d^m x \in \mathcal{R}(V_d).$$

Conseqüentemente,  $\dim(\mathcal{K}^{d+1}(A, q_1)) \leq \text{posto}(V_d) = d$ . Como  $T_d$  é não reduzida, usando uma pequena modificação do Teorema 7.4.3 de Golub e Van Loan [59], tem-se que,  $\dim(\mathcal{K}^{d+1}(A, q_1)) = d$ , e então  $r_d = 0$ .  $\square$

### O Algoritmo IRL

Sejam  $d$  e  $p$ , números pré-especificados fixos, onde,  $d$  é o número de autovalores desejados, e  $p$  o número de passos extras que serão realizados. Após  $d + p$  passos de Lanczos, tem-se uma matriz  $(d + p) \times (d + p)$ , Hermitiana e tridiagonal  $T_{d+p}$ , e uma matriz  $Q_{d+p} \in C^{n \times (d+p)}$  com colunas ortonormais, tais que

$$Q_{d+p} = Q_{d+p} T_{d+p} + r_{d+p} e_{d+p}^T \quad (5.10)$$

onde  $r_{d+p} = \beta_{d+p} q_{d+p+1}$ . Por simplicidade, vamos chamar  $Q = Q_{d+p}$ ,  $T = T_{d+p}$  e  $r = r_{d+p}$ , e aplicar o algoritmo  $QR$  com shift, a esta fatoração truncada de  $A$ . Seja  $\mu_1$  um shift, calcule a decomposição  $QR$  de  $T - \mu_1 I$ ,  $T - \mu_1 I = X_1 R_1$ , onde  $X_1, R_1 \in C^{(d+p) \times (d+p)}$ ,  $X_1 X_1^* = I$  e  $R_1$  é uma matriz triangular superior.

Então somando e subtraindo  $\mu_1 Q$  na equação (5.10), e usando que  $T - \mu_1 I = X_1 R_1$ , obtém-se:

$$\begin{aligned}
AQ - \mu_1 Q - QT + \mu_1 Q &= re_{d+p}^T \\
(A - \mu_1 I)Q - Q(T - \mu_1 I) &= re_{d+p}^T \\
(A - \mu_1 I)Q - QX_1R_1 &= re_{d+p}^T,
\end{aligned} \tag{5.11}$$

multiplicando a equação (5.11) por  $X_1$  e rearranjando os termos

$$\begin{aligned}
(A - \mu_1 I)(QX_1) - (QX_1)(R_1X_1) &= re_{d+p}^T X_1 \\
A(QX_1) - (QX_1)(R_1X_1 + \mu_1 I) &= re_{d+p}^T X_1.
\end{aligned}$$

Seja  $Q^+ = QX_1$  e  $T^+ = R_1X_1 + \mu_1 I = X_1^*TX_1$ . Então  $T^+$  é simétrica, e como  $T$  é tridiagonal,  $X_1$  é Hessenberg superior, assim,  $T^+$  também é tridiagonal. Explicitando a primeira coluna de ambos os lados da equação (5.11) tem-se

$$(A - \mu_1 I)q_1 = QX_1R_1e_1 = Q^+e_1\rho_{11} = q_1^+\rho_{11}$$

onde  $q_1 = Qe_1$ ,  $\rho_{11} = e_1^TR_1e_1$  e  $q_1^+ = Q^+e_1$ . Tem-se assim, uma relação entre  $q_1$  e  $q_1^+$ . Aplicando  $p$  shifts e usando a equação (5.11), obtém-se:

$$\begin{aligned}
AQ^+ &= Q^+T^+ + re_{d+p}^T \hat{X} \\
&= (Q^+ \quad q_{d+p+1}) \begin{pmatrix} T^+ \\ \beta_{d+p} e_{d+p}^T \hat{X} \end{pmatrix},
\end{aligned} \tag{5.12}$$

onde  $q_{d+p+1} = r/\beta_{d+p}$ ,  $Q^+ = Q\hat{X}$ ,  $T^+ = \hat{X}^*T\hat{X}$  e  $\hat{X} = X_1X_2 \dots X_p$ ,  $X_i$  é a matriz ortogonal associada ao shift  $i$ .

Observe que, como  $T^+$  permanece tridiagonal a cada shift, as matrizes  $X_i$ , são Hessenberg superiores, e os primeiros  $d-1$  elementos da linha  $d+p$  em  $\hat{X}$  serão iguais a zero. O que significa que:

$$\beta_{d+p} e_{d+p}^T \hat{X} = (0, 0, \dots, \tilde{\beta}_{d+p}, b^*), \text{ onde } \tilde{\beta}_{d+p} = e_d^T \beta_{d+p}.$$

Particionando as matrizes,

$$Q^+ = \begin{pmatrix} Q_d^+ & \hat{Q}_p^+ \end{pmatrix} \quad T^+ = \begin{pmatrix} T_d^+ & \hat{\beta}_d e_d e_1^T \\ \hat{\beta}_d e_1 e_d & \hat{T}_p \end{pmatrix},$$

com  $\hat{\beta} = e_{d+1}^T T^+ e_d$  e substituindo em (5.12), tem-se:

$$A \begin{pmatrix} Q_d^+ \\ \hat{Q}_p^+ \end{pmatrix} = \begin{pmatrix} Q_d^+ \\ \hat{Q}_p^+ \\ q_{d+p+1} \end{pmatrix} \begin{pmatrix} T_d^+ & \hat{\beta}_d e_d e_1^T \\ \hat{\beta}_d e_1 e_d & \hat{T}_p \\ \tilde{\beta}_{d+p} e_d^T & b^* \end{pmatrix}. \quad (5.13)$$

Igualando as primeiras  $d$  colunas da equação (5.13), obtém-se:

$$A Q_d^+ = Q_d^+ T_d^+ + e_d^+ e_d^T = \begin{pmatrix} Q_d^+ \\ q_{d+1}^+ \end{pmatrix} \begin{pmatrix} T_d^+ \\ \beta_d^+ e_d^T \end{pmatrix}.$$

onde  $q_{d+1}^+ = \frac{1}{\beta_d^+} r_d^+$ ,  $r_d^+ = (\hat{Q}_p e_1 \hat{\beta}_d + q_{d+p+1} \tilde{\beta}_{d+p})$  e  $\beta_d^+ = \|r_d^+\|$ . Como  $(Q_d^+)^* Q_d^+ = I_d$  e  $(Q_d^+)^* q_{d+1}^+ = 0$ , então esta é outra decomposição de Lanczos de  $A$ . O novo vetor inicial pode ser escrito como  $q_1^+ = \Psi_p(A) q_1$ , onde  $\Psi_p$  é um polinômio mônico de grau  $p$ , cujas raízes são os  $p$  shifts,  $\mu_1, \mu_2, \dots, \mu_p$ . Recomeçando a partir desta fatoração, podemos calcular  $p$  passos extras de Lanczos, o que nos remete à equação (5.10). Então, aplicando alternadamente  $p$  passos de Lanczos e  $p$  shifts, obtém-se o Método de Lanczos com Recomeços Implícitos. A cada iteração obtém-se aproximações dos autovetores, a partir de um subespaço de *Krylov*, de dimensão  $d + p$ , com um custo de  $p$  multiplicações do tipo matriz-vetor, ao invés de  $d + p$ .

### A escolha dos shifts

Os  $p$  passos  $QR$  aplicados desempenham um papel de filtragem do vetor inicial, fazendo com que este esteja num subespaço invariante apropriado,  $\text{span}(V_d)$ . O lema abaixo, extraído de [30], sugere que para eliminar o conjunto de autovalores não desejados de  $T_{d+p}$ , deve-se escolher os  $p$  shifts  $\mu_j$  como sendo os autovalores que estão na parte não desejada do espectro. Segundo Calvetti [30], estes shifts são chamados de *shifts exatos*.

**Lema 5.6.2** *Seja  $\lambda(T_{d+p}) = \{\tau_1, \dots, \tau_d\} \cup \{\mu_1, \dots, \mu_p\}$  uma partição disjunta do espectro de  $T_{d+p}$  e seja*

$$T_{d+p}^+ = \hat{X}^* T_{d+p} \hat{X} \quad (5.14)$$

onde  $\hat{X} = X_1 X_2 \dots X_p$ , e  $X_j$  é implicitamente determinado pelo shift  $\mu_j$ . Seja  $\beta_j$  o elemento da posição  $(j + 1)$  da subdiagonal de  $T_{d+p}^+$ . Se  $\beta_j \neq 0$  para  $1 \leq j < d$ , então  $\beta_d = 0$  e

$$T_{d+p}^+ = \begin{pmatrix} T_d^+ & 0 \\ 0 & \hat{T}_p \end{pmatrix},$$

onde  $\lambda(T_d^+) = \{\tau_1, \dots, \tau_d\}$  e  $\lambda(\hat{T}_p) = \{\mu_1, \dots, \mu_p\}$ . Além disso,

$$q_1^+ = Q_{d+p} \hat{X} e_1 = \sum_{j=1}^d \eta_j y_j,$$

onde  $\eta_j \in C$  e  $y_j$  é o vetor de Ritz associado ao valor de Ritz  $\tau_j$ , isto é,  $y_i = Q_{d+p} g_j$  com  $T_{d+p} g_j = g_j \tau_j$  para  $1 \leq j \leq d$ .

**Demonstração:** Para simplificar a demonstração, vamos chamar  $T = T_{d+p}$  e  $Q = Q_{d+p}$ .

Seja  $x_1^{(i)} = X_1 X_2 \dots X_i e_1$  e  $\rho_{11}^{(i)} = e_1^T$ . Quando o primeiro shift é aplicado, obtém-se  $T - \mu_1 I = X_1 R_1$ . Explicitando a primeira coluna tem-se:  $\frac{1}{\rho_{11}^{(1)}}(T - \mu_1 I) = x_1^{(1)}$ . Similarmente, aplicando  $p$  shifts chega-se a:

$$x_1 = \hat{X} e_1 = \frac{1}{\rho_{11}^{(1)} \dots \rho_{11}^{(p)}} (T - \mu_p I) \dots (T - \mu_1 I) e_1 =: \psi(T) e_1,$$

onde  $\psi(\lambda) = \frac{1}{\nu} \prod_{i=1}^p (\lambda - \mu_i)$  e  $\nu = \rho_{11}^{(1)} \dots \rho_{11}^{(p)}$ .

Seja  $T = G \Lambda G^*$  a decomposição autovalor - autovetor de  $T$ , onde

$$G^* G = I, \quad \text{e} \quad \Lambda = \text{diag}(\tau_1, \dots, \tau_d, \mu_1, \dots, \mu_p).$$

Assim,

$$(T - \mu_1 I) = G(\Lambda - \mu_1 I)G^*, \quad \dots, \quad (T - \mu_p I) = G(\Lambda - \mu_p I)G^*,$$

logo

$$\begin{aligned} (T - \mu_1 I)(T - \mu_2 I) \dots (T - \mu_p I) &= G(\Lambda - \mu_1 I)(\Lambda - \mu_2 I) \dots (\Lambda - \mu_p I)G^* = \\ &= G \begin{pmatrix} \tilde{\tau}_1 & & & & \\ & \ddots & & & \\ & & \tilde{\tau}_d & & \\ & & & 0 & \\ & & & & 0 \end{pmatrix} G^* \\ &= G \tilde{\Lambda} G^*. \end{aligned}$$

Suponha que  $e_1 = \sum_{j=1}^{d+p} \varphi_j g_j = G \bar{\varphi}$ , onde  $\bar{\varphi} = [\varphi_1, \dots, \varphi_{d+p}]^T$ . Então  $x_1 = \psi(T) e_1 = G \tilde{\Lambda} G^* G \bar{\varphi} = G \tilde{\Lambda} \bar{\varphi} = G_d \bar{\varphi}_d$ , com  $G_d = [g_1, \dots, g_d]$ ,  $g_i$ ,  $1 \leq i \leq d$  são autovetores associados à parte desejada do espectro, e  $\bar{\varphi}_d = [\varphi_1, \dots, \varphi_d]^T$ . Isto mostra que as componentes de  $e_1$  na direção dos autovetores associados a  $\mu_i$ ,  $1 \leq i \leq p$  são anuladas, então  $x_1$  pode

ser escrito como  $x_1 = \sum_{j=1}^d \eta_j g_j$ . Equacionando as primeiras  $j$  colunas da relação (5.14), tem-se uma decomposição de Lanczos para a matriz  $T$ , com vetor inicial  $x_1$ . Usando uma pequena modificação do Teorema 7.4.3 de Golub e Van Loan [59], segue que  $\beta_d = 0$ ,

portanto  $T_{d+p}^+ = \begin{pmatrix} T_d^+ & 0 \\ 0 & \hat{T}_p \end{pmatrix}$ . Além disso,

$$q_1^+ = Q\hat{X}e_1 = Qx_1 = \sum_{j=1}^d \eta_j Qg_j = \sum_{j=1}^d \eta_j y_j.$$

□

O Lema 5.6.2 mostra que a escolha dos shifts exatos é matematicamente equivalente a recomençar as iterações de Lanczos com um vetor inicial que seja uma combinação linear dos vetores de Ritz de  $A$ , correspondentes aos  $d$  autovalores desejados, isto é, como uma aproximação de uma combinação linear dos  $d$  autovetores. O método IRL pode ser usado para aproximar qualquer parte do espectro de  $A$ . Contudo, o lema 5.6.2 não apresenta nenhuma informação sobre a razão de convergência, e infelizmente sob certas condições, esta pode ser muito lenta. Por exemplo, o cálculo dos menores autovalores, requer muitas iterações até convergir, pois o método precisa trabalhar muito até vencer a influência dos maiores autovalores, os quais aparecem como shifts em muitas iterações. Em [30], encontra-se uma discussão sobre outros tipos de shifts.

A seguir, é apresentado um esboço do algoritmo IRL, para aproximar os  $d$  maiores autovalores. É assumido que

$$\tau_1 \geq \tau_2 \geq \dots \geq \tau_{d+p}$$

**5.6 Algoritmo** *Método de Lanczos com Recomeços Implícitos e Shifts Exatos (IRL-ES).* Seja  $A = H^*H$ . Dados  $r_0 \neq 0$ ,  $\beta_0 = \|r_0\|$ .

1. Faça  $d + p$  passos usando o algoritmo 5.4 (reortogonalizando quando for necessário)

Repita

2. Calcule a decomposição autovalor-autovetor de  $T_{d+p}$ , e teste a convergência de  $\beta_{(d+p)i}$

3. Se a quantidade desejada de pares de Ritz convergiu, calcule os vetores de Ritz e pare.

4. Aplique o algoritmo QR com shifts  $\mu = \tau_j$ ,  $j = d + 1, \dots, d + p$

5 Faça  $p$  passos de Lanczos, começando de (5.14), e volte para o passo 2 acima.

Como os pares de Ritz possuem diferentes razões de convergência, uma forma de aumentar a convergência do método, seria remover os vetores de Ritz que já convergiram, armazenar os de interesse e, manter os vetores de Lanczos ortogonais a estes. Com isso, a dimensão do espaço de *Krylov* usado é reduzida, e como este espaço é ortogonal aos vetores de Ritz que já convergiram, as próximas iterações do método irão estimar a parte restante desejada do espectro.

O problema de remover aproximações de autovalores e autovetores que já convergiram, é muito estudado, e existem muitas técnicas de deflação. Lehoucq e Sorensen [85], desenvolveram regras de deflação para o método IRL, baseadas em técnicas do método *QR*. Basicamente tem-se dois tipos de deflação. Se o valor de Ritz pertence a parte desejada do espectro, é necessário manter o correspondente vetor de Ritz na próxima fatoração de Lanczos. Este tipo de deflação é chamado *Locking*. E se o valor de Ritz pertence a parte não desejada do espectro, é preciso removê-lo das próximas iterações, e este é chamado de *Purging*.

### Convergência

Será feita agora uma análise da convergência do IRL-ES, considerando o caso em que se deseja aproximar apenas os  $d$  maiores autovalores. A discussão está baseada no trabalho de Sorensen [106], onde primeiramente é mostrado que a medida que ocorrem as iterações, os autovalores de  $T_d$  convergem para certos limites. Então é mostrado que estes limites são os autovalores de  $A$  e, finalmente, que são os  $d$  maiores.

Inicialmente são apresentados dois lemas, que serão usados posteriormente.

**Lema 5.6.3** *Seja  $M = \begin{pmatrix} T & \beta e_d \\ \beta e_d^T & \alpha \end{pmatrix}$ , uma matriz simétrica tridiagonal. Então as raízes da equação*

$$\beta^2 e_d^T (T - \lambda I)^{-1} e_d = \alpha - \lambda \quad (5.15)$$

*são autovalores de  $M$ .*

**Demonstração:** A demonstração segue do cálculo do polinômio característico de  $M$ .

Seja  $T = G\Lambda G^*$ , a decomposição autovalor-autovetor de  $T$ , como

$$(T - \lambda I)^{-1} = G(\Lambda - \lambda I)^{-1}G^*,$$

substituindo na equação (5.15) tem-se:

$$\beta^2 \sum_{j=1}^{\alpha} \frac{\eta_j^2}{(\tau_j - \lambda)} = \alpha - \lambda. \quad (5.16)$$

onde  $(\eta_1, \eta_2, \dots, \eta_d) = e_d^T G$  e  $\Lambda = \text{diag}(\tau_1, \dots, \tau_d)$ . Da equação (5.16), pode-se concluir que nenhum autovalor de  $M$  é igual a  $\tau_i$ ,  $1 \leq i \leq d$ . Usando a propriedade do *interlacing*, veja Golub [59, pag. 396], tem-se:

$$\hat{\mu} < \tau_d < \tilde{\tau}_d < \tau_{d-1} < \dots < \tau_1 < \tilde{\tau}_1 \quad (5.17)$$

onde  $\{\tilde{\tau}_1, \dots, \tilde{\tau}_d, \hat{\mu}\}$  são autovalores de  $M$ .

Observe também que, se os elementos da subdiagonal de  $T$  são todos não nulos, então pela discussão que segue o algoritmo 5.4, os  $\tau_j$  são distintos, e nenhum dos  $\eta_j$  é zero.

O próximo lema afirma que se  $q_1$  está em um subespaço de dimensão menor do que  $d$ , então algum  $\beta_j$  deve ser zero, e deve ocorrer a deflação.

**Lema 5.6.4** *Suponha que  $AQ = QT + re_k^T$  é uma decomposição de Lanczos de  $A$ , e seja  $\beta_j$  o  $j$ -ésimo elemento da subdiagonal de  $T$ . Se  $q_1 = v\gamma + w\zeta$ , com  $\gamma^2 + \zeta^2 = 1$ ,  $\|v\| = \|w\| = 1$ ,  $v^*w = 0$  e  $v = \sum_{j=1}^i \gamma_j \zeta_j$ ,  $Av_j = \lambda_j v_j$ , onde  $\{\lambda_j\}$  é um conjunto arbitrário de  $i$  autovalores de  $A$ , então*

$$\prod_{j=1}^i \beta_j \leq \zeta \left\| \prod_{j=1}^i (A - \lambda_j I) \right\| \quad (5.18)$$

**Demonstração:** A demonstração pode ser encontrada em [106].

Com estes dois lemas em mãos, é possível começar a mostrar que o IRL-ES converge. Serão escolhidos os  $p$  menores autovalores de  $T_{d+p}$ .

Seja  $l$  o número da iteração. Então  $q_1^{(l)}$  representa o vetor inicial na iteração  $l$ , e após  $p$  passos de Lanczos tem-se:

$$AQ_{d+p}^{(l)} - Q_{d+p}^{(l)} T_{d+p}^{(l)} = r_{d+p}^{(l)} e_{d+p}^T.$$

Seja  $T_{d+p}^{(l)} = \begin{pmatrix} T_d^{(l)} & \beta_d^{(l)} e_d e_1^T \\ \beta_d^{(l)} e_1 e_d^T & \tilde{T}^{(l)} \end{pmatrix}$ , com autovalores

$$\tau_{1,l+1} > \dots > \tau_{d,l+1} > \mu_{1,l+1} > \dots > \mu_{p,l+1}.$$

Aplicando  $p$  shifts tem-se:

$$\hat{X}^{(l)H} T^{(l)} \hat{X}^{(l)} = \begin{pmatrix} T_d^{(l+1)} & 0 \\ 0 & \hat{T}_p \end{pmatrix}, \quad Q_d^{(l+1)} = (Q_{d+p}^{(l)} \hat{X}^{(l)}) \begin{pmatrix} I_d \\ 0 \end{pmatrix},$$

onde  $\hat{X}^{(l)} = X_1^{(l)} \dots X_p^{(l)}$  são construídos aplicando os shifts  $\mu_{1,l+1}, \dots, \mu_{p,l+1}$ .

O próximo lema mostra que os autovalores  $\tau_{jl}$  de  $H_d^{(l)}$  convergem para alguns valores limites.

**Lema 5.6.5** *Cada  $\tau_{jl}, l = 1, 2, \dots$  é uma seqüência convergente para cada  $j = 1, 2, \dots, d$ .*

**Demonstração:** Como  $T_{d+p}^{(l)}$  é obtida através de  $p$  abordagens sucessivas de  $T_d^{(l)}$ , segue após  $p$  aplicações do lema 5.6.4 que

$$\tau_{jl} < \tau_{j,l+1} \quad \text{para } j = 1, \dots, d.$$

Como cada  $\tau_{jl}$  é um quociente de Rayleigh com relação a  $A$ , segue que  $\lambda_n \leq \tau_{jl} \leq \lambda_1$ , para todo  $j, l$  (veja Strang [109, pag. 266]). Ou seja, a seqüência  $\{\tau_{jl}\}$  é limitada, e como é crescente, então converge.  $\square$

A seguir, é mostrado que o limite para o qual as seqüências  $\{\tau_{jl}\}_{l=1}^{\infty}$  convergem, são autovalores de  $A$ .

**Lema 5.6.6** *Seja  $T_d^{(l)} = G^{(l)} \Lambda^{(l)} G^{(l)H}$  e  $(\eta_1^l, \eta_2^l, \dots, \eta_d^l) = e_d^T G^{(l)}$ . Assuma que  $\tau_j$  são disjuntos, onde  $\tau_{jl} \rightarrow \tau_j$ . então*

$$\beta_d^l \eta_j^l \rightarrow 0, \quad \text{quando } l \rightarrow \infty, \quad j = 1, \dots, d$$

e em conseqüência

$$\|AQ^{(l)}g_j^{(l)} - Q^{(l)}g_j^{(l)}\tau_{jl}\| = |\beta^{(l)}\eta_j^l| \rightarrow 0,$$

onde  $g_j^l = G^{(l)}e_j$  para  $j = 1, \dots, d$ .

**Demonstração:** Considere a submatriz  $(d+1) \times (d+1)$  principal de  $T_{d+p}^{(l)}$

$$M = \begin{pmatrix} T_d^{(l)} & \beta_d^{(l)} e_d \\ \beta_d^{(l)} e_d^T & \alpha^{(l)} \end{pmatrix}.$$

Do lema 5.6.3 segue que os  $d$  maiores autovalores  $\hat{\tau}_{jl}$  de  $M^{(l)}$  satisfazem

$$\tau_{jl} < \hat{\tau}_{jl} < \tau_{j,l+1},$$

onde a última desigualdade surge de  $p-1$  aplicações adicionais do lema 5.6.3.

Além disso, por manipulações algébricas da equação (5.16) obtém-se:

$$(\beta\eta_j)^2 = (\tau_j - \lambda) \left[ \frac{(\alpha - \lambda) - \beta^2 \sum_{i=j+1}^d \frac{\eta_i^2}{(\tau_i - \lambda)}}{1 + \sum_{i=1}^{j-1} \frac{\eta_i^2 (\tau_j - \lambda)}{\eta_i^2 (\tau_i - \lambda)}} \right] \quad (5.19)$$

para qualquer raiz  $\lambda$ . Substituindo da matriz  $T_d^{(l)}$  as quantidades apropriadas indexadas por  $l$ , e fazendo  $\lambda = \hat{\tau}_{jl}$  tem-se:

$$(\beta_d^{(l)} \eta_j^{(l)})^2 < |(\tau_{jl} - \hat{\tau}_{jl})| \left| (\alpha^{(l)} - \hat{\tau}_{jl}) - \beta_d^{(l)^2} \sum_{i=j+1}^d \frac{\eta_j^{(l)^2}}{(\tau_{il} - \hat{\tau}_{jl})} \right|.$$

Observe que, como  $\hat{\tau}_{jl} > \tau_{jl} > \tau_{il}$  para  $i = 1, 2, \dots, j-1$ , então o denominador na equação (5.19) é maior do que 1.

A hipótese de que os limites  $\tau_j$  são distintos implica que

$$\left| (\alpha^{(l)} - \hat{\tau}_{jl}) - \beta_d^{(l)^2} \sum_{i=j+1}^d \frac{\eta_j^{(l)^2}}{(\tau_{il} - \hat{\tau}_{jl})} \right|,$$

tem limites finitos para cada  $j$ . Assim, para  $l$  suficientemente grande, existe uma constante positiva  $K$ , tal que

$$(\beta_d^{(l)} \eta_j^{(l)})^2 < K |(\tau_{jl} - \hat{\tau}_{jl})| < K |(\tau_{jl} - \tau_{j,l+1})| \rightarrow 0, \text{ quando } l \rightarrow \infty,$$

pois cada seqüência  $\{\tau_{jl}, l = 1, 2, \dots\}$  é uma seqüência crescente convergente, para cada  $j = 1, 2, \dots, d$ .  $\square$

Uma vez mostrado que os  $\tau_{jl}$  convergem para os autovalores de  $A$ , o próximo passo é mostrar que estes são os  $d$  maiores. Observe que o vetor inicial  $q_1^{(l)}$  pode ser escrito como:

$$q_1^{(l)} = \frac{\Psi_l(A)q_1}{\|\Psi_l(A)q_1\|},$$

onde  $\psi_i(\lambda) = \nu_i \prod_{k=1}^p (\lambda - \mu_{k,i})$ ,  $\Psi_l(\lambda) = \prod_{i=1}^l \psi_i(\lambda)$  e  $q_1$  é o vetor inicial original.

**Teorema 5.6.7** *Suponha que o vetor inicial  $q_1$  satisfaz  $v_j^* q_1 = \gamma_j \neq 0$  para  $j = 1, \dots, d$ , onde  $v_j$  é o autovetor de  $A$  associado ao autovalor  $\lambda_j$ , com os autovalores de  $A$  listados em ordem decrescente. Seja  $\beta_i^{(l)}$  o elemento da posição  $i$  da subdiagonal de  $T_d^{(l)}$ , e assuma que  $\beta_i^{(l)} > \epsilon > 0$  para todo  $i, l$ . Então as seqüências*

$$\tau_{jl} \rightarrow \tau_j = \lambda_j \text{ quando } l \rightarrow \infty$$

**Demonstração:** A hipótese de  $\beta_i^{(l)} > \epsilon > 0$  assegura que a separação dos  $\tau_{jl}$  é uniforme para todo  $l$ , de forma que os limites  $\tau_j$  são distintos. Isto implica que cada  $\tau_j$  é um autovalor de  $A$ . Além disso, essa hipótese implica uma limitação inferior uniforme de  $|\eta_j^{(l)}|$  e então,  $\beta_d^{(l)} \rightarrow 0$ . Estas observações seguem do lema 5.6.3.

Seja  $\hat{p}_d^{(l)}(\lambda) = \prod_{i=1}^d (\lambda - \tau_{il})$  o polinômio característico de  $T_d^{(l)}$  e seja  $\hat{p}_d(\lambda) = \prod_{i=1}^d (\lambda - \tau_i)$  o polinômio limite de  $\hat{p}_d^{(l)}$ . Então:

$$\|\hat{p}_d^{(l)}(A)q_1^{(l)}\| = \prod_{j=1}^d \beta_d^{(l)} \rightarrow 0$$

e então

$$\hat{p}_d^{(l)}(\lambda_j)\gamma_j^{(l)} = v_j^* \hat{p}_d^{(l)}(A)q_1^{(l)} \rightarrow 0$$

Assim,

$$\hat{p}_d(\lambda_j) = 0 \text{ ou } \gamma_j^{(l)} \rightarrow 0$$

para todo  $j = 1, \dots, n$ . Isto significa que  $n - d$  dos coeficientes da expansão de  $\gamma_j^{(l)}$ , tendem para 0, quando  $l \rightarrow \infty$ . Além disso, como cada seqüência  $\{\tau_{jl}, l = 1, 2, \dots\}$  é convergente para  $j = 1, \dots, d$ , tem-se que os  $d$  coeficientes da expansão, correspondentes aos autovalores  $\tau_j$ , devem ser todos limitados inferiormente por 0, pela hipótese  $\beta_i^{(l)} > \epsilon > 0$  para todo  $j = 1, 2, \dots, d - 1$  e  $l$ . Suponha agora que  $\lambda_{j1} = \tau_1 < \lambda_1$ . Então o coeficiente da expansão

$$\gamma_{j1}^{(l)} = v_{j1}^* q_1^{(l)} = v_{j1}^* \frac{\Psi_l(A)q_1}{\|\Psi_l(A)q_1\|} = \frac{\Psi_l(\tau_1)v_{j1}^* q_1}{\|\sum_{i=1}^n \gamma_i \Psi_l(\lambda_i)v_i\|} = \frac{\gamma_{j1}\Psi_l(\tau_1)}{\sqrt{\sum_{i=1}^n \gamma_i^2 \Psi_l^2(\lambda_i)}},$$

onde  $v_{j1}$  representa o autovetor de  $A$  associado ao autovalor  $\tau_1$ , e  $q_1 = \sum_{i=1}^n \gamma_i v_i$  com  $v_i^* v_j = 0$  quando  $i \neq j$ .

Assim,

$$(\gamma_{j1}^{(l)})^2 = \frac{(\gamma_{j1}\Psi_l(\tau_1)/\Psi_l(\lambda_1))^2}{\gamma_1^2 + \sum_{i=2}^n \gamma_i^2 \Psi_l^2(\lambda_i)/\Psi_l^2(\lambda_1)} \leq \left(\frac{\gamma_{j1}\Psi_l(\tau_1)}{\gamma_1 \Psi_l(\lambda_1)}\right)^2,$$

onde os  $\gamma_j$  são os coeficientes da expansão de  $q_1$ . Agora, as raízes  $\mu_{il}$  dos polinômios filtros satisfazem,  $\lambda_n \leq \mu_{il} < \tau_{1l} < \lambda_1$ , de forma que:

$$\frac{\Psi_l(\tau_1)}{\Psi_l(\lambda_1)} = \prod_{k=1}^l \left( \prod_{i=1}^p \left( \frac{\tau_1 - \mu_{ik}}{\lambda_1 - \mu_{ik}} \right) \right) \leq \left( \frac{\tau_1 - \lambda_n}{\lambda_1 - \lambda_n} \right)^{lp} \rightarrow 0,$$

pois,  $(\tau_1 - \lambda_n)/(\lambda_1 - \lambda_n) < 1$ . O que é uma contradição. Concluímos então que  $\tau_1 = \lambda_1$ .

Um argumento similar pode ser feito para cada  $j$ , nos casos  $\tau_j < \lambda_j$ , e isso conclui a demonstração.  $\square$

## 5.7 Taxa de Convergência

Até este ponto, foi analisada a convergência do IRL, porém não foi dito nada a respeito da taxa de convergência. A seguir, são apresentados alguns resultados que descrevem condições sob os quais, obtém-se convergência rápida.

**Lema 5.7.1** *Seja  $A = B + \tau I$  uma matriz Hermitiana  $M \times M$ , onde  $B$  é Hermitiana semi-positiva definida,  $\text{posto}(B) = d$ , e  $\tau > 0$ . Se os  $d$  autovetores associados aos  $d$  maiores autovalores são distintos, e se  $q_0$  não é ortogonal a nenhum destes  $d$  autovetores, e ortogonal a no mínimo um autovetor associado a um autovalor não desejado, então  $\dim(\mathcal{K}^{d+1}(A, q_0)) = d + 1$ , ou seja, o método de Lanczos converge em no máximo  $d + 1$  passos.*

**Demonstração:** A prova deste lema decorre da propriedade:  $\mathcal{K}^m(A, q_0) = \mathcal{K}^m(A - \tau I, q_0) = \mathcal{K}^m(B, q_0)$  e  $\dim(\mathcal{K}^m(B, q_0) = d + 1$ , para todo  $m > d$  [133].  $\square$

### 5.7.1 Estimativas para $\sigma_d(H(\ell))$

Outro resultado importante em relação à convergência do método de Lanczos, aplicado ao cálculo dos  $d$  maiores valores singulares, é que a taxa de convergência depende da separação entre  $\tilde{\sigma}_d$  e  $\tilde{\sigma}_{d+1}$  (ver Teorema 6.4.1 em [65]). Quanto maior a separação, melhor é a taxa de convergência. Como nos problemas em estudo,  $\tilde{H}(\ell) = H(\ell) + E$  e  $\text{posto}(H) = d$ , da teoria de perturbação de valores singulares, segue que

$$\tilde{\sigma}_j \leq \|E\|, \quad j \geq d + 1.$$

Isto mostra que uma melhor separação entre  $\tilde{\sigma}_d$  e  $\tilde{\sigma}_{d+1}$ , ocorre quando  $\|E\| \ll \sigma_d$  [59]. Ou seja, o tamanho do  $\sigma_d$ , como uma função da dimensão do problema e das características do sinal, desempenha um papel importante na convergência do método. O seguinte teorema fornece uma estimativa para  $\sigma_d$ , no caso em que são utilizadas matrizes quadradas.

**Teorema 5.7.2** *Seja  $H$  a matriz de Hankel  $M \times M$ , descrita em (2.6) para  $\ell = 0$ . Defina  $\delta = \min_{l \neq k} |z_l - z_k|$ ,  $1 \leq l, k \leq d$ , onde  $z_l = e^{s_l \Delta t}$ , e  $\beta = \min |z_l|$ . Então um limitante inferior para  $\sigma_d(H)$  satisfaz*

$$\sigma_d(H) \geq \frac{(1 + \beta^2 + \beta^4 + \dots + \beta^{2(M-1)})}{\gamma d^2} \left[ 1 + \frac{D_M^2}{(d-1)\delta^2} \right]^{1-d} \quad (5.20)$$

onde  $\gamma = \min |r_l|$ , e  $D_M^2$  é dada no Teorema 4.1.5.

**Demonstração:** Usando a fatoração (2.7)

$$\sigma_d(H) \geq \sigma_d(R)\sigma_d(W_M)^2. \quad (5.21)$$

A demonstração termina aplicando nesta desigualdade a relação (4.11) (veja a prova do Teorema 4.8).

□

Como já foi visto anteriormente, no Lema 4.1.6,  $D_M^2 \approx 0$  quando  $M$  é suficientemente grande e  $|z_l| \approx 1$ . Além disso, em [10], é analisado o comportamento do limitante dado em (5.21), para matrizes de Hankel retangulares, e a conclusão apresentada é que a estimativa será melhor, quando a matriz de Hankel for quadrada.

O teorema 5.7.2 sugere ainda que  $\sigma_d$  deve crescer com a dimensão de  $H$ , e que a estimativa melhora quando os  $z_l$  ficam perto do círculo unitário no plano complexo mas não muito próximos uns dos outros. Portanto, deve-se utilizar matrizes de Hankel quadradas, com a maior dimensão possível, a fim de que  $\sigma_d$  possa vencer o tamanho da perturbação  $\|E\|$ , possibilitando assim, uma boa separação entre  $\tilde{\sigma}_d$  e  $\tilde{\sigma}_{d+1}$ .

Dos resultados obtidos do Lema 5.7.1 e Teorema 5.7.2, deduz-se que para obter convergência rápida do método de Lanczos, no cálculo dos  $d$  maiores valores singulares da matriz de Hankel  $\tilde{H}$  e correspondente subespaço invariante, deve-se ter  $\|E\| < \sigma_d$ , a dimensão da matriz de Hankel deve ser a maior possível, e o vetor inicial  $q_0$  deve ter fortes componentes no subespaço de interesse. Quando isto é satisfeito, pode-se esperar que a convergência seja rápida. Na seção seguinte é apresentado um exemplo numérico que mostra que se  $\|E\| \ll \sigma_d$ , então o método de Lanczos converge em pouco mais de  $d$  iterações, desde que vetor inicial seja escolhido adequadamente.

Finalizando esta seção, é apresentada uma escolha de vetor inicial para o método de Lanczos, a qual mostra resultados melhores em relação à velocidade de convergência, comparado com uma escolha aleatória. O vetor

$$r_0 = \tilde{H}(\ell)^* \tilde{\mathbf{h}}_{\ell-1}, \dots, \tilde{\mathbf{h}}_{\ell-1} = [\tilde{h}_{\ell-1}, \tilde{h}_\ell, \dots, \tilde{h}_{\ell+M-2}]^*,$$

é proposto como vetor inicial. Aqui a matriz é denotada simplesmente por  $\tilde{H}$ . Para justificar esta escolha, considere a decomposição em valores singulares (SVD) de  $\tilde{H}$ :

$$\tilde{H} = \tilde{U}\tilde{\Sigma}\tilde{V}^* = \tilde{\sigma}_1\tilde{u}_1\tilde{v}_1^* + \tilde{\sigma}_2\tilde{u}_2\tilde{v}_2^* + \dots + \tilde{\sigma}_N\tilde{u}_N\tilde{v}_N^*.$$

Usando esta decomposição, segue imediatamente que

$$\tilde{v}_j^* \tilde{H}^* \tilde{\mathbf{h}}_{\ell-1} = \tilde{\sigma}_j \tilde{u}_j^* \tilde{\mathbf{h}}_{\ell-1}, \quad j = 1, 2, \dots, N. \quad (5.22)$$

Este resultado mostra que, se  $\|E\| \ll \sigma_d$  e  $\|\tilde{\mathbf{h}}_{\ell-1} - \mathbf{h}_{\ell-1}\| \ll \|\mathbf{h}_{\ell-1}\|$ , então as componentes mais fortes do vetor  $q_0$  estão associadas com os  $d$  primeiros vetores singulares, os quais deseja-se aproximar.

## 5.8 Resultados Numéricos Sobre IRL

Para observar o desempenho do IRL em relação à escolha do vetor inicial  $q_0$ , à escolha de  $p$  e do uso da multiplicação rápida, são apresentados resultados de alguns experimentos numéricos realizados com um sinal proveniente de simulações em Ressonância Magnética Nuclear, [118]. Para contar o número de operações de ponto flutuante, foi usada a função *flops* do Matlab. O sinal foi construído pela seguinte modelagem:

$$\tilde{s}_j = \sum_{k=1}^d r_k e^{i\phi_k} e^{(\alpha_k + i\omega_k)j\Delta t} + \epsilon_j, \quad j = 1, 2, \dots, 512,$$

onde os parâmetros  $r_l$ ,  $\alpha_l$ ,  $\omega_l$  e  $\phi_l$  representam a amplitude real, o fator de decaimento, a frequência angular e a fase, respectivamente. Os  $\epsilon_j$  correspondem ao ruído, o qual, em nossos testes foi gerado usando o comando *randn* do Matlab com diferentes valores de desvio padrão ( $\sigma = 5, 10, 15$ ). O sinal mostrado possui 11 exponenciais, e os parâmetros  $r_l$ ,  $\alpha_l$ ,  $\omega_l$  e  $\phi_l$  são dados na tabela 5.1.

$l$	1	2	3	4	5	6	7	8	9	10	11
$r_l$	75	150	75	150	150	150	150	150	1400	60	500
$\alpha_l$	-50	-50	-50	-50	-50	-50	-50	-25	-286	-25	-200
$\omega_l \setminus (2\pi)$	-86	-70	-54	152	168	292	308	360	440	490	530
$\phi_l$	135	135	135	135	135	135	135	135	135	135	135

Tabela 5.1: Parâmetros do sinal, onde  $r_l$  é dada em unidades arbitrárias,  $\alpha_l$  e  $\omega_l$  em Hz e  $\phi_l$  em graus.

A partir do sinal, foram geradas matrizes de Hankel de ordem  $256 \times 256$ , de acordo com (2.6). Na figura 5.4, mostramos os 15 maiores valores singulares de  $\tilde{H}$ , para dois níveis de ruído correspondentes a  $\sigma = 5$  e  $\sigma = 15$ , bem como os 11 valores singulares não-nulos da matriz de Hankel  $H$ . Observe que para  $\sigma = 5$ , existe um gap bem notório entre  $\tilde{\sigma}_{11}$  e  $\tilde{\sigma}_{12}$ . Quando o nível de ruído é aumentado para  $\sigma = 15$ , este gap diminui, tornando o problema de calcular os 11 maiores valores e correspondentes vetores singulares mais complicado. Além disso, se o ruído introduzido for muito alto, ou seja se  $\|E\| > \sigma_{11}$ , este último valor singular pode não ser recuperado.

Para observar o desempenho do método em relação à escolha do vetor inicial, foram realizados testes numéricos com  $q_0$  aleatório e com  $q_0 = \tilde{H}^* \tilde{b}$ , para vários valores de  $p$ . Os resultados dos experimentos são mostrados na Tabela 5.2, onde  $\rho_1$  denota o quociente

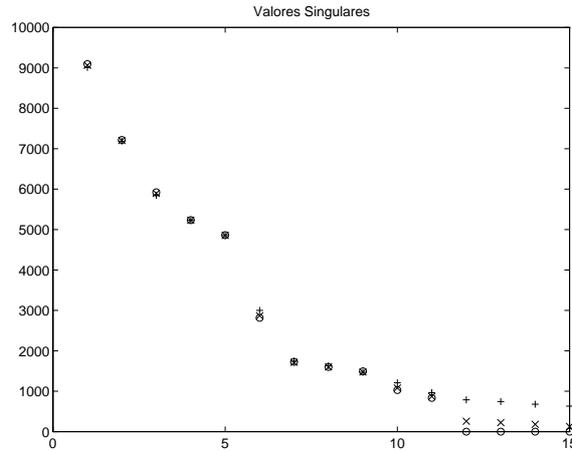


Figura 5.4: Em 'o' os 11 valores singulares de  $H$ , em 'x' e '+' os 15 maiores valores singulares de  $\tilde{H}$  para  $\sigma = 5$  e  $\sigma = 15$ , respectivamente.

$\sigma$	$p$	Recomeços $r_0$ aleatório	Recomeços $r_0 = \tilde{H}^* \tilde{b}$	$\rho_1$	$\rho_2$
5	5	3	0	3.2476	4.5251
10	7	5	0	5.6751	4.2278
15	10	8	8	1.0057	2.2862
15	11	7	0	9.7371	3.6875

Tabela 5.2: Comparação do número de operações de ponto flutuante para o IRL-ES,  $p$  denota o Número de passos Lanczos adicionais.

entre o número de flops gastos quando o método é inicializado com  $q_0$  aleatório e o número de flops gastos quando  $q_0 = \tilde{H}^* \tilde{b}$ . Analisando os resultados numéricos dessa tabela, concluímos que se o nível de ruído é alto, o valor de  $p$  deve ser tomado um pouco maior do que  $d$ , caso contrário, o valor de  $p$ , deve ser em geral, menor do que  $d$ . Além disso, verificamos que a convergência é mais rápida quando o vetor inicial é escolhido como  $q_0 = \tilde{H}^* \tilde{b}$ . A tabela apresenta também, na terceira coluna, a razão entre o número de flops do IRL-ES sem o uso da multiplicação rápida<sup>1</sup>, e com o uso de multiplicação rápida. Foram realizados os mesmos testes para outros sinais, e os resultados obtidos foram similares.

Numerosos resultados numéricos sobre recuperação de harmônicos são encontrados em Cardoso [31]. O problema de detectar a ordem do modelo exponencial foi recentemente abordado com sucesso por Bazán em [13].

<sup>1</sup>Multiplicação rápida matriz-vetor para matrizes de Hankel é feita utilizando-se a potencialidade do algoritmo da transformada rápida de Fourier, detalhes são encontrados em [8].

## 5.9 Métodos para Autovalores de Polinômios Matriciais

Conforme comentado no Capítulo 1, uma maneira de calcular os autovalores de um polinômio matricial

$$P_m(\lambda) = A_m\lambda^m + A_{m-1}\lambda^{m-1} + \cdots + A_1\lambda + A_0,$$

é através de um problema de autovalores generalizados do tipo

$$Ax = \lambda Bx,$$

com  $A$  e  $B$  com em (1.10), isto é, através de uma linearização do tipo

$$A - \lambda B = \begin{bmatrix} 0 & I_q & 0 & \cdots & 0 \\ 0 & 0 & I_q & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & I_q & 0 \\ -A_0 & -A_1 & \cdots & -A_{m-1} & 0 \end{bmatrix} - \lambda \begin{bmatrix} I_q & 0 & 0 & \cdots & 0 \\ 0 & I_q & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & I_q & 0 \\ 0 & 0 & 0 & \cdots & A_m \end{bmatrix}. \quad (5.23)$$

Se  $B$  é não singular, a técnica recomendada é transformar o problema de autovalor num problema de autovalor matricial padrão, conforme descrito no Capítulo 1, Seção 1.2, a qual funciona bem, exceto nos casos em que  $B$  é mal-condicionada. Se  $B$  é mal-condicionada, por razões de estabilidade, recomenda-se o uso da decomposição de Schur generalizada, calculada pelo método QZ. O método QZ é, em certo sentido, uma generalização do método QR, detalhes do método podem ser encontrados em [59, Cap. 7, pag. 382-386].

Em certos casos, o objetivo é calcular apenas alguns autovalores. Isto pode ser feito a partir de uma fatoração generalizada de Schur parcial, calculada através do método de Jacobi-Davidson. Detalhes podem ser encontrados em [4, Cap. 8, Seção 8.1].

A principal desvantagem dos métodos que utilizam a linearização acima para resolver o problema de autovalor polinomial matricial, é que a ordem do problema resultante torna-se grande ( $q \times m$ ). Essas dificuldades podem ser contornadas através de técnicas baseadas no método de Newton, detalhes podem ser encontradas em [112, 39]. Finalmente, uma grande variedade de métodos para o problema de autovalor polinomial matricial quadrático, utilizando diferentes linearizações, pode ser encontrada em [122].



# Apêndice A

## Conceitos Básicos da Álgebra Linear

### A.1 Notação

Em quase todo o texto, usam-se letras minúsculas para vetores, letras gregas minúsculas para escalares, e letras maiúsculas e gregas maiúsculas para matrizes. A matriz nula, o vetor nulo e o escalar zero são denotados por  $0$ . A matriz identidade é escrita como  $I$ , ou  $I_n$  se for necessário especificar a sua ordem. A transposta da matriz  $A$ , é denotada por  $A^T$ , sua conjugada transposta por  $A^*$  (alguns vezes denotada também como  $A^H$ ), isto é,  $A^* = \bar{A}^T$ . A inversa é denotada por  $A^{-1}$ , a inversa da transposta por  $A^{-T}$ , e a inversa da conjugada transposta por  $A^{-*}$ .

O espaço coluna de  $A \in \mathbb{C}^{m \times n}$ , denotado por  $\mathcal{R}(A)$ , é definido por

$$\mathcal{R}(A) = \{Ax : x \in \mathbb{C}^n\}$$

e o espaço nulo de  $A$ ,  $\mathcal{N}(A)$ , por

$$\mathcal{N}(A) = \{x : Ax = 0\}.$$

O posto de uma matriz  $A$ ,  $\text{rank}(A)$ , é definido por

$$\text{rank}(A) = \dim(\mathcal{R}(A)),$$

em que  $\dim(\mathcal{X})$  denota a dimensão do espaço  $\mathcal{X}$ . O determinante de  $A$ , é escrito como  $\det(A)$ , e o traço como  $\text{tr}(A)$ .

A norma-2 de um vetor  $x \in \mathbb{C}^n$ , também chamada de norma Euclidiana, é escrita como  $\|x\|_2$ , e definida como a raiz quadrada positiva de

$$\sum_{i=1}^n |x_i|^2 = x^*x.$$

A norma Frobenius de uma matriz  $A \in \mathbb{C}^{m \times n}$ ,  $\|A\|_F$  é definida por

$$\|A\|_F = \left( \sum_{j=1}^n \sum_{i=1}^m |a_{ij}|^2 \right)^{1/2},$$

e a norma-2 de  $A$  é definida por

$$\|A\|_2 = \sup_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}.$$

Algumas vezes a norma 2 será denotada somente por  $\|\cdot\|$ .

O subespaço gerado por um conjunto de vetores  $\{x_1, x_2, \dots, x_n\}$  é denotado por  $\text{span}\{x_1, x_2, \dots, x_n\}$ .

A seguir, são definidos alguns tipos especiais de matrizes:

**Definição A.1.1** - Seja  $A \in \mathbb{C}^{m \times n}$

- Simétrica (Hermitiana), se  $A^T = A$  ( $A^* = A$ ).
- Positiva definida (semi-positiva definida), se é Hermitiana e se  $x^*Ax > (\geq) 0$  para todo  $x \neq 0$ .
- Unitária, ou Ortogonal no caso real, se  $A^*A = AA^* = I$ .
- Normal, se  $A^*A = AA^*$ .
- Triangular Superior, se é quadrada e se  $i > j \Rightarrow a_{ij} = 0$ , isto é, se os elementos abaixo da diagonal principal são iguais a zero.
- Triangular Inferior, se é quadrada e se  $i < j \Rightarrow a_{ij} = 0$ , isto é, se os elementos acima da diagonal principal são iguais a zero.
- Hessenberg Superior, se  $a_{ij} = 0$  para qualquer par  $i, j$  tal que  $i > j + 1$ .

## A.2 Autovalores e Autovetores

Informalmente falando, um autovetor é um vetor que não muda a sua direção, quando multiplicado por  $A$ , e seu autovalor é o tamanho da sua expansão ou contração neste processo.

**Definição A.2.1** O par  $(x, \lambda)$  é chamado um *auto-par* da matriz  $A \in \mathbb{C}^{n \times n}$ , se  $x \neq 0$  e

$$Ax = \lambda x. \quad (\text{A.1})$$

O vetor  $x$  é chamado *autovetor* à direita de  $A$ , e  $\lambda$  o *autovalor* associado. Se  $y^*A = \lambda y^*$ ,  $y$  é chamado autovetor à esquerda de  $A$ , associado ao autovalor  $\lambda$ . O conjunto de todos os autovalores de  $A$  é denotado por  $\lambda(A)$ . A equação (A.1) pode ser reescrita como:

$$(A - \lambda I)x = 0, \quad x \neq 0$$

o que ocorre se e somente se  $(A - \lambda I)$  é singular, ou equivalentemente, se e somente se:

$$\det(A - \lambda I) = 0 \quad (\text{A.2})$$

A expressão (A.2) é chamada *Polinômio Característico* de  $A$ , e possui grau  $n$  em  $\lambda$ , suas raízes são os autovalores de  $A$ . Conseqüentemente, a matriz  $A$  possui exatamente  $n$  autovalores.

Diz-se que um autovalor  $\lambda$  possui multiplicidade algébrica  $\mu$ , se ele é uma raiz com multiplicidade  $\mu$  do polinômio característico. Se a multiplicidade algébrica de um autovalor for 1, ele é chamado *simples*, caso contrário é chamado *múltiplo*. Um autovalor  $\lambda$  possui multiplicidade geométrica  $\gamma$ , se o número máximo de autovetores associados a  $\lambda$  é  $\gamma$ , ou seja, a dimensão de  $\mathcal{N}(A - \lambda I)$  é  $\gamma$ .

A seguir, são apresentadas algumas propriedades dos autovalores e autovetores de  $A \in \mathbb{C}^{n \times n}$ :

- P1.  $A$  e  $A^T$  tem os mesmos autovalores.
- P2. Se  $\lambda$  é autovalor de  $A$  com autovetor  $v$  e  $A$  é não singular, então  $\lambda^{-1}$  é autovalor de  $A^{-1}$  com autovetor  $v$ .
- P4. A soma (produto) dos autovalores é igual ao traço (determinante) da matriz.
- P6. Sejam  $v_1, \dots, v_k$  autovetores de  $A$  associados a distintos autovalores  $\lambda_1, \dots, \lambda_k$ . Então  $v_1, \dots, v_k$  são *LI*.
- P7. Se  $A$  tem  $n$  autovalores distintos, então  $A$  pode ser diagonalizada, ou seja, existe uma matriz  $X$  não singular e uma matriz diagonal  $D$ , tais que  $A = X^{-1}DX$ .
- P8. Os autovalores de  $A^k$  são  $\lambda_1^k, \dots, \lambda_n^k$ , a  $k$ -ésima potência dos autovalores de  $A$ .
- P9. Seja  $\rho \in \mathbb{C}$ , se  $v$  é um autovetor de  $A$  com autovalor  $\lambda$ , então  $v$  é também um autovetor de  $A - \rho I$  com autovalor  $\lambda - \rho$ .

P10. O maior autovalor de uma matriz positiva definida é real e positivo, logo os componentes do respectivo autovetor também o são.

P11. Os autovalores de uma matriz simétrica ou de uma matriz Hermitiana são reais.

P12. Os autovetores de uma matriz simétrica real ou de uma matriz Hermitiana, correspondentes a autovalores distintos são ortogonais.

P13. Uma matriz simétrica real pode ser fatorada em  $A = Q\Lambda Q^T$ , com autovalores em  $\Lambda$  e respectivos autovetores ortogonais em  $Q$ .

P14. Se  $A$  é real, então seu polinômio característico é real, e os seus autovalores complexos devem ocorrer em pares conjugados complexos.

P15. Os autovalores de uma matriz triangular são os elementos da sua diagonal.

### A.2.1 Condicionamento dos Autovalores

Nesta subseção são apresentados alguns resultados clássicos da teoria de perturbação de autovalores, os quais fornecem medidas da sensibilidade de qualquer problema de autovalor.

**Definição A.2.2** Seja  $A \in \mathbb{C}^{n \times n}$  e seja  $A = X\Lambda X^{-1}$  uma decomposição espectral de  $A$ . Os números de condição de Jordan do problema de autovalor associado a  $A$ , são definidos por

$$\kappa_2 = \min_{D \in \mathcal{D}} \{\|XD\|_2 \|D^{-1}X^{-1}\|_2\}, \quad \kappa_F = \min_{D \in \mathcal{D}} \{\|XD\|_F \|D^{-1}X^{-1}\|_F\}, \quad (\text{A.3})$$

onde  $\mathcal{D}$  é o conjunto de todas as matrizes diagonais inversíveis.

Define-se também o número de condição associado ao autovalor  $\lambda_j$ ,  $|s_j|^{-1}$ , por

$$s_j = \frac{u_j^* v_j}{\|u_j\|_2 \cdot \|v_j\|_2}, \quad (\text{A.4})$$

onde  $u_j$  e  $v_j$  são autovetores à direita e à esquerda de  $A$ , associados ao autovalor  $\lambda_j$ .

**Observação A.2.3** Se  $\lambda_j$  é um autovalor simples, então  $x$  e  $y$  são únicos e  $|s_j|$  é unicamente determinado. O número  $|s_j|^{-1}$  fornece uma medida da sensibilidade do autovalor  $\lambda_j$ , a pequenas perturbações nas entradas da matriz  $A$ , [131].

O teorema A.2.4 fornece uma outra forma de calcular  $\kappa_F$ , quando os autovalores de  $A$  são simples.

**Teorema A.2.4** *Se todos os autovalores de  $A$  são simples, então*

$$\kappa_F = |s_1|^{-1} + |s_2|^{-1} + \cdots + |s_n|^{-1}. \quad (\text{A.5})$$

**Demonstração:** A demonstração é encontrada em [104].

O próximo teorema fornece algumas relações envolvendo o número de condição  $|s_j|^{-1}$ .

**Teorema A.2.5** *Suponha que todos os autovalores de  $A$  são simples. Então*

$$1 \leq |s_j|^{-1} \leq \left[ 1 + \frac{D^2(A)}{(n-1)\delta_j^2} \right]^{\frac{n-1}{2}}, \quad 1 \leq j \leq n, \quad (\text{A.6})$$

onde  $\delta_j = \min_{\substack{1 \leq k \leq n \\ k \neq j}} |\lambda_j - \lambda_k|$  e  $D(A)^2 = \|A\|_F^2 - \sum_{j=1}^n |\lambda_j|^2$ . Além disso o número de condição

$\kappa_F(X) = \|X\|_F \|X^{-1}\|_F$ , satisfaz

$$\sum_{j=1}^n \frac{1}{|s_j|} \leq \kappa_F(X) \leq n \cdot \max_j \frac{1}{|s_j|} \cdot \frac{\max_j \|e_j^* X^{-1}\|_2}{\min_j \|e_j^* X^{-1}\|_2}. \quad (\text{A.7})$$

**Demonstração:** A desigualdade (A.6) e a desigualdade à esquerda de (A.7) seguem do Teorema 5 e do Teorema 3, de Smith [104], respectivamente. Para provar a desigualdade à direita de (A.7), defina os vetores  $u_j^* = e_j^* X^{-1}$  e  $v_j = X e_j$ , os quais são autovetores à esquerda e à direita de  $A$ , associados ao autovalor  $\lambda_j$ , usando a definição de  $\kappa_F(X)$  e (A.4), segue que

$$\kappa_F^2(X) = \sum_{j=1}^n \|X e_j\|_2^2 \sum_{j=1}^n \|e_j^* X^{-1}\|_2^2 = \sum_{j=1}^n \frac{1}{|s_j|^2} \cdot \frac{1}{\|e_j^* X^{-1}\|_2^2} \sum_{j=1}^n \|e_j^* X^{-1}\|_2^2.$$

Basta agora, limitar os termos na soma.

□

O número  $D(A)^2$ , chamado *desvio da normalidade* de  $A$  (departure from normality), fornece uma medida de quão perto uma matriz está de ser uma matriz normal. Se  $A$  for uma matriz normal,  $D(A)^2 = 0$ . Além disto, os autovalores de uma matriz normal são perfeitamente bem condicionados.

O próximo teorema apresenta uma estimativa do erro entre os autovalores da matriz  $A$ , e os autovalores da matriz perturbada  $A + E$ , ou seja, apresenta uma estimativa do comportamento dos autovalores em função da perturbação nas entradas da matriz.

**Teorema A.2.6** *Se  $\tilde{\lambda}$  é um autovalor de  $A + E \in \mathbb{C}^{n \times n}$  e*

$$X^{-1}AX = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n),$$

então

$$\min_{\lambda \in \lambda(A)} |\tilde{\lambda} - \lambda| \leq \kappa(X) \|E\|_2.$$

**Demonstração:** A demonstração pode ser encontrada em [59].

### A.3 Decomposição em Valores Singulares (SVD)

Nesta seção é apresentada uma ferramenta da Álgebra Linear Computacional muito poderosa, que facilita o trabalho com problemas de posto incompleto, e que é de fundamental importância no desenvolvimento deste trabalho.

**Teorema A.3.1** ( Decomposição em Valores Singulares (SVD)) *Seja  $A \in \mathbb{C}^{m \times n}$ , uma matriz com posto  $d$ . Então existem matrizes ortogonais  $U \in \mathbb{C}^{m \times m}$  e  $V \in \mathbb{C}^{n \times n}$ , e uma matriz diagonal  $\Sigma \in \mathbb{C}^{m \times n}$ ; tais que:*

$$A = U\Sigma V^T \tag{A.8}$$

onde a matriz  $\Sigma$  é da forma:

$$\Sigma = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix}$$

com  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_d > \sigma_{d+1} = \dots = \sigma_n = 0$ .

**Demonstração:** Veja [59].

Os números  $\sigma_i$  são chamados valores singulares de  $A$ , enquanto os vetores  $u_i$  e  $v_i$  são chamados vetores singulares a esquerda e a direita de  $A$ , respectivamente.

Geometricamente falando, a SVD de  $A$  mostra que existem duas bases ortogonais, a saber, as colunas de  $U$  e  $V$ , tais que a aplicação  $A$ , de  $\mathbb{C}^n$  em  $\mathbb{C}^m$ , é representada por uma matriz diagonal  $\Sigma$ , cujas entradas são números reais.

A partir das relações  $A^*A = V\Sigma^2V^*$  e  $AA^* = U\Sigma^2U$ , verifica-se que a SVD está fortemente relacionada com as decomposições espectral das matrizes hermitianas  $A^*A$  e  $AA^*$ .

Se  $\text{rank}(A) = d$ , a equação (A.8) pode ser rescrita como:

$$A = [U_1 \ U_2] \Sigma \begin{bmatrix} V_1^* \\ V_2^* \end{bmatrix} = U_1 \Sigma_1 V_1^*, \quad (\text{A.9})$$

onde  $U_1$  e  $V_1$  são as  $d$  primeiras colunas das matrizes  $U$  e  $V$  respectivamente,  $U_2$  e  $V_2$  são as  $m - d$  e  $n - d$  últimas colunas das matrizes  $U$  e  $V$  respectivamente, e  $\Sigma_1 = \text{diag}(\sigma_1, \dots, \sigma_d)$ . A última decomposição é chamada *SVD reduzida*. Assim, a matriz  $A$  pode ser escrita como um produto de matrizes de posto  $d$ , ou como a soma de  $d$  matrizes de posto 1.

Tem-se ainda que:

$$\begin{aligned} \mathcal{N}(A) &= \text{span}\{v_{d+1}, \dots, v_n\}, \\ \mathcal{R}(A) &= \text{span}\{u_1, \dots, u_d\}, \\ \mathcal{N}(A^*) &= \text{span}\{u_{d+1}, \dots, u_m\}, \\ \mathcal{R}(A^*) &= \text{span}\{v_1, \dots, v_d\}. \end{aligned}$$

A SVD também fornece informação sobre a norma-2 e a norma Frobenius de  $A$

$$\begin{aligned} \|A\|_F^2 &= \sigma_1^2 + \dots + \sigma_p^2, \quad p = \min\{m, n\}, \\ \|A\|_2 &= \sigma_1, \\ \min_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} &= \sigma_n; \quad (m \geq n). \end{aligned}$$

Ainda mais, definindo-se  $\mathbf{A}_k$  como:  $\mathbf{A}_k = \sum_{i=1}^k \sigma_i u_i v_i^T$ ,  $k \leq p$ , demonstra-se que (veja, por exemplo, [59, Teor. 2.5.3]):

$$\min_{\text{rank}(B)=k} \|A - B\|_2 = \|A - \mathbf{A}_k\|_2 = \sigma_{k+1}, \quad (\text{A.10})$$

ou seja,  $A_k$  é a matriz de posto  $k$  mais próxima de  $A$ .

Abaixo, são listadas as principais projeções ortogonais associadas com a SVD. Suponha que  $A$  satisfaz as condições do teorema A.3.1, e as matrizes  $U$  e  $V$  são particionadas como em (A.9), então:

$$V_1 V_1^* = \text{projeção em } \mathcal{N}(A)^\perp = \mathcal{R}(A^*)$$

$$V_2 V_2^* = \text{projeção em } \mathcal{N}(A)$$

$$U_1 U_1^* = \text{projeção em } \mathcal{R}(A)$$

$$U_2 U_2^* = \text{projeção em } \mathcal{R}(A)^\perp = \mathcal{N}(A^*)$$

### A.3.1 Sensibilidade dos Valores Singulares

Suponha que  $A \in \mathbb{C}^{m \times n}$  possui posto  $d$ , nesta seção é feito um breve estudo dos efeitos de uma perturbação em  $A$ , nos seus valores e vetores singulares. O teorema a seguir, é um resultado clássico da teoria de perturbação de valores singulares, e afirma que os valores singulares são bem condicionados em relação a perturbações:

**Teorema A.3.2** *Seja  $A$  e  $\tilde{A} = A + E \in \mathbb{C}^{m \times n}$ ,  $m \geq n$  com valores singulares  $\sigma_i$  tal que*

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n, \quad \text{e} \quad \tilde{\sigma}_1 \geq \tilde{\sigma}_2 \geq \cdots \geq \tilde{\sigma}_n,$$

*respectivamente. Então:*

$$\begin{aligned} |\sigma_i - \tilde{\sigma}_i| &\leq \|E\|_2 \\ \sum_{i=1}^n |\sigma_i - \tilde{\sigma}_i| &\leq \|E\|_F^2 \end{aligned}$$

Deste teorema decorre que os  $n - d$  valores singulares da matriz  $\tilde{A}$  são da ordem de  $\|E\|_2$ , uma vez que  $\sigma_{d+1} = \cdots = \sigma_n = 0$ . E desta forma a matriz  $\tilde{A}$  se comporta como sendo de posto completo.

Quanto aos efeitos da perturbação nos vetores singulares, eles também são estimados na ordem de  $\|E\|_2$ , veja [105].

## A.4 Pseudo Inversa

Dada uma matriz  $A \in \mathbb{C}^{m \times n}$ , a pseudo inversa de  $A$ , também conhecida como inversa de Moore-Penrose, é definida através das seguintes condições:

$$\begin{aligned} (1) \quad AA^\dagger A &= A & (2) \quad A^\dagger AA^\dagger &= A^\dagger \\ (3) \quad (AA^\dagger)^* &= AA^\dagger & (4) \quad (A^\dagger A)^* &= A^\dagger A \end{aligned}$$

Como  $AA^\dagger = U_1 U_1^*$ , tem-se que  $AA^\dagger$  é a projeção ortogonal sobre  $\mathcal{R}(A)$ , e  $A^\dagger A = V_1 V_1^*$  é a projeção ortogonal sobre  $\mathcal{R}(A^*)$ . Em particular, a pseudo inversa pode ser definida pela SVD truncada  $d$  de  $A$ :

$$A^\dagger = V_1 \Sigma_1^{-1} U_1^* \tag{A.11}$$

em que  $U_1$ ,  $V_1$ , e  $\Sigma_1$  são dadas como em (A.9). Observe que se  $m = n$ , e se  $\text{posto}(A) = m$ , então  $A^\dagger = A^{-1}$ .

Define-se também o *número de condição* de  $A$ , por

$$\kappa = \|A\|_2 \|A^\dagger\|_2, \tag{A.12}$$

o qual fornece uma medida da sensibilidade do problema  $Ax = b$ , a perturbações em  $A$ .

## A.5 Problemas de Quadrados Mínimos Lineares

O Problema de Quadrados Mínimos Lineares (PQML) consiste em, dados um vetor  $b \in \mathbb{C}^m$  e uma matriz  $A \in \mathbb{C}^{m \times n}$ ,  $m > n$ , encontrar um vetor  $x \in \mathbb{C}^n$ , tal que  $Ax$  é a “melhor aproximação” para  $b$ , isto é, resolver:

$$\min_{x \in \mathbb{C}^n} \|Ax - b\|_2 \quad (\text{A.13})$$

O vetor  $x$  é chamado solução de quadrados mínimos, e o vetor  $r = b - Ax$ , é o vetor de resíduos.

O próximo resultado caracteriza o conjunto de todas as soluções do PQML (A.13).

**Teorema A.5.1** *Seja  $\mathcal{S} = \{x \in \mathbb{C}^n / \|Ax - b\|_2 = \min\}$  o conjunto de todas as soluções de (A.13). Então,  $x \in \mathcal{S}$  se e somente se ocorre a seguinte condição de ortogonalidade*

$$A^*(b - Ax) = 0 \quad (\text{A.14})$$

**Demonstração:** A demonstração deste teorema pode ser encontrada em [26, pag. 5].

O Teorema acima afirma que o vetor de resíduos de uma solução de quadrados mínimos,  $r = b - Ax$ , está em  $\mathcal{N}(A^*)$ . Consequentemente, qualquer solução do PQML decompõe, de maneira única o vetor  $b$  em duas componentes ortogonais

$$b = Ax + r, \quad Ax \in \mathcal{R}(A), \quad r \in \mathcal{N}(A^*).$$

Esta propriedade é ilustrada geometricamente na figura A.1.

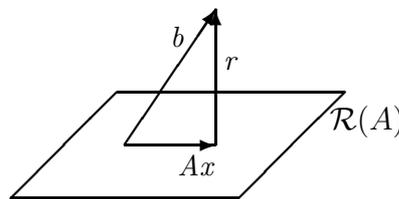


Figura A.1: Interpretação Geométrica do PQML

Da equação (A.14) segue que a solução de quadrados mínimos satisfaz as equações normais

$$A^*Ax = A^*b. \quad (\text{A.15})$$

A matriz  $A^*A \in \mathbb{C}^{n \times n}$  é Hermitiana e semi definida positiva, sendo definida positiva se e somente se  $\text{rank}(A) = n$ . Assim, se  $\text{rank}(A) = n$ , então a única solução de quadrados mínimos  $x$ , e o correspondente vetor de resíduos  $r$ , são dados por;

$$x = (A^*A)^{-1}A^*b, \quad r = b - A(A^*A)^{-1}A^*b.$$

Se  $\text{rank}(A) < n$ , então  $A$  possui um espaço nulo não trivial, e a solução de quadrados mínimos não é única, porém, entre todas as soluções existe uma única solução que minimiza  $\|x\|_2$ , veja o próximo teorema.

A SVD é uma poderosa ferramenta para resolver PQML, pois as matrizes ortogonais que transformam  $A$  em uma matriz diagonal (A.8) não alteram a norma-2 dos vetores. É apresentado a seguir um resultado que se aplica tanto a sistemas sobre-determinados como a sub-determinados.

**Teorema A.5.2** *Considere o problema de quadrados mínimos lineares geral*

$$\min_{x \in \mathcal{S}} \|x\|_2, \quad \mathcal{S} = \{x \in \mathbb{C}^n / \|b - Ax\|_2 = \min\},$$

em que  $A \in \mathbb{C}^{m \times n}$  e  $\text{rank}(A) = d \leq \min\{m, n\}$ . Então o problema sempre possui solução única, a qual pode ser escrita em termos da SVD de  $A$  como

$$x^\dagger = A^\dagger b = V \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^* b, \quad (\text{A.16})$$

onde  $\Sigma^{-1} = \text{diag}(\sigma_1^{-1}, \sigma_2^{-1}, \dots, \sigma_d^{-1})$ .

**Demonstração:** A demonstração pode ser encontrada em [26].

□

### A.5.1 Problemas de Quadrados Mínimos Totais

Em muitas aplicações existem problemas onde a matriz  $A$  não é precisamente conhecida. Por exemplo,  $A$  pode ser disponível apenas por medidas, ou pode ser uma aproximação de um operador. Assim, é necessário desenvolver métodos que levem em conta os erros tanto em  $A$  como em  $b$ . Um destes métodos é o de Quadrados Mínimos Totais (TLS).

**Definição A.5.3** Considere um sistema de equações sobredeterminado, tal que  $Ax \approx b$ . Dizemos que  $\hat{x}$  é uma solução do problema de quadrados mínimos totais, se  $\hat{x}$  é solução do sistema consistente

$$\hat{A}x = \hat{b},$$

em que  $\hat{A}$  e  $\hat{b}$  resolvem o problema

$$\min_{(\hat{A}; \hat{b})} \|(A; b) - (\hat{A}; \hat{b})\|_F \quad (\text{A.17})$$

sujeito a

$$\hat{b} \in \mathcal{R}(\hat{A}).$$

A solução de um problema TLS pode ser obtida pela SVD. De fato, seja

$$(A; b) = \hat{U} \hat{\Sigma} \hat{V}^*, \quad \hat{\Sigma} = \text{diag}(\hat{\sigma}_1, \dots, \hat{\sigma}_{n+1}),$$

onde  $\hat{U}^* \hat{U} = I_m$ ,  $\hat{V}^* \hat{V} = I_{n+1}$  e  $\hat{\sigma}_1 \geq \dots \geq \hat{\sigma}_{n+1} \geq 0$ .

Pelo Teorema 1.2.9 em [26], os valores singulares de  $A$ ,

$$\sigma_1 \geq \dots \geq \sigma_n \geq 0$$

entrelaçam os valores singulares de  $(A; b)$ , i.e.

$$\hat{\sigma}_1 \geq \sigma_1 \geq \hat{\sigma}_2 \geq \dots \geq \hat{\sigma}_n \geq \sigma_n \geq \hat{\sigma}_{n+1}.$$

Assuma que  $\text{posto}(A) = n$ , ou seja que  $\sigma_n > 0$ . Se  $\hat{\sigma}_{n+1} = 0$ , então segue que  $b \in \mathcal{R}(A)$ . Neste caso, o sistema original,  $Ax = b$ , é compatível, e pode-se tomar  $(E; r) = 0$ . Se  $\hat{\sigma}_{n+1} > 0$ , então  $b \notin \mathcal{R}(A)$  e do teorema de Eckart-Young-Mirsky, [70, pag ], segue que

$$\min_{\text{posto}(A+E, b+r) < n+1} \|(E, r)\|_F = \hat{\sigma}_{n+1}.$$

Se ocorrer

$$\hat{\sigma}_k > \hat{\sigma}_{k+1} = \dots = \hat{\sigma}_{n+1}, \quad k \leq n,$$

então o mínimo é atingido para qualquer perturbação de posto 1 da forma

$$(E; r) = -(A; b) \hat{v} \hat{v}^*, \quad \hat{v} \in \mathcal{S} = \text{span}[\hat{v}_{k+1}, \dots, \hat{v}_{n+1}],$$

onde  $\hat{v}_{k+1}, \dots, \hat{v}_{n+1}$  são os vetores singulares a direita, correspondentes a  $\hat{\sigma}_{k+1}, \dots, \hat{\sigma}_{n+1}$ .

Se  $e_{n+1} = (0, \dots, 0, 1)^*$  é ortogonal a  $\mathcal{S}$ , então o problema TLS não possui solução. Assuma que é possível encontrar um vetor  $\hat{v} \in \mathcal{S}$ , cuja  $(n+1)$ -ésima componente,  $\gamma$  é não nula. Então, com

$$\hat{v} = \begin{pmatrix} z \\ \gamma \end{pmatrix} = -\gamma \begin{pmatrix} x \\ -1 \end{pmatrix}, \quad x = -\gamma^{-1} z \quad (\text{A.18})$$

tem-se

$$-\gamma(A + E; b + r) \begin{pmatrix} x \\ -1 \end{pmatrix} = (A + E, b + r)\hat{v} = (A; b)(I - \hat{v}\hat{v}^*)\hat{v} = 0.$$

Portanto  $(A + E)x = b + r$ , o que mostra que  $x$  é solução do problema de quadrados mínimos totais. Usando (A.18), a perturbação mínima pode ser escrita como

$$(E; r) = -\gamma^2(A; b) \begin{pmatrix} x \\ -1 \end{pmatrix} (x^*, -1) = \gamma^2(\hat{r}x^*, -\hat{r})$$

onde

$$\hat{r} = b - Ax, \quad \gamma^2 = (1 + \|x\|_2^2)^{-1}.$$

O próximo teorema fornece uma condição suficiente para que o problema (A.17) possua solução única.

**Teorema A.5.4** *Sejam  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n > 0$  os valores singulares de  $A$ . Se  $\sigma_n > \hat{\sigma}_{n+1}$ , então o problema de quadrados mínimos totais possui solução única.*

**Demonstração:** A demonstração pode ser encontrada em [26] ou [70].

## A.6 Ângulo entre Subespaços

Sejam  $\mathcal{S}$  e  $\mathcal{X}$  dois subespaços em  $\mathbb{C}^m$  cujas dimensões satisfazem

$$p = \dim(\mathcal{S}) \geq \dim(\mathcal{X}) = q \geq 1.$$

Os ângulos principais entre  $\mathcal{S}$  e  $\mathcal{X}$ ,  $\theta_1, \dots, \theta_q \in [0, \pi/2]$  são definidos recursivamente por

$$\cos(\theta_k) = \max_{u \in \mathcal{S}} \max_{v \in \mathcal{X}} u^* v = u_k^* v_k$$

sujeito a:

$$\begin{aligned} \|u\| &= \|v\| = 1 \\ u^* u_i &= 0 & i = 1, \dots, k-1 \\ v^* v_i &= 0 & i = 1, \dots, k-1 \end{aligned}$$

Observe que os ângulos principais satisfazem  $0 \leq \theta_1 \leq \dots \leq \theta_q \leq \pi/2$ . Os vetores  $\{u_1, \dots, u_q\}$  e  $\{v_1, \dots, v_q\}$  são chamados *vetores principais* entre os subespaços  $\mathcal{S}$  e  $\mathcal{X}$ .

O maior ângulo principal está relacionado com a noção de distância entre subespaços, que possuem a mesma dimensão. Se  $p = q$ , então  $\text{dist}(\mathcal{S}, \mathcal{X}) = \sqrt{1 - \cos(\theta_p)^2} = \text{sen}(\theta_p)$ .

Se as colunas de  $Q_{\mathcal{S}} \in \mathbb{C}^{m \times p}$  e  $Q_{\mathcal{X}} \in \mathbb{C}^{m \times q}$  definem bases ortonormais para  $\mathcal{S}$  e  $\mathcal{X}$  respectivamente, então:

$$\max_{\substack{u \in \mathcal{S} \\ \|u\|=1}} \max_{\substack{v \in \mathcal{X} \\ \|v\|=1}} u^* v = \max_{\substack{y \in \mathbb{C}^p \\ \|y\|=1}} \max_{\substack{z \in \mathbb{C}^q \\ \|z\|=1}} y^* (Q_{\mathcal{S}}^* Q_{\mathcal{X}}) z.$$

A relação entre os ângulos principais e a SVD é dada no teorema abaixo:

**Teorema A.6.1** *Suponha que  $Q_{\mathcal{S}} \in \mathbb{C}^{m \times p}$  e  $Q_{\mathcal{X}} \in \mathbb{C}^{m \times q}$  são bases ortogonais dos subespaços  $\mathcal{S}$  e  $\mathcal{X}$ . Considere a SVD*

$$M = Q_{\mathcal{S}}^* Q_{\mathcal{X}} = Y C Z^*, \quad C = \text{diag}(\sigma_1, \dots, \sigma_p)$$

onde  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p$ ,  $Y^* Y = Z^* Z = I_q$ . Então os ângulos e vetores principais são dados por

$$\cos \theta_k = \sigma_k, \quad U = Q_{\mathcal{S}} Y, \quad V = Q_{\mathcal{X}} Z.$$

**Demonstração:** A demonstração pode ser encontrada em [26].

□



# Bibliografia

- [1] R. ALLEMANG E D. L. BROWN, *A unified matrix polynomials approach to modal identification*. J. Sound and Vibration, 21(3), 323-333, 1998.
- [2] E. ANDERSON ET AL., *LAPACK Users' Guide 3rd ed.*, SIAM, Philadelphia, 1999.
- [3] R. J. ASTLEY E A. CUMMINGS, *A finite element scheme for attenuation in ducts lined with porous material; Comparison with experiment*, J. Sound Vibration, 2, p.239-263, 1987.
- [4] Z. BAI, J. DEMMEL, J. DONGARRA, A. RUHE E H. VAN DER VORST, *Templates for the solution of Algebraic eigenvalue Problems: A Practical Guide*, SIAM, Philadelphia, 2000.
- [5] Z. BAI E G. W. STEWART, *SRRIT: a FORTRAN subroutine to calculate the dominant invariant subspace of a nonsymmetric matrix*, Computer Science Dept. Report TR 2908, University of Maryland, April 1992.
- [6] F. S. V. BAZÁN E C. BAVASTRI, *An Optimized Pseudo-Inverse Algorithm (OPIA) For Multi-Input Multi-Output Modal Parameter Identification*, Mechanical Systems and Signal Processings, 10 (1996), pp. 365–380.
- [7] F. S. V. BAZÁN E L. H. BEZERRA, *On Zero Location of Predictor Polynomials*, Numer. Linear Algebra With Applications, 4(6), p. 459-468, 1997.
- [8] F. S. V. BAZÁN E M. I. CARDOSO, *Álgebra Linear Computacional para Problemas de Posto Incompleto*, Notas de Minicurso, XXIII CNMAC, Santos-SP, 2000.
- [9] F. S. V. BAZÁN E PH. L. TOINT AND M. C. ZAMBALDI, *A Conjugate-Gradients Based Method for Harmonic Retrieval Problems that Does Not Use Explicit Signal Subspace Computation*, Technical report 16, Department of Mathematics, FUNDP, Namur, Belgium, November 1997.

- [10] F. S. V. BAZÁN E PH. L. TOINT, *Singular value analysis of Predictor Matrices*, Mechanical Systems and Signal Processing, 15(4), p. 667-683, 2001.
- [11] F. S. V. BAZÁN, *Conditioning of Rectangular Vandermonde Matrices with nodes in the Unit Disk*, SIAM Matrix Analysis and Applications, Vol 21(2) pp. 679-693, 2000.
- [12] F. S. V. BAZÁN, *Sensitivity Eigenanalysis of Single Shift-Invariant Subspace-Based Methods*, Signal Processing, 80 pp 89-100, 2000.
- [13] F. S. V. BAZÁN, *CGLS-GCV: a Hybrid algorithm for solving low-rank-deficient problems*. To appear in Appl. Numer. Math.
- [14] F. S. V. BAZÁN, *Error analysis of signal zeros: a projected companion matrix approach*, Lin. Alg. and Appl., 369, p. 153-167, 2003.
- [15] F. S. V. BAZÁN, *Eigensystem realization algorithm: reformulation and system pole perturbation analysis*. Submetido, Julho 2002.
- [16] F. S. V. BAZÁN E PH. L. TOINT, *Conditioning of Infinite Hankel matrices of finite rank*. Systems and Control Letters, 41, 347-359, 2000.
- [17] F. S. V. BAZÁN, *Error Analysis of Signal Zeros from a Related Companion Matrix Eigenvalue Problem*. Appl. Math. Letters, 14, 859-866, 2001.
- [18] F. S. V. BAZÁN, *Application of Generalized Cross-Validation to the Solution of Perturbed Rank-deficient Problems*. Proceedings IMACS2000, Laussane-Switzerland, 2000.
- [19] R. DE BEER E D. VAN ORMONDT, *Analysis of NMR Data Using Time Domain Fitting Procedure*. NMR Basic Principles and Progress, vol 26, 1992.
- [20] R. DE BEER, *c59 Quantitative In Vivo NMR (lecture notes)*. University of Technology Delft, Department of Applied Physics, The Netherlands, 1994.
- [21] L. H. BEZERRA AND F. S. V. BAZÁN, *Eigenvalue Locations of Generalized Companion Predictor Matrices*. SIAM J. Matrix Anal. Appl. Vol. 19(4) pp. 886-897, 1998.
- [22] L. H. BEZERRA AND F. S. V. BAZÁN, *Métodos de Decomposição Parcial do Espectro de uma Matriz*, Notas de Minicurso, XXIV CNMAC, Belo Horizonte MG, Brasil, 2001.

- [23] L. H. BEZERRA, C. TOMEI, AND R. ALAN MCCOY, *Möbius Transforms and Solvers for Large Sparse Generalized Nonsymmetric Eigenvalue Problems*, TR/PA/98/03, CERFACS, Toulouse, France, 1998.
- [24] L. H. BEZERRA AND C. TOMEI, *Spectral Transformation Algorithms for Computing Unstable Modes*. Comp. Appl. Math. 18(1), 1–14, 1999.
- [25] A. BERMÚDEZ, R. G. DURÁN, R. RODRÍGUEZ, E J. SOLOMIN, *Finite element analysis of a quadratic eigenvalue problem arising in dissipative acoustics*, SIAM J. Numerical Anal., 38 (2000), p. 267-291.
- [26] Å. BJÖRCK, *Numerical Methods for Least Squares Problems*. SIAM, Philadelphia, 1996.
- [27] I. BRÁS E T. P. LIMA, *A spectral approach to polynomial matrices solvents*. Appl. Math. Lett. Vol. 9(4), 27-33, 1996.
- [28] K. BRENAN, S. S. CAMPBELL E L. PETZOLD, *Numerical Solution of Initial Value Problems in Differential Algebraic Equations*. North Holland , New York, 1989.
- [29] T. J. BRIDGES E P. J. MORRIS, *Differential eigenvalue problems in which the parameter appears non linearly*, J. Comput. Phys., 55, p. 437-460, 1984.
- [30] D. CALVETTI, L REICHEL E D. C. SORENSEN, *An Implicitly Restarted Method for Large Symmetric Eigenvalue Problems*, Electronic Transactions on Numerical Analysis, 2:1-21, 1994.
- [31] M. I. CARDOSO, *Métodos de Identificação Paramétrica*, Tese de Mestrado, Departamento de Matemática, Universidade Federal de Santa Catarina, Florianópolis SC, Brasil, 2000.
- [32] R. CARIÉRE E R. L. MOSES, *High Resolution Tadar Target Modeling Using a Modified Peony Estimator*, IEEE Transaction on Antenas e Propagation, vol 40.1, 1992.
- [33] F. CHAITIN-CHATELIN E V. FRANYSEÉ, *Lectures on Finite Precision Computations*, SIAM Philadelphia 1996.
- [34] C-T., CHEN, *Linear System Theory and Design*, Oxford University press, 1999.
- [35] M. CLINT AND A. JENNINGS, *A simultaneous iteration method for the unsymmetric eigenvalue problem*, J. Inst. Math. Appl. 8: 111-121, 1971.

- [36] C. E. DAVILA, *A subspace approach to estimation of autoregressive parameters from noisy measurements*. IEEE Trans. Signal Processing, 46 (1998), p. 531-534.
- [37] C. DAVIES, *£ 5m bill to mend wobbly Millenium Bridge*, Electronic Telegraph, 17 November 2000, <http://www.telegraph.co.uk>.
- [38] P. J. DAVIS, *Circulant Matrices*. John Wiley & Sons, 1979.
- [39] G. J. DAVIS, *Numerical solution of a quadratic matrix equation*. SIAM J. Sci. Statist. Comput., 2(1961), p. 164-175.
- [40] G. J. DAVIS, *algorithm 598: An algorithm to compute solvents of the matrix equation  $AX^2 + BX + C = 0$* . Trans. Math. Software, 9(1983), p. 246-254.
- [41] P. DEIFT, T. NANDA, AND C. TOMEI, *Ordinary Differential Equations and the Symmetric Eigenvalue Problem*, SIAM J. Numer. Anal. 20: 1-22, 1983.
- [42] J. W. DEMMEL, *Applied Numerical Linear Algebra*. SIAM, Philadelphia, 1997.
- [43] J. E. DENNIS, J. F. TRAUB E R. P. WEBER, *The algebraic theory of matrix polynomials*. SIAM J. Numer. Anal., 13(1976), p. 831-845.
- [44] J. E. DENNIS, J. F. TRAUB E R. P. WEBER, *Algorithms for solvents of matrix polynomials*. SIAM J. Numer. Anal., 15(1978), p. 523-533.
- [45] D. J. EWINS, *Modal Testing and practice*, John Wiley & Sons Inc, New York, 1989.
- [46] J. EISENFELD, *Operator equations and nonlinear eigen parameter problems*. J. Funct. Anal., 12 (1973), p. 475-490.
- [47] A. FERIANI, F. PEROTTI E V. SIMONCINI, *Iterative Systems Solvers for frequency analysis of linear mechanical systems*, Comput. Methods Appl. Mech. Engrg., 190, North-Holland, Amsterdam, p. 1719-1739, 2000.
- [48] R. A. FRAZER, W. J. DUNCAN E A. R. COLLAR, *Elementary matrices*, 2<sup>a</sup>Edit., Cambridge Univ. Press, London and New York, 1955. Reprint of 1938 edition.
- [49] W. GANDER, G. H. GOLUB, E U. VON MATT, *A constrained eigenvalue problem*. Linear Alg. Appl. Vol. 114-115 (1989), p 815-839.
- [50] W. GANDER, *Least squares with a quadratic constraint*. Numer. Math., 36(181), p.291-307.

- [51] W. GANDER, G. H. GOLUB, E R. STREBER, *Least-squares fitting of circles and ellipses*. BIT, 34 (1994), p. 558-578.
- [52] F. GANTMACHER, *The theory of matrices*, Vol 1, 2, Chelsea, New York, 1960.
- [53] W. GAUTSCHI, *Numerical Analysis -An Introduction-*. Birkhäuser, Boston 1997.
- [54] W. GAWRONSKI E H. G. NATKE, *Order estimation of AR and ARMA Models*. Int. J. Systems Sci. 1988, Vol. 19, No 7, 1143-1148.
- [55] P. E. GILL, W. MURRAY E M. H. WRIGHT, *Practical Optimization*. Academic press, London e New York, 1981.
- [56] W. GIVENS, *Computations of Plane Unitary Rotations Transforming a General Matriz to Triangular Form*. SIAM J. Appl. Math., 6:26-50,1958.
- [57] I. GOHBERG, P. LANCASTER E L. RODMAN, *Matrix Polynomials*. Academic Press, New York, 1982.
- [58] G. H. GOLUB, M. T. HEATH E G. WAHBA, Generalized Cross-Validation as a method for choosing a good ridge parameter. Technometrics, 21:215-223, 1979.
- [59] G. H. GOLUB E C. F. VAN LOAN, *Matrix Computations*, Johns Hopkins University Press, Baltimore, third edition, 1996.
- [60] R. S. HEEG E B. J. GEURTS, *Spatial instabilities of the incompressible-attachment-line flow using sparse matrix Jacobi-Davidson techniques*, Appl., Sci. Res., 59, p. 315-329, 1998.
- [61] N. J. HIGHAM E HYUN-MIN KIM, *Solving a Quadratic Matrix Equation by Newton's Method with Exact Line Searches* SIAM J. Matrix Anal. Appl., 23(2): 303-316, 2001.
- [62] J. J. HOLLKAMP E S. M. BATILL, *Automated parameter identification and order reduction for discrete time series*. AIAA Journal 29, 96-103, 1991.
- [63] R. HORN E C. R. JONHSON, *topics in matrix analysis*. Cambridge University Press, Cambridge, Uk. 1991.
- [64] C. MCGUILLEN E G. COOPER, *Continuous & and Discrete Signals and System Analysis*. Saunders College Publishing, Third Edition, USA 1991.
- [65] P. C. HANSEN, *Rank-Deficient and Discrete Ill-Posed Problems*. SIAM, Philadelphia, 1997.

- [66] PETER S. R. HANSEN, *Signal Subspace Methods for Speech Enhancement*. PH.D Thesis, Department of Mathematical Modeling, Technical University of Denmark, 1997.
- [67] H. J. HIGHAM E F. TISSEUR, *Bounds for Eigenvalues of Matrix Polynomials*. To appear in *Linear Alg. and Its Appl.*
- [68] A. S. HOUSEHOLDER, *Unitary Triangularization of a Nonsymmetric Matrix*. *Journal of the ACM*, 5:339-342, 1958.
- [69] Y. HUA E T. P. SARKAR, *Matrix Pencil Method for estimating parameters of exponentially damped/undamped sinusoids in noise*. *IEEE Trans. On Acoust. Speech and Signal Processings*, ASSP - 38(5):814-824, May 1990.
- [70] S. VAN HUFFEL E J. VANDEWALLE, *The Total Least Squares Problem - Computational Aspects and Analysis*. SIAM, Philadelphia, 1991.
- [71] S. VAN HUFFEL, H. CHEN, C. DECANNIERE E P. VAN HECKE, *Algorithm for time-domain NMR data fitting based on Total Least Squares*. *Journal of Magnetic Resonance A*, 110:228-237, 1994.
- [72] H. C. Ibrahimbegovic, E. L. Chen, E. L. Wilson e R. L. Taylor, *Ritz method for dynamic analysis of large discrete linear systems with non-proportional damping*, *Earthquake Engrg. Strut. Dyn.*, 19, p. 877-889, 1990.
- [73] D. J. INMAN, *Vibration with Control Measurements and Stability*, Prentice Hall, Inc. Englewoods, Cliffs NJ. 1989.
- [74] T. KAILATH, *Linear Systems*, Prentice-Hall, Inc. Englewoods, Cliffs NJ. 1980.
- [75] MD. K. HASAN, MD. J. HOSSAIN E MD. A. HAQUE, *Parameter estimation of multichannel autoregressive process in noise*, *Signal Processing*, 3, p. 603-610, MArch, 2003.
- [76] S. M. KAY, *Noise compensation for autoregressive spectral estimates*, *IEEE Trans. Acoust. Speech Signal Process.* ASSP-28, p. 292-303, 1980.
- [77] R. KUMARESAN E D. W. TUFTS, *Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise*. *IEEE trans. Acoust. Speech and Sig. Proc.*, 30:833-840, 1982.

- [78] R. KUMARESAN, D. W. TUFTS E L. L. SCHARF, *A Prony Method for noisy data: Choosing the signal components and selecting the order in Exponential Signal Models*. Proceedings of the IEEE, Vol. 72:230-233, february, 1984.
- [79] S.Y. KUNG, *A New Identification and Model Reduction Algorithm via Singular Value Decomposition*. Twelfth Asilomar Conf. on Circuits, Systems and Comp., (Asilomar,CA), 705-714, Nov. 1978.
- [80] M. KUZUOGLU E R. MITRA, *Finite element solution of electromagnetic problems over a wide frequency range via the Padé approximations*, Comput. Methods. Appl. Mech. Engrg., 169, p. 263-277, 1999.
- [81] P. LANCASTER, *Lambda-Matrices and Vibrating Systems*. Pergamon Press, Oxford, 1966.
- [82] A. J. LAUB, *Efficient multivariable frequency response computations*. IEEE Trans. Automat. Control, 26 (1981), p. 407-508.
- [83] R. B. LEHOUCQ, *Analysis and Implementation of an Implicitly Restarted Arnoldi Iteration*. PhD Thesis, Rice University, Houston, Texas, 1995.
- [84] R. B. LEHOUCQ, D. C. SORENSEN AND C. YANG, *ARPACK Users' Guide: Solution of Large Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*, SIAM, Philadelphia, 1998.
- [85] R. B. LEHOUCQ E D. C. SORENSEN, *Deflation techniques for an Implicitly Restarted Arnoldi Iteration*. SIAM J. Matrix Anal. Appl., 17(4):789-821, 1996.
- [86] J.-S. LEW, J.-N. JUANG E R. W. LONGMAN, *Comparison of Several system Identification Methods For Flexible Structures*. J. Sound and Vibration 167(3) p. 461-480, 1993.
- [87] E. LUNDSTRÖM, *Singular Value Computations for Toeplitz Matrices*. Lic. Thesis, Linköping University, ISBN 91-7871-719-1, May 1996.
- [88] M. MOONEN, B. DE MOOR, L. VANDENBERGHE, J. VANDEWALLE. *On- and Off-line identification of linear state-space models*. Int. J. On Control, Vol. 49 No. 1, 219-232, 1989.
- [89] R. B. MORGAN, *On Restarting the Arnoldi Method for Large Nonsymmetric Eigenvalue Problems*, Math. Comp. 65: 1213-1230, 1996.

- [90] N. K. NICHOLS E J. KAUTSKY, *Robust eigenstructure Assignment in Quadratic Polynomials: Nonsingular Case*. J Matrix Anal. Appl. Vol 23(1), pp. 77-102, 2001.
- [91] V. Y. PAN, *solving polynomial equations: some history and recent progress*. SIAM Review, 33(1997), p. 187-220.
- [92] PETER VAN OVERSCHE E B. DE MOOR, *Subspace Identification for Linear Systems*. Kluwer Academic Publishers, 1996 USA.
- [93] C. PAIGE, *The Computation of Eigenvalues and Eigenvectors of Very Large Sparse Matrices*. PhD Thesis, University of London, 1971.
- [94] C. PAIGE, *Error Analysis of the Lanczos Algorithm for Tridiagonalizing a Symmetric Matrix*. J. Inst. Maths Applics, 18:341-349, 1976.
- [95] B. N. PARLETT, *The Symmetric Eigenvalue Problem*. SIAM, Philadelphia, 1997.
- [96] B. N. PARLETT E D. S. SCOTT, *The Lanczos Algorithm with Selective Orthogonalization*. Math. Comp.,33(145):217-238,1979.
- [97] B. N. PARLETT, D. TAYLOR AND Z. LIU, *A Look-Ahead Lanczøs Algorithm for Unsymmetric Matrices*, Math. Comp. 44: 105–124, 1985.
- [98] E. PEREIRA, *On Solvents of Matrix polynomials*. Aceito para Publicação em Applied Numerical Math.
- [99] P. M. PRENTER, *Splines and variational methods*, Wiley-Interscience, New York, 1975.
- [100] R. DE PRONY, *Essai expérimental et analytique sur les loi de la dilatabilité et sur celles de la force expansive de la vapeur de l'eau et de la capeur de l'alkool, à différentes températures*, J. de l'Ecole Polytechnique, 1:24-76, 1795.
- [101] W. E. ROTH, *A solução da matrix equation  $P(X) = A$* , Tran. Amer. Math. Soc. 30(1928), p. 579-596.
- [102] W. E. ROTH, *On the unilateral equation in matrices*, Tran. Amer. Math. Soc. 32(1930), p. 61-80.
- [103] H. D. SIMON, *Analysis of the Symmetric Lanczos Algorithm with reorthogonalization Methods*, Linear Algebra Appl.,61:101-131,1984.

- [104] R. A. SMITH, *The Condition Numbers of the Matrix Eigenvalue Problem*, Numerische Mathematik, 10, 132:240, 1967.
- [105] G. W. STEWART, *Perturbation Theory for the Singular Value Decomposition*, SVD and Signal Processing II: Algorithms, Analysis and Applications, 99-109, Elsevier, 1991.
- [106] D. C. SORENSEN, *Implicit Application of Polynomial Filters in a  $k$ -step Arnoldi Method*, SIAM J. matrix Appl., 13:357-385, 1992.
- [107] G. W. STEWART, *Error and perturbation bounds for subspaces associated with certain eigenvalue problems*, SIAM Review. 32:727-764, 1973.
- [108] G. W. STEWART E JI-GUANG SUNG, *Matrix Perturbation Theory*, Academic Press, Inc. New York, 1990. 32:727-764, 1973.
- [109] G. STRANG, *Linear Algebra and its Applications*, Academic Press, 1992.
- [110] G. STRANG E G. J. FIX, *An analysis of the Finite Element Method*, Prentice Hall, Englewood Cliffs, NJ, 1973.
- [111] H. RUTISHAUSER, *Simultaneous Iteration Method for Real Symmetric Matrices*. Numer. Math. 16, 205-223, 1970.
- [112] A. RUHE, *Algorithms for the nonlinear eigenvalue problem*, SIAM J. Numer. Anal., 10, p. 674-689, 1973.
- [113] A. RUHE E P. A. WEDIN, *Algorithms for separable nonlinear least squares problems*. SIAM Rev. 22(3) 1980.
- [114] Y. SAAD, *Variations of Arnoldi's Method for Computing Eigenelements of Large Unsymmetric Matrices*, Lin. Alg. and Its Applic. 34: 269-295, 1980.
- [115] Y. SAAD, *Practical Use of Some Krylov Subspace Methods for Solving Indefinite and Nonsymmetric Linear Systems*, SIAM J. Sci. and Stat. Comp. 6:865-882, 1984.
- [116] Y. SAAD, *Chebyshev Acceleration Techniques for Solving Nonsymmetric Eigenvalue Problems*, Math. Comp. 42: 567-566, 1984.
- [117] Y. SAAD, *Numerical Methods for Large Eigenvalue Problems: Theory and Algorithms*, Wiley, New York, 1992.

- [118] J. TOTZ, A. V. D. BOOGAART, S. V. HUFFEL, D. GRAVERON-DEMILLY, I. DOLOGKOU, R. HEIDLER AND D. MICHEL, *The Use of Continuous Regularization in the Automated Analysis of MRS Time-Domain Data*. Journal of Magnetic Resonance. 124, 400-409, 1997.
- [119] RICHARD J. VACCARO, *A second perturbation expansion for the SVD*. SIAM Journal on Matrix Analysis and Applications, 15(2):661-671, Abril, 1994.
- [120] SWINDLEHURST, R. B. ROY, B. OTTERSTEN, T. K. KAILATH, *A subspace fitting method for identification of linear state space models*. IEEE Trans. Autom. Control 40(2) Feb. 1995 316-331.
- [121] F. TISSEUR, *Backward error and condition of polynomial eigenvalue problems*. Linear Alg. And Its Appl. 309, 339-361, 2000.
- [122] F. TISSEUR E K. MEERBERGEN, *The quadrátic Eigenvalue Problem*. Siam Review, Vol. 43, No. 2, p. 235-286, 2001.
- [123] LL. N. TREFETHEN, D. BAU, III, *Numerical Linear Algebra*. SIAM Philadelphia, 1997.
- [124] J. M. VARAH, *On fitting exponentials by non linear least squares*. J. SIAM J. Sci. Statist. Comput. 6(1985) 30-44.
- [125] A. VAN DER VEEN, E. F. DEPRETTERE E A. LEE SWINDLEHURST, *Subspace-based Signal Analysis using Singular Value Decomposition*. Proceedings of the IEEE, 81(9):1277-1309, Setembro, 1993.
- [126] M. VIBERG E B. OTTERSTEN, *Sensor Array Procesing Based on Subspace Fitting*, IEEE Trans. Signal Processing, vol 39, pp. 1110-1121, 1991.
- [127] H. VOLD, J. KUNDRATH, T. ROCKLIN E R. RUSSELL, *A multi-input modal estimation alforithm for mini-computers*, SAE Transactions 91, 815-821, 1982.
- [128] GRACE WAHBA, *Spline Functions for Observational Data*. CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 59, SIAM, Philadelphia, 1990.
- [129] D. S. WATKINS, *Fundamentals of Matrix Computations*. John Wiley & Sons, New York, 1191.
- [130] G. WALKER, *On periodicity in series of related terms*. Proc. Royal Soc. London Ser. A, 131A:518-532, 1931.

- [131] J. H. WILKINSON, *The Algebraic Eigenvalue Problem*. Oxford University Press, Oxford, Great Britain, 1965.
- [132] E. L. WILSON, M.-W. YUAN E J. M. DICKENS, *Dynamic analysis by direct superposition of Ritz vectors*, Earthquake Engrg. Dyn., 10, p. 813-821, 1982.
- [133] G. XU E T. KAILATH, *Fast Estimation od Principal Eigenspace using Lanczos Algorithm*, SIAM Jornal on Matrix Analysis and Applications, 15:974-994, 1994.
- [134] G. U. YULE, *On a method of investigating periodices in disturbed series, with special reference to Wolfer's sinspot numbers*. Philos. Trans. Roy. Soc. London Ser. A, 226A:267:298, 1927.
- [135] H. P. ZEIGER E A. J. MAC-EWEN, *Aproximate Linear Realization of Given Dimension via Ho's Algorithm*, IEEE, Trans. Aut. Control AC-19, pp 153, 1974.