



Universidade Estadual de Campinas  
Faculdade de Engenharia Elétrica e de Computação  
Departamento de Comunicações

MÍNIMOS QUADRADOS TOTAIS  
&  
MÁXIMA VEROSSIMILHANÇA  
EM  
ESTIMAÇÃO DE FREQUÊNCIAS

Tese submetida à Faculdade de Engenharia Elétrica e de Computação da  
Universidade Estadual de Campinas, Departamento de Comunicações,  
como parte dos requisitos exigidos para a obtenção do título de  
Doutor em Engenharia Elétrica

por

*Rodrigo Pinto Lemos* ✱  
Mestre em Engenharia Elétrica - UNICAMP - 1995

Orientador  
*Amauri Lopes* ✱  
Doutor em Engenharia Elétrica - UNICAMP - 1982

CAMPINAS  
1997

a examinar o respectivo a redação final da tese  
apresentada por **RODRIGO PINTO LEMOS**  
e aprovada pela Comissão  
adure em 27 / 06 / 1997.  
*Amauri Lopes*  
Orientador

UNIDADE	BC
N.º CHAMADA:	
T.º	Unicamp
L554m	
E.º	
T.º 5	31160
F.º	281197
C.º	0   1   0   X
PREÇO	R\$ 11,00
DATA	24/07/97
N.º OPD	

CM-0009970 1-1

FICHA CATALOGRÁFICA ELABORADA PELA  
BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

L554m

Lemos, Rodrigo Pinto

Mínimos quadrados totais e máxima verossimilhança em estimação de frequências / Rodrigo Pinto Lemos.-- Campinas, SP: [s.n.], 1997.

Orientador: Amauri Lopes.

Tese (doutorado) - Universidade Estadual de Campinas Faculdade de Engenharia Elétrica e de Computação.

1. Processamento de sinais - Técnicas digitais.<sup>x</sup> 2. Detecção de sinais.<sup>x</sup> 3. Máxima e mínima.<sup>x</sup> 4. Otimização matemática.<sup>x</sup> 5. Discriminadores de frequência.<sup>x</sup> I. Lopes, Amauri. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título.

*Com amor para  
Luís Guilherme*

## Abstract

This work establishes an unifying framework to the use of linear prediction for estimating the unknown frequencies of signals corrupted by additive noise. The noise and signal subspaces associated to the data are detailed using the singular value decomposition. The noise effects are studied in order to achieve its maximum reduction. A new method is proposed that optimizes linear prediction in total least squares sense and a signal subspace approach to it is developed. This method is shown to be competitive to Principal Components, although both methods still perform a little worse than Maximum Likelihood. In order to make linear prediction perform like Maximum Likelihood, four recent linear prediction methods take into account the structure imposed by the linear prediction filter on the data matrix. These methods are shown to minimize the same objective function and yield equivalent solutions. A new method is proposed that generalizes these methods to the fixed elements case. Another three methods are also proposed that take into account the data matrix structure but using a least squares optimization approach. These methods are shown to be equivalent and to perform better than Principal Components method.

## Sumário

Este trabalho revisa a aplicação de predição linear para a estimação de frequências em meio ruidoso, estabelecendo uma formulação unificada. Utiliza-se a decomposição em valores singulares para detalhar os subespaços de sinal e de ruído, associados aos dados, buscando compreender a extensão da atuação ruidosa, de forma a reduzi-la à mínima possível. Propõe-se um novo método que otimiza a predição linear no sentido dos mínimos quadrados totais, e para ele se desenvolve uma solução pelo subespaço de sinal. Este método é competitivo com o método Componentes Principais, embora seus desempenhos sejam inferiores ao obtido por Máxima Verossimilhança. Descrevem-se quatro métodos recentes que otimizam a predição linear no sentido dos mínimos quadrados totais e que, considerando a estrutura inserida na matriz de dados pelo filtro de predição, alcançam o desempenho da Máxima Verossimilhança. Demonstra-se analiticamente a equivalência entre esses métodos e se comparam seus desempenhos. Propõe-se um novo método que generaliza os anteriores para o caso de elementos fixos, e mais três métodos para a otimização no sentido dos mínimos quadrados convencional. Demonstra-se a equivalência entre estes últimos e verifica-se que seu desempenho supera o do Componentes Principais.

# Agradecimentos

Muito vivi nestes anos de UNICAMP. Foram anos que mudaram definitivamente minha vida, que me viram começar e recomeçar, que me encontraram sozinho e me trouxeram uma família. Foram tempos de produção: de conhecimentos e de uma nova vida, a de meu filho. Muito, muito aprendi, seja com os livros, artigos, professores, amigos, colegas e principalmente com meu Amor verdadeiro, minha esposa, Lucimeire. Nestas poucas linhas pretendo ser capaz de me lembrar um pouquinho e de agradecer a todos aqueles que me auxiliaram na conquista deste sonho adolescente. Àqueles que, porventura, faltam uma palavra, peço a compreensão, pois, felizmente, são tantos a quem agradecer e tanto o que agradecer, que não excluo a possibilidade de um eventual esquecimento. Porém, saibam eles que em algum momento que me vierem à mente as boas lembranças, certamente contarão com minha gratidão.

Primeiramente, agradeço a Deus a oportunidade de progredir, de reencarnar e viver. Agradeço também aos espíritos laboriosos e amigos que, a serviço da Divina Providência, socorreram-me em meus momentos de aflição. A inspiração bendita que me concederam, aliada ao trabalho incessante e apaixonado, é grande responsável pela conclusão deste documento.

É com muita alegria e orgulho que mais uma vez venho tecer meus agradecimentos ao Prof. Dr. Amauri Lopes que, mesmo frente aos compromissos próprios da Direção da FEEC, sempre dedicou parte de seu escasso tempo a este trabalho, enriquecendo-o com seus comentários preciosos e precisos. Já disse isto no Mestrado mas insisto em repetir: orgulho-me de nossa amizade.

Agradeço especialmente ao amigo Engenheiro Antônio Augusto Teixeira Pinto de Moraes, que por quantas vezes me socorreu, brilhantemente apresentando meus trabalhos nos eventos internacionais de que não pude participar devido à proximidade do nascimento de meu filho. Sou-lhe eternamente grato e digo-lhe que poderá contar sempre com minha amizade.

Meus muitíssimos agradecimentos ao amigo Prof. Dr. Jorge Pedraza Arpasi que muito me auxiliou emprestando-me seus conhecimentos da língua espanhola e que muitíssimo me honra com sua amizade. Nossas conversas sempre foram semente de questionamentos e amadurecimento. Mesmo nas horas de amenidades, de um cafezinho, foi sempre um prazer desfrutar de sua culta presença. Os cafés jamais serão os mesmos sem polêmica e sem sotaque.

Agradeço ao amigo Prof. José Edson Vargas Bautista pelo apoio e paciência dispensados a mim em nossa viagem à Argentina. Aqueles dias serão sempre inesquecíveis, pelas alegrias, confusões, surpresas e risadas.

Muitíssimo tenho a agradecer ao Prof. Dr. Weber Martins e ao Prof. Sérgio de Araújo Figueiredo, que desde há muito tempo têm sido meus bons amigos e que muito se empenharam

para que fosse concedido meu afastamento, conseguindo conciliar gregos e troianos, ou melhor, CPPD e PRPPG. Sem seu apoio, talvez este trabalho jamais tivesse sido concluído.

Agradeço também aos amigos Prof. Alberto José Centeno Filho, Prof. Enes Gonçalves Marra e Prof. Antônio César Baleeiro Alves, que tantas orientações me trouxeram a respeito das regras de funcionamento da UFG e do Serviço Público Federal, auxiliando-me em minha nova carreira.

Meus muitos agradecimentos à Universidade Federal de Goiás pela confiança e crédito em mim depositados, permitindo manter-me afastado de minhas funções docentes, mesmo durante o estágio probatório, para a conclusão deste trabalho.

Gostaria de agradecer também ao Prof. Dr. Paulo Morelato França pela confiança em mim depositada quando da incerteza dos momentos de transição que vivi no ano passado.

Agradeço fortemente à Coordenação de Aperfeiçoamento do Pessoal de Nível Superior - CAPES, pelo suporte financeiro a este trabalho desde seu início. Sua seriedade e eficiência são notórias dentre os órgãos financiadores da pesquisa científica neste país.

Meus agradecimentos também ao FAEP/ FUNCAMP que muito prontamente financiou minhas participações em eventos científicos internacionais, possibilitando não somente a divulgação do conhecimento aqui adquirido, bem como do papel capital e distinto da UNICAMP no panorama científico brasileiro e internacional.

Muito agradeço à Sociedade Brasileira de Telecomunicações, na pessoa do Prof. Dr. João Marcos Travassos Romano, pelo suporte financeiro concedido à minha participação no 13o. Simpósio Brasileiro de Telecomunicações.

Meus muitos agradecimentos aos Professores com quem tomei contato nestes anos. Sua atuação impecável será sempre exemplar em minha carreira. Muito orgulhosamente fui seu aluno e hoje tenho a honra de poder chamá-los de colegas.

Aos companheiros do DECOM meu muito obrigado pelos momentos alegres que passamos juntos. Aos funcionários Mário, Renata e Ricardo, meus agradecimentos pela sua ajuda sempre oportuna e boa vontade.

Agradeço a minha mãe, sempre carinhosa, sempre esta figura extraordinária de coragem, disposição e justiça. Ela é e sempre será a pessoa em quem eu procuro me espelhar. Dedico-lhe minha profunda admiração e gratidão pelo seu inestimável apoio.

A minha irmã, familiares, sogros e cunhados, meus eternos agradecimentos e o orgulho de dizer que somos parte de uma mesma família.

Agora muito especialmente agradeço à minha esposa Lucimeire, que sempre me brindou com sua compreensão, sua dedicação e carinho. Seu estímulo foi fundamental para que eu pudesse me lançar nesta empreitada. Se faço estas coisas por alguém, é por ela, que deu novo sentido à minha vida e deu-me um filho de que me orgulho. Nada que eu diga ou que faça será jamais suficiente para agradecer a sua presença luminosa em minha vida. Somente posso agradecer a Deus por ela existir e dedicar-lhe meu Amor eterno.

Rodrigo Pinto Lemos  
Campinas, Junho de 1997.

# Índice

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Máxima Verossimilhança</b>	<b>4</b>
2.1	Sinal com uma Única Exponencial Complexa . . . . .	4
2.2	Sinal com Múltiplas Exponenciais Complexas . . . . .	6
<b>3</b>	<b>Limite de Cramér-Rao</b>	<b>9</b>
3.1	Sinal com uma Única Exponencial Complexa . . . . .	9
3.2	Sinal com Múltiplas Exponenciais Complexas . . . . .	12
<b>4</b>	<b>Estimação de Frequências por Predição Linear</b>	<b>17</b>
4.1	Predição Linear Forward-Backward - FBLP . . . . .	17
4.2	O Método FBLP para Estimação de Frequências . . . . .	22
4.2.1	A Função de Transferência do Filtro FBLP . . . . .	22
4.3	O Caso Ruidoso . . . . .	23
<b>5</b>	<b>SVD e Subespaços</b>	<b>26</b>
5.1	Definições . . . . .	26
5.2	Decomposição em Valores Singulares . . . . .	27
5.3	A SVD da Matriz $\mathbf{A}$ e o Vetor $\mathbf{w}'$ . . . . .	27
5.4	Os Subespaços de Sinal e de Ruído . . . . .	28
5.5	Dissecando os Subespaços . . . . .	29
5.5.1	A Matriz de Dados Exatos $\mathbf{A}_E$ . . . . .	29
5.5.2	Considerações sobre a SVD de $\mathbf{A}$ . . . . .	30
5.5.3	A SVD de $\mathbf{A}$ em termos da SVD de $\mathbf{A}_E$ . . . . .	32
5.5.4	Limitações no Conhecimento de $\mathbf{A}_E$ . . . . .	33
5.6	Resumindo . . . . .	34

5.7	Considerações Estatísticas . . . . .	35
5.8	Considerações de Ordem Prática . . . . .	36
<b>6</b>	<b>FBLP Modificado</b>	<b>37</b>
6.1	Mínimos Quadrados com Posto Reduzido - LSr . . . . .	37
6.2	Melhorando o FBLP . . . . .	38
6.3	Desempenho . . . . .	39
<b>7</b>	<b>FBLP e Mínimos Quadrados Totais - TLS-FBLP</b>	<b>41</b>
7.1	Mínimos Quadrados Totais - TLS . . . . .	41
7.1.1	Análise pela Teoria de Perturbações . . . . .	42
7.1.2	A Solução dos Mínimos Quadrados Totais . . . . .	43
7.1.3	As Soluções TLS pelos Subespaços de Sinal e de Ruído . . . . .	45
7.2	Comparação entre TLS e LS . . . . .	47
7.2.1	Interpretação Geométrica dos Mínimos Quadrados e Mínimos Quadrados Totais . . . . .	48
7.2.2	A Solução LS Equivalente . . . . .	51
7.2.3	Solução FBLP Modificado Expressa em termos do TLS . . . . .	53
7.3	Desempenhos em Perspectiva . . . . .	55
7.4	Conclusões . . . . .	60
<b>8</b>	<b>Métodos TLS com Desempenho ML</b>	<b>61</b>
8.1	Máxima Verossimilhança Quadrática Iterativa - IQML . . . . .	61
8.1.1	Convergência . . . . .	67
8.1.2	Desempenho . . . . .	69
8.2	Mínimos Quadrados Totais Restrito - CTLS . . . . .	71
8.3	Mínimos Quadrados Totais Branqueado - WTLS . . . . .	74
8.3.1	Desempenho . . . . .	76
8.4	Conclusões . . . . .	76
<b>9</b>	<b>Mínimos Quadrados Totais Estruturado - STLS</b>	<b>79</b>
9.1	Revisando o TLS . . . . .	79
9.1.1	Derivadas . . . . .	80
9.1.2	Eliminação de $\tilde{\mathbf{X}}$ . . . . .	81
9.1.3	Normalização . . . . .	81

9.2	Mínimos Quadrados Totais Estruturado (STLS)	81
9.2.1	Derivadas	83
9.2.2	Eliminação de $\tilde{X}$	83
9.2.3	Normalização	84
9.2.4	STLS na forma de uma SVD Riemanniana	84
9.3	Equivalência entre STLS e CTLS	86
9.4	Conclusões	89
<b>10</b>	<b>Estruturando o LS</b>	<b>90</b>
10.1	Mínimos Quadrados Restrito - CLS	91
10.2	Mínimos Quadrados Totais Estruturado com Elementos Fixos	92
10.3	Mínimos Quadrados Estruturado - SLS	97
10.4	Mínimos Quadrados Branqueado - WLS	101
10.4.1	Desempenho	102
10.5	Conclusões	103
<b>11</b>	<b>Conclusões e Sugestões</b>	<b>104</b>
11.1	Conclusões	104
11.2	Questões para Prosseguimento	106
<b>A</b>	<b>Diferenciação do Lagrangiano</b>	<b>108</b>
A.1	Definições Básicas	108
A.1.1	Derivada Complexa de uma Função Escalar	108
A.1.2	Extensão para o Caso Vetorial	110
A.2	Diferenciando o Lagrangiano do TLS	112
A.3	Diferenciando o Lagrangiano do STLS	114
<b>B</b>	<b>Quadro Resumo dos Métodos Estudados</b>	<b>116</b>
<b>C</b>	<b>Glossário de Símbolos</b>	<b>117</b>
	<b>Bibliografia</b>	<b>120</b>

# Lista de Figuras

4.1	Filtro de erro de predição forward. . . . .	18
4.2	Filtro de erro de predição backward. . . . .	18
4.3	Zeros da função de transferência, $L = 24$ , $\text{SNR} = \infty$ . . . . .	23
4.4	Zeros da função de transferência, $L = 24$ , $\text{SNR} = 10\text{dB}$ . . . . .	24
4.5	Comparação de desempenho entre os métodos ML e FBLP. . . . .	25
6.1	Zeros da função de transferência FBLP, $N = 25$ amostras, $L = 17$ , $\text{SNR} = 10\text{dB}$ . . . . .	39
6.2	Zeros da função de transferência FBLP Modificado, $N = 25$ amostras, $L = 17$ , $\text{SNR} = 10\text{dB}$ . . . . .	39
6.3	Comparação de desempenho entre os métodos ML, FBLP e FBLP Modificado. . . . .	40
7.1	Diferenças no ajuste da reta $w$ pelos critérios LS ou TLS. . . . .	49
7.2	O ajuste LS. . . . .	50
7.3	O ajuste TLS. . . . .	50
7.4	Superfície de desempenho para o método TLS-FBLP. . . . .	57
7.5	Superfície de desempenho para o método FBLP Modificado. . . . .	57
7.6	Linhas de nível para a superfície de desempenho TLS-FBLP (vista de topo). . . . .	58
7.7	Linhas de nível para a superfície de desempenho FBLP Modificado(vista de topo). . . . .	58
7.8	Comparação entre os melhores desempenhos TLS-FBLP e FBLP Modificado. . . . .	59
8.1	Desempenho do método IQML, $N = 25$ , $L = 2$ , $M = 2$ , 20 experimentos. . . . .	70
8.2	Desempenhos dos métodos IQML, FBLP Modificado, TLS-FBLP e ML. . . . .	70
8.3	Desempenho do método WTLS, na versão FBLP, obtido para 20 experimentos. . . . .	77
8.4	Desempenhos dos métodos WTLS e IQML, para 20 experimentos. . . . .	77
10.1	Desempenho do método WLS, na versão FBLP, para $L = 24$ . . . . .	103

# Capítulo 1

## Introdução

A estimação de frequências é uma dos temas principais de Processamento de Sinais. Esta importância e o progresso das pesquisas estão associados a grande diversidade de aplicações, abrangendo problemas de Análise Espectral, Modelagem e Identificação de Sistemas, Processamento de Sinais Biomédicos, Fotônica, Geofísica, filtragem espaço-temporal efetuada por Arranjos de Antenas, Radar, Sonar, Radioastronomia, etc. Perspectivas de novas aplicações incluem Comunicações Pessoais e a Detecção e Localização Automática de Falhas em Máquinas e Motores, Controle Óptico de Qualidade em Manufatura e Caracterização de Formas de Objetos em Tomografia [45].

O problema de estimação de frequências torna-se mais crítico à medida em que se reduz o número de amostras disponíveis para análise e se diminui a relação sinal-ruído. Isto fica claro observando os métodos convencionais, que utilizam o espectro de potência do sinal para estimar as frequências a partir das posições dos picos espectrais. Quando o intervalo de observação, representado pelo número finito de amostras do sinal, for inferior ao recíproco da diferença entre as frequências a serem estimadas, estes métodos não apresentam resolução espectral suficiente para distinguir entre estas frequências. Assim, quanto menor o número de amostras disponíveis, pior a resolução. Isto é fundamental em aplicações como as que utilizam arranjos de antenas, onde o limite físico de comprimento do arranjo, ou seja, a abertura, pode resultar em um número pequeno de sensores. Além disso, a presença de ruído aditivo corrompe as características do sinal estudado, distorcendo a informação nele contida. Portanto, quanto maior a potência do ruído, pior a qualidade das estimativas das frequências.

Em um sistema de estimação é importante conhecermos números que indiquem a melhor estimação possível com os dados disponíveis, ou observações. O valor mínimo possível da variância do erro de estimação define o limite de desempenho de Cramér-Rao (CRB), apresentado por Rife e Boorstin [85] para o caso de sinais senoidais em ruído branco. Um estimador não-polarizado cuja variância aproxima-se ou se iguala ao CRB, é dito ser ótimo.

Embora o melhor desempenho seja alcançado por métodos baseados em Máxima Verossimilhança, o esforço computacional proibitivo associado tem motivado a busca de soluções que aproximem esse desempenho máximo, porém mantendo a factibilidade de aplicação a problemas práticos.

Em 1795, Gaspard Riche, o Barão de Prony, propôs um método [80] para o ajuste de curvas

a um conjunto de pontos, relativos a valores medidos da dilatação de fluidos. Este método contém os conceitos básicos da estimação de frequências por predição linear [32], permitindo estimar satisfatoriamente as frequências na ausência de ruído, mesmo para um número pequeno de amostras disponíveis. Nas últimas décadas, foram propostas diversas variantes à proposta de Prony, desenvolvidas para combater os efeitos do ruído sobre a estimação.

Ulrych e Clayton [93] e Nuttall [72] em 1976, propuseram o método de predição linear *forward-backward* (FBLP) que utiliza um filtro de predição *forward* e outro *backward*, otimizados conjuntamente. Este método fornece melhores estimativas para as frequências do que os métodos paramétricos convencionais, quando existem poucas amostras disponíveis e relações sinal-ruído (SNR) grandes. Entretanto, para SNR's pequenas, ele se revela muito mal condicionado.

Tufts e Kumaresan [91], em seu método Componentes Principais, ou FBLP Modificado, utilizaram a decomposição em valores singulares para separar a informação dos dados observados em subespaços de sinal e de ruído. Utiliza-se, então, apenas a informação do subespaço de sinal, descartando a contribuição do subespaço de ruído. Isto ocorre pela exclusão das parcelas correspondentes a esse subespaço, na recomposição da matriz de dados, obtendo-se uma matriz corrigida. A esta matriz é aplicado o método de Prony. Os autores mostraram que o seu método garante um desempenho satisfatório para valores de SNR muito menores do que os demais métodos. Porém, ainda neste caso, o desempenho cai rapidamente para valores de SNR abaixo de 7 dB, enquanto o desempenho da Máxima Verossimilhança sustenta-se até 3 dB de SNR.

Em 1980, Golub e Van Loan [28] apresentaram o critério de otimização dos mínimos quadrados totais (TLS), que atua na minimização do erro de predição linear considerando perturbações não somente no vetor desejado, como efetuado pelo critério de mínimos quadrados (LS), mas também na matriz de dados. Em 1995, Lemos e Lopes [53] propuseram a otimização do método FBLP no sentido TLS, originando o método TLS-FBLP que, em 1996, foi modificado pelos autores [60] para utilizar apenas o subespaço de sinal. Esse método mostrou-se competitivo com o FBLP Modificado, porém seu desempenho, assim como o deste último, ainda é significativamente inferior ao da Máxima Verossimilhança.

Acreditamos que essa sub-otimalidade se deve ao fato de tanto o FBLP Modificado quanto o TLS-FBLP, através do truncamento da decomposição em valores singulares, não preservarem a estrutura que é inserida pelo filtro de predição linear na formulação da matriz de dados. Quando descartamos a contribuição do subespaço de ruído da matriz de dados, produzindo uma matriz corrigida, descartamos também a estrutura original da matriz. Embora esta correção proporcione significativa melhora no desempenho da estimação de frequências, ela desestrutura o esquema de predição linear.

No sentido de obter desempenhos mais competitivos em relação ao obtido pela Máxima Verossimilhança, foram propostos, recentemente, diversos métodos baseados no critério TLS que permitem preservar a estrutura da predição ao descartar a contribuição do subespaço de ruído: em 1990, Hua e Sarkar propuseram o método Mínimos Quadrados Totais Branqueado (WTLS) [39]; em 1991, Abatzoglou e Mendel apresentaram o método Mínimos Quadrados Totais Restrito (CTLS) [2] e, em 1993, De Moor introduziu o método Mínimos Quadrados Estruturado (STLS) [18]. Alternativamente, em 1986, Bresler e Macovski derivaram uma expressão exata para o

critério de máxima verossimilhança em termos do polinômio de predição do sinal isento de ruído, originando o método Máxima Verossimilhança Quadrática Iterativa (IQML) [6].

Este trabalho pretende apresentar uma revisão de todos estes métodos, estabelecendo uma formulação unificada em termos de notação, abordagem e desenvolvimento que, até onde temos notícia, ainda não havia sido realizada. Nos Capítulos 2 e 3 descrevemos as formulações do critério de Máxima Verossimilhança e do Limite de Cramér-Rao. No Capítulo 4 introduzimos a predição linear como ferramenta para a estimação de frequências. No Capítulo 5 discutimos as limitações da decomposição em valores singulares quando o sinal é corrompido por ruído aditivo. No Capítulo 6 descrevemos o método FBLP Modificado e apresentamos seu desempenho. No Capítulo 7 introduzimos o método TLS-FBLP, detalhando e comparando sua formulação e seu desempenho àquele do FBLP Modificado. No Capítulo 8 apresentamos de forma detalhada os métodos IQML, CTLS e WTLS, comparando suas estruturas e desempenhos e mostrando que, na verdade, eles são equivalentes entre si. No Capítulo 9 descrevemos de forma destacada o método STLS, demonstrando sua equivalência ao método CTLS.

A partir da formulação unificada, no Capítulo 10 propomos uma generalização de todos estes métodos, que permite ampliar a diversidade de problemas a serem tratados, aumentando a gama de aplicações da teoria aqui exposta. Finalizando, particularizamos esta visão geral, desenvolvendo versões LS para os métodos CTLS, STLS e WTLS. Estas propostas vêm ao encontro do esforço de pesquisa internacional, abrindo perspectivas para a busca de algoritmos mais simples e robustos. No Capítulo 11, apresentamos as conclusões e diversas questões para prosseguimento, demonstrando de quão fértil terreno ainda se dispõe em estimação de frequências.

# Capítulo 2

## Máxima Verossimilhança

Neste capítulo estudaremos a estimação de Máxima Verossimilhança (ML) dos parâmetros de exponenciais complexas imersas em ruído branco gaussiano complexo. A ML reconhecidamente fornece as estimativas mais precisas e com maior robustez aos efeitos do ruído. Entretanto, veremos a seguir que a implementação da ML é complicada e que o esforço computacional por ela exigido é proibitivo.

### 2.1 Sinal com uma Única Exponencial Complexa

Inicialmente, analisaremos a estimação da frequência de uma única exponencial complexa. Neste caso, a estimativa ML pode ser obtida simplesmente da transformada de Fourier dos dados. Particularmente, a estimativa ML será dada pela frequência na qual o periodograma assume seu valor máximo. Dessa forma, podemos modelar uma sequência  $x[n]$  de  $N$  amostras uniformemente espaçadas de um sinal exponencial complexo corrompido pelo ruído  $y[n]$  branco aditivo gaussiano complexo:

$$x[n] = a_1 \exp[j(\omega_1 n + \phi_1)] + y[n], \quad n = 0, \dots, N - 1 \quad (2.1)$$

onde  $\omega_1$  é a frequência,  $a_1$  é a amplitude e  $\phi_1$  é a fase da exponencial complexa, sendo todos parâmetros desconhecidos. Equivalentemente podemos escrever:

$$\mathbf{x} = \mathbf{s} + \mathbf{y} \quad (2.2)$$

onde  $\mathbf{s} = a_1 \exp[j\phi_1] \begin{bmatrix} 1 & \exp[j\omega_1] & \dots & \exp[j(N-1)\omega_1] \end{bmatrix}^T$ . Supondo conhecido o vetor  $\mathbf{s}$  de amostras unicamente do sinal, a função densidade de probabilidade (fdp) de um sinal em ruído branco gaussiano complexo,  $\mathbf{y} = \mathbf{x} - \mathbf{s}$ , de média nula como em (2.2), é dada por [42]:

$$p(\mathbf{x} - \mathbf{s}) = \frac{1}{\pi^N \det(\hat{\mathbf{R}}_{yy})} \exp \left[ -(\mathbf{x} - \mathbf{s})^H \hat{\mathbf{R}}_{yy}^{-1} (\mathbf{x} - \mathbf{s}) \right] \quad (2.3)$$

onde  $\hat{\mathbf{R}}_{yy}$  é a matriz de autocorrelação do ruído. Como o ruído é branco, sua matriz de autocorrelação pode ser expressa por:

$$\hat{\mathbf{R}}_{yy} = \sigma_y^2 \mathbf{I} \quad (2.4)$$

podemos escrever a fdp em (2.3) como [42]:

$$p(\mathbf{x} - \mathbf{s}) = \left( \frac{1}{\pi \sigma_y^2} \right)^N \exp \left[ -\frac{1}{\sigma_y^2} (\mathbf{x} - \mathbf{s})^H (\mathbf{x} - \mathbf{s}) \right] \quad (2.5)$$

Falando agora em estimação, nosso objetivo é estimar o vetor  $\mathbf{s}$  através de um vetor  $\hat{\mathbf{s}}$ , tal que o erro de estimação  $(\mathbf{x} - \hat{\mathbf{s}})$  seja mínimo. Então, como a fdp em (2.5) é gaussiana de média nula, minimizar o erro de estimação corresponde a maximizar esta fdp sobre  $\hat{a}_1$ ,  $\hat{\phi}_1$  e  $\hat{\omega}_1$ . Isto é realizado minimizando-se  $(\mathbf{x} - \hat{\mathbf{s}})^H (\mathbf{x} - \hat{\mathbf{s}})$  no expoente da exponencial em (2.5), o que é equivalente a minimizar a norma do erro de estimação:

$$\min(\mathbf{x} - \hat{\mathbf{s}})^H (\mathbf{x} - \hat{\mathbf{s}}) \quad (2.6)$$

Denominando  $\hat{a}_{c1} = \hat{a}_1 \exp[j\hat{\phi}_1]$  e  $\mathbf{e}_1 = [1 \ \exp[j\hat{\omega}_1] \ \cdots \ \exp[j(N-1)\hat{\omega}_1]]^T$ , podemos escrever:

$$\min(\mathbf{x} - \hat{a}_{c1} \mathbf{e}_1)^H (\mathbf{x} - \hat{a}_{c1} \mathbf{e}_1) \quad (2.7)$$

Supondo que conhecêssemos  $\hat{\omega}_1$ , a estimativa de máxima verossimilhança para a amplitude complexa poderia ser obtida almejando-se que:

$$\min(\mathbf{x} - \hat{a}_{c1} \mathbf{e}_1)^H (\mathbf{x} - \hat{a}_{c1} \mathbf{e}_1) = 0 \quad (2.8)$$

logo:

$$\begin{aligned} \mathbf{x} - \hat{a}_{c1} \mathbf{e}_1 &= \mathbf{0} \\ \hat{a}_{c1} \mathbf{e}_1 &= \mathbf{x} \end{aligned} \quad (2.9)$$

e, multiplicando por  $\mathbf{e}_1^H$ , obtemos:

$$\hat{a}_{c1} \mathbf{e}_1^H \mathbf{e}_1 = \mathbf{e}_1^H \mathbf{x} \quad (2.10)$$

Por fim, obtemos:

$$\hat{a}_{c1} = \frac{\mathbf{e}_1^H \mathbf{x}}{\mathbf{e}_1^H \mathbf{e}_1} \quad (2.11)$$

Substituindo a amplitude em (2.7) por seu valor em (2.11), a minimização sobre a frequência resulta em:

$$\begin{aligned}
\min(\mathbf{x} - \hat{a}_{c1}\mathbf{e}_1)^H(\mathbf{x} - \hat{a}_{c1}\mathbf{e}_1) &= \min \left[ \mathbf{x}^H (\mathbf{x} - \hat{a}_{c1}\mathbf{e}_1) - \hat{a}_{c1}^* \mathbf{e}_1^H (\mathbf{x} - \hat{a}_{c1}\mathbf{e}_1) \right] \\
&= \min \left[ \mathbf{x}^H (\mathbf{x} - \hat{a}_{c1}\mathbf{e}_1) - \hat{a}_{c1}^* (\mathbf{e}_1^H \mathbf{x} - \hat{a}_{c1} \mathbf{e}_1^H \mathbf{e}_1) \right] \\
&= \min \left[ \mathbf{x}^H (\mathbf{x} - \hat{a}_{c1}\mathbf{e}_1) - \hat{a}_{c1}^* \left( \mathbf{e}_1^H \mathbf{x} - \frac{\mathbf{e}_1^H \mathbf{x}}{\mathbf{e}_1^H \mathbf{e}_1} \mathbf{e}_1^H \mathbf{e}_1 \right) \right] \\
&= \min \left[ \mathbf{x}^H (\mathbf{x} - \hat{a}_{c1}\mathbf{e}_1) - \hat{a}_{c1}^* (\mathbf{e}_1^H \mathbf{x} - \mathbf{e}_1^H \mathbf{x}) \right] \\
&= \min \left[ \mathbf{x}^H (\mathbf{x} - \hat{a}_{c1}\mathbf{e}_1) \right] \\
&= \min \left[ \mathbf{x}^H \mathbf{x} - \hat{a}_{c1} \mathbf{x}^H \mathbf{e}_1 \right] \\
&= \min \left[ \mathbf{x}^H \mathbf{x} - \frac{\mathbf{e}_1^H \mathbf{x}}{\mathbf{e}_1^H \mathbf{e}_1} \mathbf{x}^H \mathbf{e}_1 \right] \\
&= \min \left[ \mathbf{x}^H \mathbf{x} - \frac{1}{N} |\mathbf{e}_1^H \mathbf{x}|^2 \right] \tag{2.12}
\end{aligned}$$

Portanto, para minimizar (2.12), devemos maximizar:

$$\frac{1}{N} |\mathbf{e}_1^H \mathbf{x}|^2 = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] \exp[-j\hat{\omega}_1 n] \right|^2 \tag{2.13}$$

sobre  $\hat{\omega}_1$ . Finalmente, como (2.13) é o periodograma de  $x[n]$ , dizemos que a estimativa de máxima verossimilhança para a frequência de uma exponencial complexa corrompida por ruído branco gaussiano complexo é dada pela frequência relativa ao valor máximo do periodograma do sinal.

## 2.2 Sinal com Múltiplas Exponenciais Complexas

Para obtermos as estimativas de máxima verossimilhança dos parâmetros de múltiplas exponenciais complexas corrompidas por ruído branco gaussiano complexo, procedemos da mesma forma que na seção anterior, exceto que agora devemos modelar o sinal como uma soma de  $M$  exponenciais complexas corrompidas por ruído branco aditivo gaussiano complexo:

$$x[n] = \sum_{i=1}^M a_i \exp[j(\omega_i n + \phi_i)] + y[n], \quad n = 0, \dots, N-1 \tag{2.14}$$

onde  $\omega_i$  são as frequências desconhecidas,  $a_i$  são as amplitudes e  $\phi_i$  as fases das exponenciais complexas. Equivalentemente podemos escrever:

$$\mathbf{x} = \mathbf{E}\mathbf{a}_c + \mathbf{y} \tag{2.15}$$

onde  $\mathbf{x}$  é um vetor  $N \times 1$  de dados,  $\mathbf{E}$  é a matriz  $N \times M$  de Vandermonde:

$$\mathbf{E} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \exp[j\omega_1] & \exp[j\omega_2] & \cdots & \exp[j\omega_M] \\ \exp[j2\omega_1] & \exp[j2\omega_2] & \cdots & \exp[j2\omega_M] \\ \vdots & \vdots & & \vdots \\ \exp[j(N-1)\omega_1] & \exp[j(N-1)\omega_2] & \cdots & \exp[j(N-1)\omega_M] \end{bmatrix} \quad (2.16)$$

$\mathbf{a}_c$  é um vetor  $M \times 1$  de amplitudes complexas:

$$\mathbf{a}_c = \left[ a_1 \exp[j\phi_1] \quad a_2 \exp[j\phi_2] \quad \cdots \quad a_M \exp[j\phi_M] \right] \quad (2.17)$$

e  $\mathbf{y}$  é um vetor  $N \times 1$  de ruído complexo com partes real e imaginária decorrelacionadas, cada uma com variância  $\sigma_y^2/2$ .

Levando em conta que o ruído é branco e gaussiano, as estimativas ML das frequências  $\omega_i$  do sinal e das amplitudes  $a_{ci}$  são dadas pela solução do seguinte problema de mínimos quadrados não-lineares:

$$\min_{\mathbf{E}, \mathbf{a}_c} \|\mathbf{x} - \mathbf{E}\mathbf{a}_c\|^2 = \min_{\mathbf{E}, \mathbf{a}_c} (\mathbf{x} - \mathbf{E}\mathbf{a}_c)^H (\mathbf{x} - \mathbf{E}\mathbf{a}_c) \quad (2.18)$$

Observando que, para uma matriz  $\mathbf{E}$  dada, encontrar o vetor  $\mathbf{a}_c$  ótimo é um problema de mínimos quadrados lineares com solução dada por [41]:

$$\hat{\mathbf{a}}_c = \mathbf{E}^\# \mathbf{x} = (\mathbf{E}^H \mathbf{E})^{-1} \mathbf{E}^H \mathbf{x} \quad (2.19)$$

onde " $\#$ " simboliza a pseudo inversa de Moore e Penrose [79], podemos substituir (2.19) em (2.18) e então obter:

$$\begin{aligned} \min_{\mathbf{E}} (\mathbf{x} - \mathbf{E}\hat{\mathbf{a}}_c)^H (\mathbf{x} - \mathbf{E}\hat{\mathbf{a}}_c) &= \min_{\mathbf{E}} \left[ \mathbf{x}^H (\mathbf{x} - \mathbf{E}\hat{\mathbf{a}}_c) - \hat{\mathbf{a}}_c^H \mathbf{E}^H (\mathbf{x} - \mathbf{E}\hat{\mathbf{a}}_c) \right] \\ &= \min_{\mathbf{E}} \left[ \mathbf{x}^H (\mathbf{x} - \mathbf{E}\hat{\mathbf{a}}_c) - \hat{\mathbf{a}}_c^H (\mathbf{E}^H \mathbf{x} - \mathbf{E}^H \mathbf{E}\hat{\mathbf{a}}_c) \right] \\ &= \min_{\mathbf{E}} \left[ \mathbf{x}^H (\mathbf{x} - \mathbf{E}\hat{\mathbf{a}}_c) - \hat{\mathbf{a}}_c^H (\mathbf{E}^H \mathbf{x} - \mathbf{E}^H \mathbf{E} (\mathbf{E}^H \mathbf{E})^{-1} \mathbf{E}^H \mathbf{x}) \right] \\ &= \min_{\mathbf{E}} \left[ \mathbf{x}^H (\mathbf{x} - \mathbf{E}\hat{\mathbf{a}}_c) - \hat{\mathbf{a}}_c^H (\mathbf{E}^H \mathbf{x} - \mathbf{E}^H \mathbf{x}) \right] \\ &= \min_{\mathbf{E}} \left[ \mathbf{x}^H (\mathbf{x} - \mathbf{E}\hat{\mathbf{a}}_c) \right] \\ &= \min_{\mathbf{E}} \left[ \mathbf{x}^H \mathbf{x} - \mathbf{x}^H \mathbf{E} \mathbf{E}^\# \mathbf{x} \right] \\ &= \min_{\mathbf{E}} \left[ \mathbf{x}^H \mathbf{x} - \mathbf{x}^H \mathbf{E} (\mathbf{E}^H \mathbf{E})^{-1} \mathbf{E}^H \mathbf{x} \right] \end{aligned} \quad (2.20)$$

Então, (2.18) pode ser escrita em duas formulações equivalentes:

$$\min_{\mathbf{E}} (\mathbf{x} - \mathbf{E}\hat{\mathbf{a}}_c)^H (\mathbf{x} - \mathbf{E}\hat{\mathbf{a}}_c) = \min_{\mathbf{E}} \left\{ \mathbf{x}^H \left[ \mathbf{I} - \mathbf{E} (\mathbf{E}^H \mathbf{E})^{-1} \mathbf{E}^H \right] \mathbf{x} \right\} \quad (2.21)$$

ou

$$\min_{\mathbf{E}} (\mathbf{x} - \mathbf{E}\hat{\mathbf{a}}_c)^H (\mathbf{x} - \mathbf{E}\hat{\mathbf{a}}_c) = \max_{\mathbf{E}} \left[ \mathbf{x}^H \mathbf{E} (\mathbf{E}^H \mathbf{E})^{-1} \mathbf{E}^H \mathbf{x} \right] \quad (2.22)$$

Entretanto, como (2.21) e (2.22) são altamente não-lineares em relação às frequências desconhecidas  $\{\omega_1, \dots, \omega_M\}$ , minimizar (2.21) ou maximizar (2.22) exigiria uma busca sobre um espaço  $M$ -dimensional. Esta busca é impraticável e computacionalmente muito pesada para valores grandes de  $M$ . Alternativamente, veremos no Capítulo 8 que ela pode ser aproximada numericamente e de forma iterativa, porém sem garantia de convergência para um mínimo global e sob um esforço computacional muitas vezes proibitivo. Nos capítulos seguintes, mostraremos como técnicas de Predição Linear podem ser empregadas para a estimação de parâmetros, requerendo esforço computacional muito inferior àquele exigido pela Máxima Verossimilhança.

No próximo capítulo definiremos uma expressão para o valor mínimo da variância do erro de estimação de um estimador não-polarizado, conhecido como limite de Cramér-Rao.

# Capítulo 3

## Limite de Cramér-Rao

Em um sistema de estimação, é importante conhecermos números que indiquem a melhor estimação possível de ser alcançada com os dados disponíveis, ou observações. Os erros médios quadráticos são importantes e freqüentemente utilizados como medida da imprecisão de um sistema. A polarização das estimativas, por sua vez, já é de importância secundária, embora geralmente seja desejável minimizá-la. Verificaremos neste capítulo que, para os nossos propósitos, a polarização pode geralmente ser desprezada. Então, interessa-nos saber qual é o valor mínimo possível da variância do erro de estimação. Neste capítulo definiremos este limite de desempenho, denominado como Limite de Cramér-Rao (CRB), para o caso de sinais senoidais em ruído branco. Este CRB foi derivado por Rife [84] e apresentado por Rife e Boorstin [85]. Salientamos, entretanto, que o CRB aplica-se somente a estimadores não-polarizados, de forma que um estimador cuja variância aproxima-se ou se iguala ao CRB, é dito ser ótimo.

### 3.1 Sinal com uma Única Exponencial Complexa

Inicialmente, escrevemos a função densidade de probabilidade de um sinal  $x[n]$  composto por  $N$  amostras de uma exponencial complexa corrompidas por ruído branco aditivo gaussiano complexo, de média zero e variância  $\sigma_y^2$ , quando a estimativa do vetor de parâmetros é  $\hat{\theta}$ . Sua expressão é dada por [42]:

$$p(\mathbf{x}; \hat{\theta}) = \left( \frac{1}{\pi\sigma_y^2} \right)^N \exp \left[ -\frac{1}{\sigma_y^2} (\mathbf{x} - \hat{\mathbf{s}})^H (\mathbf{x} - \hat{\mathbf{s}}) \right] \quad (3.1)$$

na qual  $\hat{s}[n] = \hat{a} \exp[j(\hat{\omega}n + \hat{\phi})]$ . Outra maneira mais conveniente de escrever (3.1) é:

$$p(\mathbf{x}; \hat{\theta}) = \left( \frac{1}{\pi\sigma_y^2} \right)^N \exp \left\{ -\frac{1}{\sigma_y^2} \sum_{n=0}^{N-1} [\text{Re}(x[n]) - \hat{\mu}[n]]^2 + [\text{Im}(x[n]) - \hat{\nu}[n]]^2 \right\} \quad (3.2)$$

onde, se  $\hat{\omega}$ ,  $\hat{a}$  e  $\hat{\phi}$  forem todos desconhecidos, temos que estimar:

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\omega} \\ \hat{a} \\ \hat{\phi} \end{bmatrix} \quad (3.3)$$

$$\hat{\mu}[n] = \hat{a} \cos(\hat{\omega}n + \hat{\phi}) \quad (3.4)$$

$$\hat{\nu}[n] = \hat{a} \sin(\hat{\omega}n + \hat{\phi}) \quad (3.5)$$

Os CRB's são dados pelos elementos diagonais da inversa da matriz  $\mathbf{J}$  de informação de Fisher [42], cujos elementos típicos são dados por:

$$\mathbf{J}_{ik} = E \{ \mathbf{H}_{\hat{\theta}_i} \mathbf{H}_{\hat{\theta}_k} \} = -E \{ \mathbf{H}_{\hat{\theta}_i \hat{\theta}_k} \} \quad (3.6)$$

onde a esperança diz respeito ao vetor de amostras  $\mathbf{x}$ , e  $\mathbf{H}_{\hat{\theta}_i}$  é dado por:

$$\mathbf{H}_{\hat{\theta}_i} = \frac{\partial}{\partial \hat{\theta}_i} \ln [p(\mathbf{x}; \hat{\boldsymbol{\theta}})] \quad (3.7)$$

ou seja,

$$\begin{aligned} \mathbf{H}_{\hat{\theta}_i} &= \frac{\partial}{\partial \hat{\theta}_i} \ln \left[ \left( \frac{1}{\pi \sigma_y^2} \right)^N \exp \left\{ -\frac{1}{\sigma_y^2} \sum_{n=0}^{N-1} [\operatorname{Re}(x[n]) - \hat{\mu}[n]]^2 + [\operatorname{Im}(x[n]) - \hat{\nu}[n]]^2 \right\} \right] \\ &= -\frac{1}{\sigma_y^2} \sum_{n=0}^{N-1} \left\{ \frac{\partial}{\partial \hat{\theta}_i} [\operatorname{Re}(x[n]) - \hat{\mu}[n]]^2 + \frac{\partial}{\partial \hat{\theta}_i} [\operatorname{Im}(x[n]) - \hat{\nu}[n]]^2 \right\} \\ &= \frac{2}{\sigma_y^2} \sum_{n=0}^{N-1} \left\{ [\operatorname{Re}(x[n]) - \hat{\mu}[n]] \frac{\partial \hat{\mu}[n]}{\partial \hat{\theta}_i} + [\operatorname{Im}(x[n]) - \hat{\nu}[n]] \frac{\partial \hat{\nu}[n]}{\partial \hat{\theta}_i} \right\} \end{aligned} \quad (3.8)$$

Dai, temos que:

$$\begin{aligned} \mathbf{J}_{ik} &= E \{ \mathbf{H}_{\hat{\theta}_i} \mathbf{H}_{\hat{\theta}_k} \} \\ &= \frac{4}{\sigma_y^4} \sum_{n=0}^{N-1} E \left\{ [\operatorname{Re}(x[n]) - \hat{\mu}[n]]^2 \right\} \frac{\partial \hat{\mu}[n]}{\partial \hat{\theta}_i} \frac{\partial \hat{\mu}[n]}{\partial \hat{\theta}_k} \\ &\quad + E \left\{ [\operatorname{Im}(x[n]) - \hat{\nu}[n]]^2 \right\} \frac{\partial \hat{\nu}[n]}{\partial \hat{\theta}_k} \frac{\partial \hat{\nu}[n]}{\partial \hat{\theta}_i} \\ &\quad + E \left\{ [\operatorname{Re}(x[n]) - \hat{\mu}[n]] [\operatorname{Im}(x[n]) - \hat{\nu}[n]] \right\} \frac{\partial \hat{\mu}[n]}{\partial \hat{\theta}_i} \frac{\partial \hat{\nu}[n]}{\partial \hat{\theta}_k} \\ &\quad + E \left\{ [\operatorname{Re}(x[n]) - \hat{\mu}[n]] [\operatorname{Im}(x[n]) - \hat{\nu}[n]] \right\} \frac{\partial \hat{\mu}[n]}{\partial \hat{\theta}_k} \frac{\partial \hat{\nu}[n]}{\partial \hat{\theta}_i} \end{aligned} \quad (3.9)$$

Como as partes real e imaginária do ruído são decorrelacionadas entre si e as variâncias das partes real e imaginária são iguais a  $\sigma_y^2/2$ , temos:

$$\mathbf{J}_{ik} = \frac{2}{\sigma_y^2} \sum_{n=0}^{N-1} \left( \frac{\partial \hat{\mu}[n]}{\partial \hat{\theta}_i} \frac{\partial \hat{\mu}[n]}{\partial \hat{\theta}_k} + \frac{\partial \hat{\nu}[n]}{\partial \hat{\theta}_i} \frac{\partial \hat{\nu}[n]}{\partial \hat{\theta}_k} \right) \quad (3.10)$$

Agora, calculamos os valores de cada uma dessas derivadas, obtendo:

$$\begin{aligned} \frac{\partial \hat{\mu}[n]}{\partial \hat{\theta}_1} &= -\hat{a} \operatorname{sen}(\hat{\omega}n + \hat{\phi})n & \frac{\partial \hat{\nu}[n]}{\partial \hat{\theta}_1} &= \hat{a} \cos(\hat{\omega}n + \hat{\phi})n \\ \frac{\partial \hat{\mu}[n]}{\partial \hat{\theta}_2} &= \cos(\hat{\omega}n + \hat{\phi}) & \frac{\partial \hat{\nu}[n]}{\partial \hat{\theta}_2} &= \operatorname{sen}(\hat{\omega}n + \hat{\phi}) \\ \frac{\partial \hat{\mu}[n]}{\partial \hat{\theta}_3} &= -\hat{a} \operatorname{sen}(\hat{\omega}n + \hat{\phi}) & \frac{\partial \hat{\nu}[n]}{\partial \hat{\theta}_3} &= \hat{a} \cos(\hat{\omega}n + \hat{\phi}) \end{aligned} \quad (3.11)$$

Desse modo, os elementos da matriz  $\mathbf{J}$  são expressos por:

$$\begin{aligned} \mathbf{J}_{11} &= \frac{2}{\sigma_y^2} \sum_{n=0}^{N-1} \hat{a}^2 n^2 [\operatorname{sen}^2(\hat{\omega}n + \hat{\phi}) + \cos^2(\hat{\omega}n + \hat{\phi})] \\ &= \frac{2\hat{a}^2}{\sigma_y^2} \sum_{n=0}^{N-1} n^2 = \frac{\hat{a}^2 N(N-1)(2N-1)}{3\sigma_y^2} \end{aligned} \quad (3.12)$$

$$\begin{aligned} \mathbf{J}_{12} &= \frac{2}{\sigma_y^2} \sum_{n=0}^{N-1} -\hat{a} n \operatorname{sen}(\hat{\omega}n + \hat{\phi}) \cos(\hat{\omega}n + \hat{\phi}) + \hat{a} n \cos(\hat{\omega}n + \hat{\phi}) \operatorname{sen}(\hat{\omega}n + \hat{\phi}) \\ &= 0 \end{aligned} \quad (3.13)$$

$$\begin{aligned} \mathbf{J}_{13} &= \frac{2}{\sigma_y^2} \sum_{n=0}^{N-1} \hat{a}^2 n [\operatorname{sen}^2(\hat{\omega}n + \hat{\phi}) + \cos^2(\hat{\omega}n + \hat{\phi})] \\ &= \frac{2\hat{a}^2}{\sigma_y^2} \sum_{n=0}^{N-1} n = \frac{\hat{a}^2 N(N-1)}{\sigma_y^2} \end{aligned} \quad (3.14)$$

$$\begin{aligned} \mathbf{J}_{22} &= \frac{2}{\sigma_y^2} \sum_{n=0}^{N-1} [\operatorname{sen}^2(\hat{\omega}n + \hat{\phi}) + \cos^2(\hat{\omega}n + \hat{\phi})] \\ &= \frac{2}{\sigma_y^2} \sum_{n=0}^{N-1} 1 = \frac{2N}{\sigma_y^2} \end{aligned} \quad (3.15)$$

$$\begin{aligned} \mathbf{J}_{23} &= \frac{2}{\sigma_y^2} \sum_{n=0}^{N-1} -\hat{a} \cos(\hat{\omega}n + \hat{\phi}) \operatorname{sen}(\hat{\omega}n + \hat{\phi}) + \hat{a} \operatorname{sen}(\hat{\omega}n + \hat{\phi}) \cos(\hat{\omega}n + \hat{\phi}) \\ &= 0 \end{aligned} \quad (3.16)$$

$$\begin{aligned} \mathbf{J}_{33} &= \frac{2}{\sigma_y^2} \sum_{n=0}^{N-1} \hat{a}^2 [\operatorname{sen}^2(\hat{\omega}n + \hat{\phi}) + \cos^2(\hat{\omega}n + \hat{\phi})] \\ &= \frac{2\hat{a}^2}{\sigma_y^2} N \end{aligned} \quad (3.17)$$

A matriz  $\mathbf{J}$  de informação de Fischer é então escrita como:

$$\mathbf{J} = \begin{bmatrix} \frac{\hat{a}^2 N(N-1)(2N-1)}{3\sigma_y^2} & 0 & \frac{\hat{a}^2 N(N-1)}{\sigma_y^2} \\ 0 & \frac{2N}{\sigma_y^2} & 0 \\ \frac{\hat{a}^2 N(N-1)}{\sigma_y^2} & 0 & \frac{2\hat{a}^2 N}{\sigma_y^2} \end{bmatrix} \quad (3.18)$$

e sua inversa é dada por:

$$\mathbf{J}^{-1} = \begin{bmatrix} \frac{6\sigma_y^2}{\hat{a}^2 N(N^2-1)} & 0 & -\frac{3\sigma_y^2}{\hat{a}^2 N} \\ 0 & \frac{\sigma_y^2}{2N} & 0 \\ -\frac{3\sigma_y^2}{\hat{a}^2 N} & 0 & \frac{(2N-1)\sigma_y^2}{\hat{a}^2 N(N+1)} \end{bmatrix} \quad (3.19)$$

Logo, concluímos que para o caso de uma única exponencial complexa corrompida por ruído branco gaussiano, os limites de Cramér-Rao para os parâmetros são dados por:

$$\text{var} \{ \hat{\omega} \} \geq \frac{6\sigma_y^2}{\hat{a}^2 N (N^2 - 1)} \quad (3.20)$$

$$\text{var} \{ \hat{a} \} \geq \frac{\sigma_y^2}{2N} \quad (3.21)$$

$$\text{var} \{ \hat{\phi} \} \geq \frac{(2N-1)\sigma_y^2}{\hat{a}^2 N (N+1)} \quad (3.22)$$

Como esperado, o CRB para a estimação da frequência é inversamente proporcional à relação sinal-ruído (SNR), ou seja:

$$\text{var} \{ \hat{\omega} \} \geq \frac{6}{N(N^2-1)SNR} \quad (3.23)$$

onde:

$$SNR = \frac{\hat{a}^2}{\sigma_y^2}$$

Podemos observar também que o CRB para as frequências diminui cubicamente com o comprimento do registro de dados, indicando a importância do número de amostras disponível para a estimação.

## 3.2 Sinal com Múltiplas Exponenciais Complexas

Agora discutiremos a situação na qual o sinal é composto por múltiplas, digamos  $M$ , exponenciais complexas imersas em ruído. Neste caso, a função densidade de probabilidade desse sinal é dada por [42]:

$$p(\mathbf{x}; \hat{\boldsymbol{\theta}}) = \frac{1}{\pi^N \det(\mathbf{R}_{yy})} \exp \left[ -(\mathbf{x} - \hat{\mathbf{s}})^H \mathbf{R}_{yy}^{-1} (\mathbf{x} - \hat{\mathbf{s}}) \right] \quad (3.24)$$

onde  $\mathbf{R}_{yy}$  é a matriz de autocorrelação do ruído. Considerando que o ruído é branco, temos que  $\mathbf{R}_{yy} = \sigma_y^2 \mathbf{I}$ . Assim como no caso de uma única exponencial complexa, uma forma mais conveniente de escrever (3.24) é:

$$p(\mathbf{x}; \hat{\boldsymbol{\theta}}) = \left( \frac{1}{\pi \sigma_y^2} \right)^N \exp \left\{ -\frac{1}{\sigma_y^2} \sum_{n=0}^{N-1} [\operatorname{Re}(x[n]) - \hat{\mu}[n]]^2 + [\operatorname{Im}(x[n]) - \hat{\nu}[n]]^2 \right\} \quad (3.25)$$

onde, se  $\hat{\omega}_i$ ,  $\hat{a}_i$  e  $\hat{\phi}_i$ ,  $i = 1, \dots, M$ , forem todos desconhecidos, devemos estimar:

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\omega}_1 \\ \hat{a}_1 \\ \hat{\phi}_1 \\ \vdots \\ \hat{\omega}_M \\ \hat{a}_M \\ \hat{\phi}_M \end{bmatrix} \quad (3.26)$$

$$\hat{\mu}[n] = \sum_{i=1}^M \hat{a}_i \cos(\hat{\omega}_i n + \hat{\phi}_i) \quad (3.27)$$

$$\hat{\nu}[n] = \sum_{i=1}^M \hat{a}_i \operatorname{sen}(\hat{\omega}_i n + \hat{\phi}_i) \quad (3.28)$$

Assim como no caso de uma única exponencial, os CRB's são dados pelos elementos diagonais da inversa da matriz  $\mathbf{J}$  de informação de Fisher, cujos elementos são dados por:

$$\mathbf{J}_{fg} = \frac{2}{\sigma_y^2} \sum_{n=0}^{N-1} \left( \frac{\partial \hat{\mu}[n]}{\partial \hat{\theta}_f} \frac{\partial \hat{\mu}[n]}{\partial \hat{\theta}_g} + \frac{\partial \hat{\nu}[n]}{\partial \hat{\theta}_f} \frac{\partial \hat{\nu}[n]}{\partial \hat{\theta}_g} \right), \quad \text{para } f, g = 1, \dots, 3M \quad (3.29)$$

onde  $\hat{\theta}_{3i-2} = \hat{\omega}_i$ ,  $\hat{\theta}_{3i-1} = \hat{a}_i$  e  $\hat{\theta}_{3i} = \hat{\phi}_i$ ,  $1 \leq i \leq M$ . Como as derivadas  $\partial/\partial \hat{\theta}_{3i-2}$ ,  $\partial/\partial \hat{\theta}_{3i-1}$  e  $\partial/\partial \hat{\theta}_{3i}$  somente se aplicam às  $i$ -ésimas parcelas de  $\hat{\mu}[n]$  e  $\hat{\nu}[n]$ , temos que os elementos da matriz  $\mathbf{J}$  são dados por:

$$\mathbf{J}_{3i-2, 3k-2} = \frac{2\hat{a}_i \hat{a}_k}{\sigma_y^2} \sum_{n=0}^{N-1} n^2 \cos[(\hat{\omega}_i - \hat{\omega}_k)n + (\hat{\phi}_i - \hat{\phi}_k)] \quad (3.30)$$

$$\mathbf{J}_{3i-2, 3k-1} = -\frac{2\hat{a}_i \hat{a}_k}{\sigma_y^2} \sum_{n=0}^{N-1} n \operatorname{sen}[(\hat{\omega}_i - \hat{\omega}_k)n + (\hat{\phi}_i - \hat{\phi}_k)] \quad (3.31)$$

$$\mathbf{J}_{3i-2, 3k} = \frac{2\hat{a}_i \hat{a}_k}{\sigma_y^2} \sum_{n=0}^{N-1} n \cos[(\hat{\omega}_i - \hat{\omega}_k)n + (\hat{\phi}_i - \hat{\phi}_k)] \quad (3.32)$$

$$\mathbf{J}_{3i-1,3k-2} = \frac{2\hat{a}_k}{\sigma_y^2} \sum_{n=0}^{N-1} n \text{sen} \left[ (\hat{\omega}_i - \hat{\omega}_k) n + (\hat{\phi}_i - \hat{\phi}_k) \right] \quad (3.33)$$

$$\mathbf{J}_{3i-1,3k-1} = \frac{2}{\sigma_y^2} \sum_{n=0}^{N-1} \cos \left[ (\hat{\omega}_i + \hat{\omega}_k) n + (\hat{\phi}_i + \hat{\phi}_k) \right] \quad (3.34)$$

$$\mathbf{J}_{3i-1,3k} = \frac{2\hat{a}_k}{\sigma_y^2} \sum_{n=0}^{N-1} \text{sen} \left[ (\hat{\omega}_i - \hat{\omega}_k) n + (\hat{\phi}_i - \hat{\phi}_k) \right] \quad (3.35)$$

$$\mathbf{J}_{3i,3k-2} = \frac{2\hat{a}_i\hat{a}_k}{\sigma_y^2} \sum_{n=0}^{N-1} n \cos \left[ (\hat{\omega}_i - \hat{\omega}_k) n + (\hat{\phi}_i - \hat{\phi}_k) \right] \quad (3.36)$$

$$\mathbf{J}_{3i,3k-1} = -\frac{2\hat{a}_i}{\sigma_y^2} \sum_{n=0}^{N-1} \text{sen} \left[ (\hat{\omega}_i - \hat{\omega}_k) n + (\hat{\phi}_i - \hat{\phi}_k) \right] \quad (3.37)$$

$$\mathbf{J}_{3i,3k} = \frac{2\hat{a}_i\hat{a}_k}{\sigma_y^2} \sum_{n=0}^{N-1} \cos \left[ (\hat{\omega}_i - \hat{\omega}_k) n + (\hat{\phi}_i - \hat{\phi}_k) \right] \quad (3.38)$$

para  $i, k = 1, \dots, M$ . Então, a matriz  $\mathbf{J}$  pode ser particionada como:

$$\mathbf{J} = \begin{bmatrix} \mathbf{Z}_{11} & \cdots & \mathbf{Z}_{1M} \\ \vdots & \ddots & \vdots \\ \mathbf{Z}_{M1} & \cdots & \mathbf{Z}_{MM} \end{bmatrix} \quad (3.39)$$

tal que as submatrizes  $\mathbf{Z}_{ik}$  têm dimensões  $3 \times 3$  e são dadas por:

$$\mathbf{Z}_{ik} = \begin{bmatrix} \frac{2\hat{a}_i\hat{a}_k}{\sigma_y^2} \sum_{n=0}^{N-1} n^2 \cos \Delta_{ik}(n) & -\frac{2\hat{a}_i}{\sigma_y^2} \sum_{n=0}^{N-1} n \text{sen} \Delta_{ik}(n) & \frac{2\hat{a}_i\hat{a}_k}{\sigma_y^2} \sum_{n=0}^{N-1} n \cos \Delta_{ik}(n) \\ \frac{2\hat{a}_k}{\sigma_y^2} \sum_{n=0}^{N-1} n \text{sen} \Delta_{ik}(n) & \frac{2}{\sigma_y^2} \sum_{n=0}^{N-1} \cos \Delta_{ik}(n) & \frac{2\hat{a}_k}{\sigma_y^2} \sum_{n=0}^{N-1} \text{sen} \Delta_{ik}(n) \\ \frac{2\hat{a}_i\hat{a}_k}{\sigma_y^2} \sum_{n=0}^{N-1} n \cos \Delta_{ik}(n) & -\frac{2\hat{a}_i}{\sigma_y^2} \sum_{n=0}^{N-1} \text{sen} \Delta_{ik}(n) & \frac{2\hat{a}_i\hat{a}_k}{\sigma_y^2} \sum_{n=0}^{N-1} \cos \Delta_{ik}(n) \end{bmatrix} \quad (3.40)$$

onde  $\Delta_{ik}(n) = (\omega_i - \omega_k)n + (\phi_i - \phi_k)$ . Assim,  $\mathbf{Z}_{ik}$  pode ser decomposta como:

$$\begin{aligned} \mathbf{Z}_{ik} &= \frac{2}{\sigma_y^2} \begin{bmatrix} \hat{a}_i & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \hat{a}_i \end{bmatrix} \times \\ &\begin{bmatrix} \sum_{n=0}^{N-1} n^2 \cos \Delta_{ik}(n) & -\sum_{n=0}^{N-1} n \text{sen} \Delta_{ik}(n) & \sum_{n=0}^{N-1} n \cos \Delta_{ik}(n) \\ \sum_{n=0}^{N-1} n \text{sen} \Delta_{ik}(n) & \sum_{n=0}^{N-1} \cos \Delta_{ik}(n) & \sum_{n=0}^{N-1} \text{sen} \Delta_{ik}(n) \\ \sum_{n=0}^{N-1} n \cos \Delta_{ik}(n) & -\sum_{n=0}^{N-1} \text{sen} \Delta_{ik}(n) & \sum_{n=0}^{N-1} \cos \Delta_{ik}(n) \end{bmatrix} \begin{bmatrix} \hat{a}_k & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \hat{a}_k \end{bmatrix} \\ &= \frac{2}{\sigma_y^2} \mathbf{D}_i \mathbf{P}_{ik} \mathbf{D}_k \end{aligned} \quad (3.41)$$

Daí, a matriz  $\mathbf{J}$  de informação de Fischer fica expressa por:

$$\begin{aligned}
 \mathbf{J} &= \frac{2}{\sigma_y^2} \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_M \end{bmatrix} \begin{bmatrix} \mathbf{P}_{11} & \cdots & \mathbf{P}_{1M} \\ \vdots & \ddots & \vdots \\ \mathbf{P}_{M1} & \cdots & \mathbf{P}_{MM} \end{bmatrix} \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_M \end{bmatrix} \\
 &= \frac{2}{\sigma_y^2} \mathbf{DPD}
 \end{aligned} \tag{3.42}$$

Dessa forma, sua inversa pode ser escrita como:

$$\mathbf{J}^{-1} = \frac{\sigma_y^2}{2} \mathbf{D}^{-1} \mathbf{P}^{-1} \mathbf{D}^{-1} \tag{3.43}$$

O caso de uma única exponencial complexa ocorre para  $M = 1$  e  $\Delta_{11}(n) = 0$ . Notemos também que se  $\mathbf{P}_{ik} = \mathbf{0}$  para  $i \neq k$ , tal que a matriz  $\mathbf{P}$  seja diagonal em blocos, então os CRB's serão iguais àqueles do caso de uma única exponencial. Este desacoplamento vale aproximadamente se os espaçamentos  $\Delta\omega_{ik}$  entre as frequências satisfizerem  $|\Delta\omega_{ik}| \gg 2\pi/N$ ,  $\forall i, k$ , e  $N \gg 2\pi$ . Neste caso,  $\mathbf{P}_{ik} \approx \mathbf{0}$  para  $i \neq k$  pois, para  $m = 0, 1, 2$ , temos que:

$$\begin{aligned}
 \frac{1}{N} \sum_{n=0}^{N-1} n^m \exp[j\Delta_{ik}(n)] &= \frac{1}{N} \sum_{n=0}^{N-1} n^m \exp(j\Delta\omega_{ik}n + j\Delta\phi_{ik}) \\
 &= j^{-m} \exp(j\Delta\phi_{ik}) \left[ \frac{\partial^m}{\partial \Delta\omega^m} \frac{1}{N} \sum_{n=0}^{N-1} \exp(j\Delta\omega_{ik}n) \right]
 \end{aligned} \tag{3.44}$$

O termo entre colchetes será aproximadamente zero para  $|\Delta\omega_{ik}| \gg 2\pi/N$ . Para este caso especial, a extração de picos do periodograma resultará na estimativa ML [41] e este estimador acompanha o limite de Cramér-Rao.

Para apenas duas exponenciais complexas, a matriz  $\mathbf{J}$  é dada por:

$$\begin{aligned}
 \mathbf{J} &= \frac{2}{\sigma_y^2} \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{bmatrix} \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{bmatrix} \\
 &= \frac{2}{\sigma_y^2} \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{bmatrix} \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{12}^T & \mathbf{P}_{11} \end{bmatrix} \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{bmatrix}
 \end{aligned} \tag{3.45}$$

e sua inversa terá a forma:

$$\mathbf{J}^{-1} = \frac{\sigma_y^2}{2} \begin{bmatrix} \mathbf{D}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{Q} & \mathbf{K} \\ \mathbf{K}^T & \mathbf{T} \end{bmatrix} \begin{bmatrix} \mathbf{D}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2^{-1} \end{bmatrix} \tag{3.46}$$

na qual:

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{Q} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_{11} & -\mathbf{Q}_{12} & \mathbf{Q}_{13} \\ -\mathbf{Q}_{21} & \mathbf{Q}_{22} & -\mathbf{Q}_{23} \\ \mathbf{Q}_{31} & -\mathbf{Q}_{32} & \mathbf{Q}_{33} \end{bmatrix} \tag{3.47}$$

tal que  $\mathbf{T}_{ii} = \mathbf{Q}_{ii}$ . Além disso, a partir de  $\mathbf{J}\mathbf{J}^{-1} = \mathbf{I}$ , podemos afirmar que:

$$\mathbf{Q} = [\mathbf{P}_{11} - \mathbf{P}_{12}\mathbf{P}_{11}^{-1}\mathbf{P}_{12}^T]^{-1} \quad (3.48)$$

Desse modo, os CRB's para as frequências de duas exponenciais complexas são dados por:

$$\text{var}(\hat{\omega}_1) \geq \frac{\sigma_y^2}{2a_1^2} \mathbf{Q}_{11} \quad (3.49)$$

$$\text{var}(\hat{\omega}_2) \geq \frac{\sigma_y^2}{2a_2^2} \mathbf{Q}_{11} \quad (3.50)$$

Podemos observar que, mesmo para o caso simples de apenas duas exponenciais complexas, a obtenção de uma expressão analítica explícita para os CRB's demanda muito esforço. Normalmente, os CRB's são calculados somente numericamente. As expressões (3.49) e (3.50) permitem-nos observar que, quando as amplitudes das exponenciais são iguais, seus CRB's também são iguais. Além disso, os CRB's não dependem dos valores individuais das frequências e fases, mas sim de suas diferenças expressas em  $\Delta_{12}(n)$ . Também verificamos que, para frequências muito afastadas, os CRB's confundem-se com aquele referente ao caso de uma única exponencial. Por fim, conforme [85], podemos afirmar que os CRB's para duas exponenciais complexas são sempre maiores ou iguais ao CRB para uma única exponencial complexa.

## Capítulo 4

# Estimação de Frequências por Predição Linear

Um problema especialmente interessante em processamento de sinais é o de predição, onde se utiliza um conjunto de amostras de um processo estacionário para prever o valor que será assumido pelo mesmo em algum instante futuro ou, genericamente, para extrapolar os valores do processo fora do intervalo de observação. A filtragem de Wiener aplica-se a este problema, utilizando a autocorrelação para conhecer o comportamento estatístico do sinal. Entretanto, como dispomos apenas de um conjunto finito de amostras do sinal, não conhecemos completamente a sua autocorrelação. Logo, utilizamos estimativas da mesma. Para evitar distorções na interpretação do sinal pelo filtro, restringiremos a predição ao intervalo de tempo em que o filtro encontrar-se totalmente preenchido. Porém, conforme a ordem do filtro e o número de amostras disponíveis, isto pode dar origem a um sistema com quantidades diferentes de incógnitas e de equações. Assim, efetuaremos a otimização de um filtro de erro de predição linear segundo o critério dos mínimos quadrados, desenvolvido no contexto de predição *forward-backward*. O filtro ótimo resultante desta minimização aplica-se especificamente à estimação de sinais senoidais não amortecidos. As frequências dos sinais senoidais podem, então, ser estimadas a partir das raízes da função de transferência do filtro otimizado ou a partir dos zeros do módulo de sua resposta em frequência. Encerrando, apresentaremos gráficos que permitirão avaliar o desempenho desta abordagem para a estimação de frequências, comparando o seu desempenho ao dos métodos baseados em Máxima Verossimilhança.

### 4.1 Predição Linear Forward-Backward - FBLP

Seja um conjunto de  $N$  amostras de um sinal com valores complexos  $x[0], x[1], \dots, x[N-1]$ , observados respectivamente nos instantes  $t_0, t_1, \dots, t_{(N-1)}$ , uniformemente espaçados. Aplica-se este sinal aos filtros de erro de predição *forward* e *backward* de ordem  $L < N$ , apresentados nas figuras 4.1 e 4.2, onde  $(*)$  simboliza conjugado complexo.

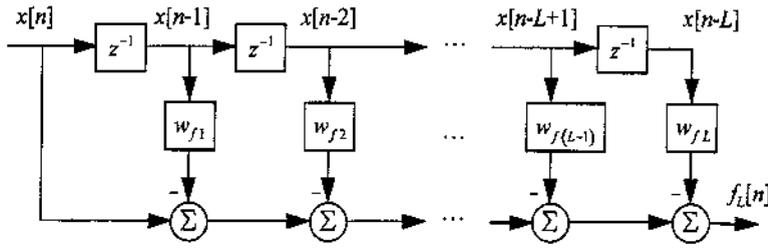


Figura 4.1: Filtro de erro de predição forward.

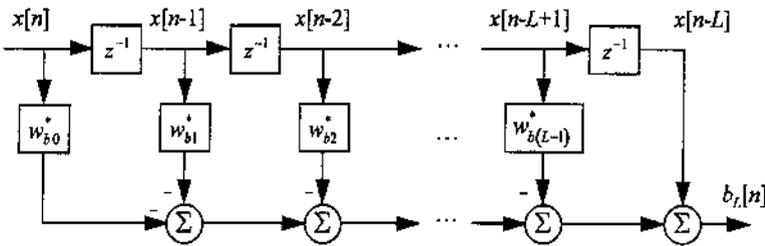


Figura 4.2: Filtro de erro de predição backward.

Estes filtros tentam prever uma amostra do sinal em função de uma combinação linear das demais amostras presentes no filtro. Daí o nome *Predição Linear*, sendo os filtros denominados Filtros de Erro de Predição Linear (FEP), pois  $f_L[n]$  e  $b_L[n]$  representam estes erros.

O filtro *forward* procura prever a amostra de entrada  $x[n]$  a partir das amostras anteriores do sinal, enquanto que o filtro *backward* tenta estimar a amostra de saída  $x[n-L]$  a partir do conhecimento das amostras posteriores do sinal.

Os erros de predição *forward* e *backward* são obtidos por:

$$f_L[n] = x[n] - \sum_{k=1}^L w_{fk} \cdot x[n-k] \quad (4.1)$$

e

$$b_L[n] = x[n-L] - \sum_{k=0}^{L-1} w_{bk}^* \cdot x[n-k] \quad (4.2)$$

para  $n = L, \dots, N-1$ .

Os vetores de coeficientes dos filtros *forward*  $((L+1) \times 1)$  e *backward*  $((L+1) \times 1)$  são definidos como:

$$\mathbf{w}_f = \begin{bmatrix} -1 \\ w_{f1} \\ w_{f2} \\ \vdots \\ w_{fL} \end{bmatrix} \quad \text{e} \quad \mathbf{w}_b = \begin{bmatrix} -1 \\ w_{b(L-1)} \\ w_{b(L-2)} \\ \vdots \\ w_{b0} \end{bmatrix} \quad (4.3)$$

dos quais extraímos seus respectivos vetores ( $L \times 1$ ) de coeficientes a serem determinados:

$$\mathbf{w}'_f = \begin{bmatrix} w_{f1} \\ w_{f2} \\ \vdots \\ w_{fL} \end{bmatrix} \quad \text{e} \quad \mathbf{w}'_b = \begin{bmatrix} w_{b(L-1)} \\ w_{b(L-2)} \\ \vdots \\ w_{b0} \end{bmatrix} \quad (4.4)$$

Nas aplicações em análise espectral e estimação de frequências, o filtro *forward* é projetado minimizando-se a energia do erro de predição *forward* produzido enquanto o filtro encontra-se totalmente preenchido por amostras de sinal:

$$\varepsilon(\mathbf{w}'_f) = \sum_{n=L}^{N-1} |f_L[n]|^2 \quad (4.5)$$

De forma similar, o filtro *backward* é projetado minimizando-se a energia do erro de predição *backward*:

$$\varepsilon(\mathbf{w}'_b) = \sum_{n=L}^{N-1} |b_L[n]|^2 \quad (4.6)$$

Combinando-se os métodos de predição linear *forward* e *backward*, obtém-se um terceiro método que utiliza melhor a informação contida nas amostras de sinal disponíveis, denominado *forward-backward* (FBLP). Nele, o critério de otimização consiste na minimização da soma das energias dos erros de predição *forward* e *backward* através do ajuste dos coeficientes dos filtros:

$$\varepsilon_{fb}(\mathbf{w}'_{fb}) = \varepsilon_f(\mathbf{w}'_f) + \varepsilon_b(\mathbf{w}'_b) \quad (4.7)$$

$$\varepsilon_{fb}(\mathbf{w}'_{fb}) = \sum_{k=L}^{N-1} [|f_L[k]|^2 + |b_L[k]|^2] \quad (4.8)$$

onde  $\mathbf{w}'_{fb}$  é o vetor de coeficientes *forward-backward* ( $L \times 1$ ) a serem determinados.

Embora as soluções  $\mathbf{w}'_f$  e  $\mathbf{w}'_b$  sejam normalmente distintas quando  $N$  é finito, no método FBLP a igualdade é assumida como restrição para a otimização. Se não impusermos a igualdade entre  $\mathbf{w}'_f$  e  $\mathbf{w}'_b$  para  $N$  finito, obteremos erros com características distintas, pois cada filtro trata com um conjunto distinto de amostras na geração do vetor de erro. Assim, teríamos vetores de coeficientes ótimos diferentes. Isto levaria a questionarmos qual deles deveria ser utilizado, uma vez que são resultantes da mesma minimização. Para evitar tal ambigüidade é que se assume a igualdade dos  $\mathbf{w}'$ . A otimização conjunta dos filtros com a restrição  $\mathbf{w}'_f = \mathbf{w}'_b$  permite atingir um único vetor ótimo e fornece, em geral, melhores resultados na análise espectral e estimação de frequências [3][93][72].

Adicionalmente a essas justificativas, a restrição que impusemos é satisfeita assintoticamente com  $N$ , justificando a sua utilização. Isto é demonstrado pelos argumentos a seguir.

Analisando as equações normais *forward* e *backward*, verificamos que [3]:

a) No caso *forward*:

$$\mathbf{A}_f^H \mathbf{A}_f \cdot \mathbf{w}'_f = \mathbf{A}_f^H \cdot \mathbf{b}_f \quad (4.9)$$

ou

$$\hat{\mathbf{R}}_f \cdot \mathbf{w}'_f = \hat{\mathbf{r}}_f \quad (4.10)$$

b) no caso *backward* :

$$\mathbf{A}_b^H \mathbf{A}_b \cdot \mathbf{w}'_b = \mathbf{A}_b^H \cdot \mathbf{b}_b \quad (4.11)$$

ou

$$\hat{\mathbf{R}}_b \cdot \mathbf{w}'_b = \hat{\mathbf{r}}_b \quad (4.12)$$

onde:

- (  $\hat{\quad}$  ) indica estimação polarizada [32] e (  $^H$  ) conjugado transposto.
- $\mathbf{A}_f$  e  $\mathbf{A}_b$  são, respectivamente, as matrizes de dados *forward* e *backward* ( $(N - L) \times L$ ).

$$\mathbf{A}_f = \begin{bmatrix} x[L-1] & x[L-2] & \cdots & x[0] \\ x[L] & x[L-1] & \cdots & x[1] \\ \vdots & \vdots & \ddots & \vdots \\ x[N-2] & x[N-3] & \cdots & x[N-L-1] \end{bmatrix} \quad (4.13)$$

$$\mathbf{A}_b = \begin{bmatrix} x^*[1] & x^*[2] & \cdots & x^*[L] \\ x^*[2] & x^*[3] & \cdots & x^*[L+1] \\ \vdots & \vdots & \ddots & \vdots \\ x^*[N-L] & x^*[N-L+1] & \cdots & x^*[N-1] \end{bmatrix} \quad (4.14)$$

- $\mathbf{b}_f$  e  $\mathbf{b}_b$  são, respectivamente, os vetores desejados *forward* e *backward* ( $(N - L) \times 1$ ).

$$\mathbf{b}_f = \begin{bmatrix} x[L] \\ x[L+1] \\ \vdots \\ x[N-1] \end{bmatrix} \quad \text{e} \quad \mathbf{b}_b = \begin{bmatrix} x^*[0] \\ x^*[1] \\ \vdots \\ x^*[N-L-1] \end{bmatrix} \quad (4.15)$$

- $\hat{\mathbf{R}}_f$  e  $\hat{\mathbf{R}}_b$  são, respectivamente, as matrizes de correlação estimadas *forward* e *backward*.
- $\hat{\mathbf{r}}_f$  e  $\hat{\mathbf{r}}_b$  são, respectivamente, os vetores de correlação estimados *forward* e *backward*.

Admitindo ergodicidade, quando  $N \rightarrow \infty$  obtemos [32]:

$$\begin{aligned} \hat{\mathbf{R}}_f &\rightarrow \mathbf{R} & \text{e} & \hat{\mathbf{r}}_f \rightarrow \mathbf{r} \\ \hat{\mathbf{R}}_b &\rightarrow \mathbf{R} & \text{e} & \hat{\mathbf{r}}_b \rightarrow \mathbf{r} \end{aligned}$$

Assim,

$$\mathbf{R} \cdot \mathbf{w}'_f = \mathbf{r} \quad (4.16)$$

e

$$\mathbf{R} \cdot \mathbf{w}'_b = \mathbf{r} \quad (4.17)$$

Logo

$$\mathbf{w}'_f = \mathbf{w}'_b = \mathbf{w}' \quad (4.18)$$

ou seja, para (  $T$  ) simbolizando transposto:

$$\begin{bmatrix} w_{f1} & w_{f2} & \cdots & w_{fL} \end{bmatrix}^T = \begin{bmatrix} w_{b(L-1)} & w_{b(L-2)} & \cdots & w_{b0} \end{bmatrix}^T \quad (4.19)$$

Então, das Figs. 4.1 e 4.2, assintoticamente, os filtros apresentarão coeficientes simétricos conjugados e os erros produzidos apresentarão as mesmas características estatísticas. Logo, sob este critério, os filtros são semelhantes.

A minimização da soma das energias dos erros de predição fica expressa como:

$$\frac{\partial \varepsilon_f(\mathbf{w}'_f)}{\partial \mathbf{w}'_f} + \frac{\partial \varepsilon_b(\mathbf{w}'_b)}{\partial \mathbf{w}'_b} = \mathbf{0} \quad (4.20)$$

que, explicitando as derivadas de  $\varepsilon_f(\mathbf{w}'_f)$  e  $\varepsilon_b(\mathbf{w}'_b)$ , conforme o Apêndice A, pode ser escrita como:

$$\begin{aligned} (\mathbf{A}_f^H \mathbf{A}_f \mathbf{w}'_f - \mathbf{A}_f^H \mathbf{b}_f)^* + (\mathbf{A}_b^H \mathbf{A}_b \mathbf{w}'_b - \mathbf{A}_b^H \mathbf{b}_b)^* &= \mathbf{0} \\ \mathbf{A}_f^H \mathbf{A}_f \mathbf{w}'_f + \mathbf{A}_b^H \mathbf{A}_b \mathbf{w}'_b &= \mathbf{A}_f^H \mathbf{b}_f + \mathbf{A}_b^H \mathbf{b}_b \\ \mathbf{A}_f^H \mathbf{A}_f \mathbf{w}' + \mathbf{A}_b^H \mathbf{A}_b \mathbf{w}' &= \mathbf{A}_f^H \mathbf{b}_f + \mathbf{A}_b^H \mathbf{b}_b \\ \left[ \mathbf{A}_f^H \mid \mathbf{A}_b^H \right] \left[ \frac{\mathbf{A}_f}{\mathbf{A}_b} \right] \mathbf{w}' &= \left[ \mathbf{A}_f^H \mid \mathbf{A}_b^H \right] \left[ \frac{\mathbf{b}_f}{\mathbf{b}_b} \right] \end{aligned}$$

$$\mathbf{A}^H \mathbf{A} \cdot \mathbf{w}' = \mathbf{A}^H \mathbf{b} \quad (4.21)$$

onde  $\mathbf{A}$  é  $2(N - L) \times L$ , definida como:

$$\mathbf{A} = \left[ \frac{\mathbf{A}_f}{\mathbf{A}_b} \right] \quad (4.22)$$

e  $\mathbf{b}$  é  $2(N - L) \times 1$ , definido como:

$$\mathbf{b} = \left[ \frac{\mathbf{b}_f}{\mathbf{b}_b} \right] \quad (4.23)$$

É demonstrado em [3] que a solução  $\mathbf{w}'$  da equação normal FBLP (4.21) é dada, no sentido dos mínimos quadrados, por:

$$\mathbf{w}' = \mathbf{A}^\# \mathbf{b}. \quad (4.24)$$

onde o símbolo (  $\#$  ) significa a aplicação de pseudo-inversa [79][3], a qual será definida no Capítulo 5.

## 4.2 O Método FBLP para Estimação de Frequências

Definamos agora o sinal de entrada  $x[n]$  composto por  $M$  exponenciais complexas, inicialmente isento de ruído, da forma:

$$x[n] \triangleq \sum_{k=1}^M a_k \cdot [\exp j(\omega_k \cdot n + \phi_k)], \quad n = 0, 1, \dots, N - 1 \quad (4.25)$$

onde:

- $a_k$  são as amplitudes das exponenciais.
- $\omega_k$  são as frequências das exponenciais;  $\omega_k = 2\pi f_k$ , e  $f_k$  é a frequência normalizada em relação à frequência de amostragem.
- $\phi_k$  são as fases das exponenciais.

### 4.2.1 A Função de Transferência do Filtro FBLP

Aplicando o procedimento FBLP a  $x[n]$ , obtemos um vetor de coeficientes  $\mathbf{w}'$  que define um polinômio  $W(z)$  de ordem  $L$  da forma:

$$W(z) = -1 + w_1 z^{-1} + w_2 z^{-2} + \dots + w_L z^{-L} \quad (4.26)$$

**Teorema 4.1:** *Se  $L$  satisfaz a desigualdade  $M \leq L \leq (N - M/2)$ , e se  $\mathbf{w}'$  satisfaz  $\mathbf{A}\mathbf{w}' = \mathbf{b}$ , então  $W(z)$  tem  $M$  de seus zeros na circunferência de raio unitário (CRU) em  $z = \exp[j\omega_k]$ ,  $k = 1, \dots, M$ . Caso  $(M/2)$  não seja inteiro, este valor é arredondado para o maior inteiro mais próximo.*

**Teorema 4.2:** *Os  $(L - M)$  zeros restantes de  $W(z)$  localizam-se no interior da CRU, se o vetor  $\mathbf{w}'$  é de norma mínima.*

As demonstrações destes teoremas podem ser encontradas em [3].

Estas afirmações podem ser verificadas na figura 4.3 a seguir, referente ao caso em que  $M = 2$ ,  $L = 24$ ,  $N = 25$  amostras e as amplitudes e frequências são:

$$\begin{aligned} a_1 &= 1.0, \quad \omega_1 = (1, 00)\pi \quad \text{e} \quad \phi_1 = (-1, 00)\pi. \\ a_2 &= 1.0, \quad \omega_2 = (1, 04)\pi \quad \text{e} \quad \phi_2 = (-0, 79)\pi. \end{aligned}$$

Os zeros de  $W(z)$  que se posicionam sobre a CRU, são denominados zeros de sinal, enquanto os demais são os zeros estranhos. Dado que os zeros de sinal distinguem-se dos zeros estranhos e indicam os valores dos  $\omega_k$  através de sua posição angular, concluímos que o método FBLP permite estimar as  $M$  frequências de  $x[n]$ . Porém, tal método aplica-se somente à estimação de exponenciais não amortecidas pois, caso contrário, perde-se a informação das amplitudes  $a_i$ ,  $i = 1, \dots, M$ .

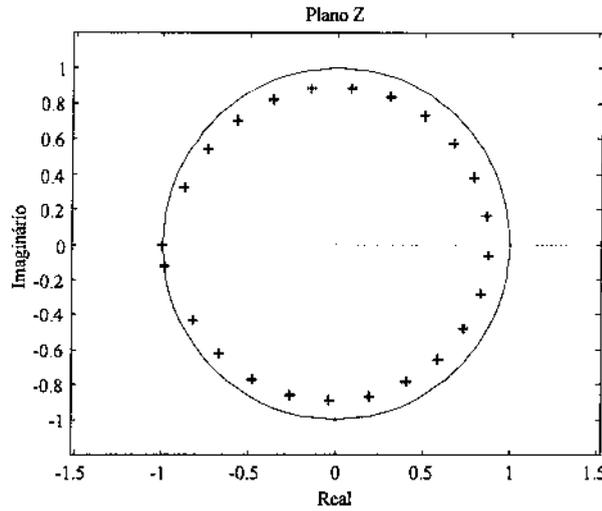


Figura 4.3: Zeros da função de transferência,  $L = 24$ ,  $SNR = \infty$ .

### 4.3 O Caso Ruidoso

Redefinamos o sinal  $x[n]$ , agora considerando-o corrompido pela inserção de ruído aditivo:

$$x[n] \triangleq \sum_{k=1}^M a_k \cdot [\exp j(\omega_k \cdot n + \phi_k)] + y[n] \quad n = 0, 1, \dots, N - 1 \quad (4.27)$$

onde  $y[n]$  é um ruído branco complexo com variâncias  $\sigma_{yr}^2 = \sigma_{yi}^2 = \sigma_y^2/2$ .

A relação sinal-ruído para cada exponencial complexa é definida como:

$$SNR_{dB} \triangleq 10 \cdot \log \left( \frac{a_k^2}{\sigma_y^2} \right) \quad (4.28)$$

onde  $\sigma_y^2/2$  é a variância da parte real ou da parte imaginária de  $y$ .

Enquanto que no caso sem ruído temos  $\text{posto}(\mathbf{A}) = M$ , agora, o  $\text{posto}(\mathbf{A}) = \min[2(N-L), L]$  e os teoremas 4.1 e 4.2 não mais se verificam. A distribuição espacial dos zeros de sinal e dos zeros estranhos de  $W(z)$  já não apresenta mais a mesma simetria anterior. Isto pode ser observado na figura 4.4 que se refere ao mesmo sinal da figura 4.3, mas agora com uma SNR de 10 dB para cada exponencial complexa.

Esta figura apresenta a superposição dos resultados de 50 experimentos, onde a cada um mantemos o mesmo sinal, porém alteramos o ruído. O objetivo é evidenciar a variância do posicionamento dos zeros produzida pelo ruído.

Em [3] verifica-se que, sob a influência de ruído, o vetor de coeficientes  $\mathbf{w}'$  define uma função  $W(z)$  tal que:

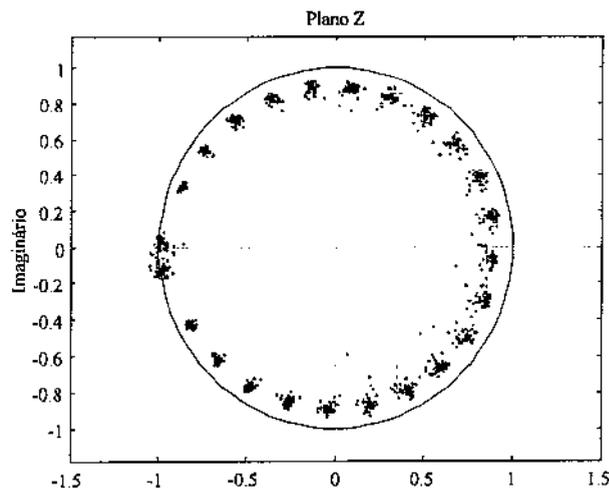


Figura 4.4: Zeros da função de transferência,  $L = 24$ ,  $\text{SNR} = 10\text{dB}$ .

- $M$  de seus zeros posicionam-se aleatoriamente na vizinhança dos zeros de sinal e tanto mais próximos dos mesmos quanto maior a relação sinal-ruído (SNR).
- Os  $(L - M)$  zeros restantes posicionam-se aleatoriamente ao longo da vizinhança da CRU, podendo eventualmente se posicionar sobre ou mesmo fora dela.
- A variância das estimativas dos zeros de sinal e dos zeros estranhos depende da relação entre o número de amostras disponíveis e a ordem da predição ( $N$  e  $L$ , respectivamente) e cresce com a redução da SNR.

De modo geral, a estratégia de identificação dos zeros de sinal consiste em associá-los aos  $M$  zeros mais próximos da CRU. O ruído insere uma aleatoriedade na abordagem anterior, prejudicando a estimação dos  $\omega_k$ 's através dos zeros de sinal e pode dar margem a estimativas drasticamente erradas. Apenas quando a SNR assume valores elevados é que os zeros de  $W(z)$  fornecem boas estimativas para os  $\omega_k$ 's, pois neste caso a situação é mais próxima daquela sem ruído.

À medida que a SNR decresce, as flutuações dos zeros de sinal ao redor de suas posições ideais produzem erros com variância crescente nas estimativas. Este comportamento altera-se drasticamente para SNR's abaixo de um determinado limiar, uma vez que a influência do ruído faz com que os zeros estranhos possam se posicionar mais próximos da CRU do que os zeros de sinal. Quando tal ocorre, um zero estranho é tomado como estimador de frequência e um grande erro é cometido. O desempenho do método degrada rapidamente dando origem ao fenômeno denominado efeito de limiar. A seguir apresentamos uma abordagem quantitativa da degradação da estimação sob diversas SNR's e ordens de predição  $L$ . Para tanto, conforme [91], definimos a medida de desempenho:

$$\text{var}\langle \hat{f}_1 \rangle \triangleq \frac{1}{\text{num}} \sum_{k=1}^{\text{num}} \left( \frac{\hat{\omega}_{1k} - \omega_1}{2\pi} \right)^2 \quad (4.29)$$

onde:

- $num$  é o número de experimentos realizados (utilizamos  $num = 500$ ).
- $\hat{\omega}_{1k}$  é a estimativa de  $\omega_1$  no  $k$ -ésimo experimento.

Para a representação das curvas de desempenho, utilizamos  $10 \cdot \log(1/\text{var}(\hat{f}_1))$ , com o objetivo de medir com maior sensibilidade as pequenas flutuações da variância. Esta medida é traçada em função da SNR [dB]. A figura 4.5 apresenta o limite de Cramér-Rao juntamente com as curvas de desempenho FBLP para as ordens  $L = 8, 12$  e  $24$ . Esta figura apresenta ainda a curva de desempenho do da Máxima Verossimilhança [3][41] para efeito de comparação.

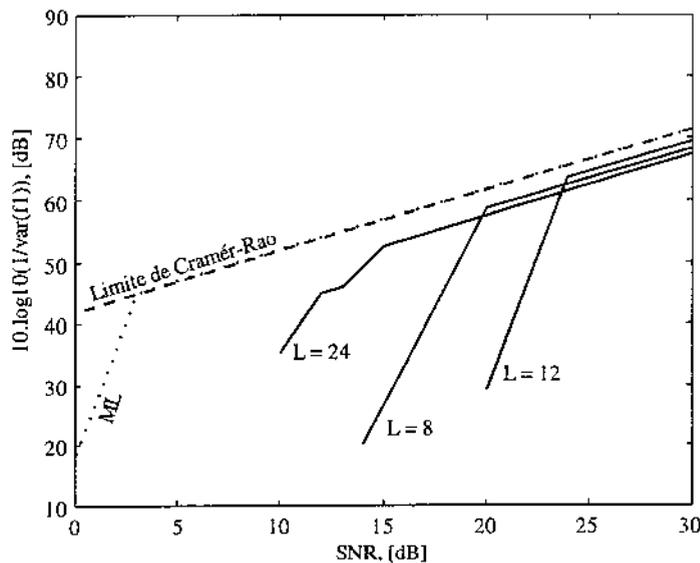


Figura 4.5: Comparação de desempenho entre os métodos ML e FBLP.

Podemos verificar a degradação contínua do desempenho à medida em que a SNR decresce, correspondente ao aumento das flutuações dos zeros de sinal em torno de suas posições ideais e conseqüente aumento da imprecisão das estimativas. Isto se verifica até o surgimento do efeito de limiar, identificado por uma mudança brusca da inclinação das curvas de desempenho. Estes limiares ocorrem para diferentes SNR's conforme a ordem de estimação adotada. Pode ser observado que o menor valor de limiar ocorre para  $L = 24$ , correspondendo a uma SNR de cerca de 15 dB. Observamos também que, mesmo acima do limiar, embora o desempenho do método FBLP se mostre inferior ao da Máxima Verossimilhança (ML), seu comportamento acompanha o limite de Cramér-Rao.

Estes efeitos podem ser atenuados pelo uso da Decomposição em Valores Singulares (SVD - *Singular Value Decomposition*), que separa a informação nos subespaços de sinal e de ruído. A exclusão do subespaço de ruído permite gerar estimativas mais precisas. No próximo capítulo analisaremos em detalhe estes procedimentos e seu poder de solução.

# Capítulo 5

## SVD e Subespaços

A presença do ruído degrada seriamente a estimação de frequências próximas entre si à medida que a relação sinal-ruído diminui. Esta degradação caracteriza-se pela redução de resolução e pelo surgimento de zeros estranhos mais próximos da CRU do que os zeros de sinal. Estes efeitos podem ser atenuados pelo uso da Decomposição em Valores Singulares (SVD - Singular Value Decomposition), que separa a informação nos subespaços de sinal e de ruído. A exclusão do subespaço de ruído permite gerar estimativas mais robustas tanto quanto às degradações produzidas pelas flutuações dos zeros de sinal ao redor de suas posições ideais, como quanto ao efeito de limiar.

Neste capítulo, inicialmente, apresentamos a SVD e definimos o que denominamos por subespaços de sinal e de ruído. A seguir, passamos a um estudo detalhado da SVD da matriz de dados  $\mathbf{A}$ . Verificaremos que não é possível recuperar toda a informação do sinal quando ele foi corrompido por ruído aditivo. No capítulo seguinte aplicaremos estes conceitos à estimação de frequências.

### 5.1 Definições

Dada uma matriz  $\mathbf{Z}$   $m \times n$ , existe uma função  $f$  definida por  $f(\mathbf{x}) = \mathbf{Z} \cdot \mathbf{x}$  tal que  $f : \mathbb{C}^n \rightarrow \mathbb{C}^m$ , ou seja,  $f$  é uma transformação linear que mapeia  $\mathbb{C}^n$  em  $\mathbb{C}^m$ . Assim, uma transformação linear pode ser caracterizada em termos da matriz que a representa:

- O domínio de  $f$  é definido como o espaço vetorial  $\mathcal{D}(\mathbf{Z})$  gerado pelas combinações lineares das linhas de  $\mathbf{Z}$ .
- A imagem ou *range* de  $f$  é definida como o espaço vetorial  $\mathcal{I}(\mathbf{Z})$  gerado pelas combinações lineares das colunas de  $\mathbf{Z}$ .
- O espaço nulo de  $f$  é a partição do seu domínio correspondente à imagem nula, ou seja, o espaço de todos  $\mathbf{x}$  tal que  $\mathbf{Z} \cdot \mathbf{x} = \mathbf{0}$ . Para tal utilizamos a representação  $\mathcal{N}(\mathbf{Z})$ .

É importante recordar também que o posto de uma matriz  $\mathbf{Z}$  é definido como o número de colunas ou linhas linearmente independentes, ou seja, é a dimensão da  $\mathcal{I}(\mathbf{Z})$  ou de  $\mathcal{D}(\mathbf{Z})$ . Após esta breve revisão, vamos agora abordar a SVD.

## 5.2 Decomposição em Valores Singulares

**Teorema 5.1:** *Toda matriz complexa  $\mathbf{Z}$   $m \times n$  pode ser fatorada como  $\mathbf{Z} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$  onde  $\mathbf{U}$  é  $m \times m$  e  $\mathbf{U}\mathbf{U}^H = \mathbf{U}^H\mathbf{U} = \mathbf{I}$ ,  $\mathbf{V}$  é  $n \times n$  e  $\mathbf{V}\mathbf{V}^H = \mathbf{V}^H\mathbf{V} = \mathbf{I}$ , e  $\mathbf{\Sigma}$  é uma matriz real  $m \times n$  com seus únicos elementos não nulos na diagonal principal. Estes elementos são chamados valores singulares e são ordenados como  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ , onde  $r$  é o posto de  $\mathbf{Z}$ . As colunas de  $\mathbf{U}$  e  $\mathbf{V}$ ,  $\mathbf{u}_i$  e  $\mathbf{v}_i$ , são, respectivamente, os vetores singulares de  $\mathbf{Z}$  à esquerda e à direita.*

A demonstração deste teorema e maiores detalhes sobre SVD podem ser obtidos em [3], [30] e [88].

Entretanto, devemos recordar que, interpretando  $\mathbf{Z}$  como uma transformação linear, a SVD decompõe esta transformação em dois conjuntos de bases vetoriais ortonormais: um conjunto  $\mathbf{U}$  gerador da imagem de  $\mathbf{Z}$  e um conjunto  $\mathbf{V}$  gerador do domínio de  $\mathbf{Z}$ . Os valores singulares, por sua vez, constituem-se nos módulos das projeções de  $\mathbf{Z}$  segundo cada direção definida por  $\mathbf{u}_i$  e  $\mathbf{v}_i$  em cada base.

## 5.3 A SVD da Matriz $\mathbf{A}$ e o Vetor $\mathbf{w}'$

Seja a matriz de dados  $\mathbf{A}$  ( $2(N-L) \times L$ ), definida no contexto da predição linear *forward-backward* de ordem  $L$ , aplicada à estimação de  $M$  exponenciais complexas a partir de  $N$  amostras de  $x[n]$ , conforme exposto no Capítulo 4. Analisando a SVD de  $\mathbf{A}$ , podemos escrever:

$$\mathbf{A} = \mathbf{U}_A \mathbf{\Sigma}_A \mathbf{V}_A^H = \mathbf{U}_A \cdot \left[ \begin{array}{c|c} \mathbf{\Sigma}_{A1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array} \right] \cdot \mathbf{V}_A^H \quad (5.1)$$

onde  $\mathbf{\Sigma}_{A1}$  é  $r \times r$ , para  $r = \min\{2(N-L), L\}$ , o posto de  $\mathbf{A}$ . Definimos, então, a pseudo-inversa da matriz  $\mathbf{A}$  como [79]:

$$\mathbf{A}^\# = \mathbf{V}_A \cdot \left[ \begin{array}{c|c} \mathbf{\Sigma}_{A1}^{-1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array} \right] \cdot \mathbf{U}_A^H$$

Dessa forma, a solução da equação normal (4.21), dada no sentido dos mínimos quadrados pode agora ser expressa a partir da SVD de  $\mathbf{A}$ :

$$\mathbf{w}' = \mathbf{A}^\# \cdot \mathbf{b} = \sum_{k=1}^{\min\{2(N-L), L\}} \frac{1}{\sigma_{Ak}} \mathbf{v}_{Ak} \mathbf{u}_{Ak}^H \mathbf{b} \quad (5.2)$$

Verificamos que esta solução sofre influência do ruído, o qual prejudica a estimação de  $w'$  à medida que a potência do ruído aumenta. Para reduzir estes danos, vamos estudar, a partir de agora, a atuação do ruído sobre o sinal.

## 5.4 Os Subespaços de Sinal e de Ruído

Sejam as seguintes partições na SVD de  $\mathbf{A}$ :

$$\mathbf{A} = \left[ \mathbf{U}_{\mathbf{A}1} \mid \mathbf{U}_{\mathbf{A}2} \right] \cdot \left[ \begin{array}{c|c} \boldsymbol{\Sigma}_{\mathbf{A}1} & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{A}2} \end{array} \right] \cdot \left[ \begin{array}{c} \mathbf{V}_{\mathbf{A}1}^H \\ \mathbf{V}_{\mathbf{A}2}^H \end{array} \right]. \quad (5.3)$$

onde:

- $\mathbf{U}_{\mathbf{A}1}$  é uma partição de  $\mathbf{U}_{\mathbf{A}}$  contendo as suas  $M$  primeiras colunas e  $\mathbf{U}_{\mathbf{A}2}$  contendo as suas  $2(N-L) - M$  últimas colunas;
- $\mathbf{V}_{\mathbf{A}1}$  é uma partição de  $\mathbf{V}_{\mathbf{A}}$  contendo as suas  $M$  primeiras colunas e  $\mathbf{V}_{\mathbf{A}2}$  contendo as suas  $L - M$  últimas colunas;
- $\boldsymbol{\Sigma}_{\mathbf{A}1}$  é uma partição quadrada  $M \times M$  de  $\boldsymbol{\Sigma}_{\mathbf{A}}$  contendo os  $M$  primeiros elementos de sua diagonal e  $\boldsymbol{\Sigma}_{\mathbf{A}2}$  retangular,  $[2(N-L) - M] \times [L - M]$ , contendo os  $\min[2(N-L), L] - M$  últimos elementos da diagonal de  $\boldsymbol{\Sigma}_{\mathbf{A}}$ .

No caso sem ruído, o posto de  $\mathbf{A}$  é  $M$  e  $\boldsymbol{\Sigma}_{\mathbf{A}2}$  é nula [3]. Logo, há  $M$  valores singulares não nulos, associados aos vetores singulares de  $\mathbf{U}_{\mathbf{A}1}$  e  $\mathbf{V}_{\mathbf{A}1}$ . Eles são denominados, respectivamente, por valores e vetores singulares de sinal.

No caso com ruído, o posto de  $\mathbf{A}$  é completo e igual a  $\min[2(N-L), L]$  e os  $\min[2(N-L), L] - M$  valores singulares que anteriormente eram nulos, tornam-se positivos. À medida que a SNR decresce, tais valores singulares aproximam-se mais e mais dos  $M$  primeiros, os valores singulares de sinal. Portanto, os menores  $\min[2(N-L), L] - M$  valores e vetores singulares associados estão relacionados unicamente ao ruído, sendo então denominados valores e vetores singulares de ruído.

Assim, podemos separar o espaço vetorial do sinal  $x[n]$  em dois subespaços:

- O Subespaço de Sinal: relacionado aos  $M$  maiores valores singulares de  $\mathbf{A}$  e aos vetores singulares associados e, portanto, relativos às exponenciais complexas contidas no sinal  $x[n]$ .
- O Subespaço de Ruído: relacionado aos  $\min[2(N-L), L] - M$  menores valores singulares de  $\mathbf{A}$  e aos vetores singulares correspondentes, isto é, associado à parcela de ruído,  $y[n]$ .

## 5.5 Dissecando os Subespaços

Conhecendo os subespaços de sinal e de ruído, a primeira idéia que nos ocorre é a diminuição da influência do ruído em  $\mathbf{w}'$ , o vetor de coeficientes ótimos de predição FBLP, pela exclusão dos valores e vetores singulares correspondentes ao subespaço de ruído na expressão (5.2). Este procedimento foi adotado por Tufts e Kumaresan em [91] e será apresentado no Capítulo 6. Entretanto, na referência citada, não foi feito um estudo mais detalhado da decomposição em valores singulares, de seu potencial nem das condições necessárias à sua aplicação para a estimação de frequências, conforme apresentado a seguir neste capítulo.

### 5.5.1 A Matriz de Dados Exatos $\mathbf{A}_E$

Uma vez que consideramos o sinal corrompido por ruído aditivo, podemos escrever a matriz  $\mathbf{A}$  como a soma de uma matriz  $\mathbf{A}_E$ , composta apenas por amostras de sinal, a uma matriz  $\mathbf{Y}$  unicamente de ruído, de mesmas dimensões que  $\mathbf{A}$ :

$$\mathbf{A} = \mathbf{A}_E + \mathbf{Y} \quad (5.4)$$

Como  $\mathbf{Y}$  contém amostras de ruído branco, suas colunas são independentes, logo seu posto é completo. A matriz  $\mathbf{A}_E$  equivale à matriz  $\mathbf{A}$  do caso sem ruído e seu posto é  $M$  e, como  $M \leq \min[2(N-L), L]$ , o posto de  $\mathbf{A}_E$  é incompleto.  $\mathbf{A}$  tem posto completo, conforme apresentado na Seção 4.3. Desse modo, a SVD de  $\mathbf{A}_E$  fica expressa por:

$$\mathbf{A}_E = \mathbf{U}_E \cdot \Sigma_E \cdot \mathbf{V}_E^H \quad (5.5)$$

ou seja,

$$\mathbf{A}_E = \left[ \mathbf{U}_{E1} \mid \mathbf{U}_{E2} \right] \cdot \left[ \begin{array}{c|c} \Sigma_{E1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array} \right] \cdot \left[ \begin{array}{c} \mathbf{V}_{E1}^H \\ \hline \mathbf{V}_{E2}^H \end{array} \right] \quad (5.6)$$

onde  $\Sigma_{E1}$  é uma matriz diagonal,  $M \times M$ , com os valores singulares de  $\mathbf{A}_E$  em sua diagonal principal. Além disso, os vetores  $\{\mathbf{u}_1, \dots, \mathbf{u}_M\}$  formam uma base ortonormal para a imagem de  $\mathbf{A}_E$ ,  $\mathcal{I}(\mathbf{A}_E)$ . Podemos observar ainda que os vetores  $\{\mathbf{v}_1, \dots, \mathbf{v}_M\}$  constituem uma base ortonormal para o domínio,  $\mathcal{D}(\mathbf{A}_E)$ , enquanto os vetores  $\{\mathbf{v}_{M+1}, \dots, \mathbf{v}_L\}$  formam uma base ortonormal para o espaço nulo de  $\mathbf{A}_E$ ,  $\mathcal{N}(\mathbf{A}_E)$ , pois  $\mathbf{A}_E \mathbf{V}_{E2} = \mathbf{0}$ .

Concluindo, a SVD de  $\mathbf{A}_E$  permite-nos escrever:

$$\mathbf{A}_E = \mathbf{U}_{E1} \Sigma_{E1} \mathbf{V}_{E1}^H \quad (5.7)$$

### 5.5.2 Considerações sobre a SVD de $\mathbf{A}$

Nos desenvolvimentos a seguir, buscaremos estabelecer condições suficientes para podermos expressar a SVD de  $\mathbf{A}$  em termos da SVD de  $\mathbf{A}_E$ , de forma que possamos extrair as informações de sinal da matriz de dados ruidosos  $\mathbf{A}$ .

Tomando a relação (5.4) entre  $\mathbf{A}$  e  $\mathbf{A}_E$  e lembrando que o Teorema 4.1 da SVD assegura que  $\mathbf{V}_E \mathbf{V}_E^H = \mathbf{I}$ , podemos escrever:

$$\mathbf{A} = (\mathbf{A}_E + \mathbf{Y}) \mathbf{V}_E \mathbf{V}_E^H \quad (5.8)$$

Particionando  $\mathbf{V}_E$  como em (5.6) temos:

$$\mathbf{A} = (\mathbf{A}_E + \mathbf{Y}) \left[ \mathbf{V}_{E1} \mid \mathbf{V}_{E2} \right] \begin{bmatrix} \mathbf{V}_{E1}^H \\ \mathbf{V}_{E2}^H \end{bmatrix} \quad (5.9)$$

logo,

$$\mathbf{A} = (\mathbf{A}_E + \mathbf{Y})(\mathbf{V}_{E1} \mathbf{V}_{E1}^H + \mathbf{V}_{E2} \mathbf{V}_{E2}^H) \quad (5.10)$$

Como  $\mathbf{A}_E \mathbf{V}_{E2} = \mathbf{0}$ , temos:

$$\mathbf{A} = (\mathbf{A}_E \mathbf{V}_{E1} + \mathbf{Y} \mathbf{V}_{E1}) \mathbf{V}_{E1}^H + (\mathbf{Y} \mathbf{V}_{E2}) \mathbf{V}_{E2}^H \quad (5.11)$$

A partir da expressão anterior, com alguma manipulação algébrica, podemos chegar a uma expressão para a decomposição em valores singulares de  $\mathbf{A}$  em termos dos vetores singulares à direita de  $\mathbf{A}_E$ . Para tanto, definimos as seguintes SVD's:

$$(\mathbf{A}_E \mathbf{V}_{E1} + \mathbf{Y} \mathbf{V}_{E1}) \triangleq \mathbf{P}_1 \mathbf{S}_1 \mathbf{Q}_1^H \quad (5.12)$$

e

$$(\mathbf{Y} \mathbf{V}_{E2}) \triangleq \mathbf{P}_2 \mathbf{S}_2 \mathbf{Q}_2^H \quad (5.13)$$

Substituindo as expressões anteriores na expressão (5.11), podemos escrever:

$$\mathbf{A} = \mathbf{P}_1 \mathbf{S}_1 \mathbf{Q}_1^H \mathbf{V}_{E1}^H + \mathbf{P}_2 \mathbf{S}_2 \mathbf{Q}_2^H \mathbf{V}_{E2}^H \quad (5.14)$$

ou seja,

$$\mathbf{A} = \left[ \mathbf{P}_1 \mid \mathbf{P}_2 \right] \cdot \begin{bmatrix} \mathbf{S}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_2 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{Q}_1^H \mathbf{V}_{E1}^H \\ \mathbf{Q}_2^H \mathbf{V}_{E2}^H \end{bmatrix} \quad (5.15)$$

Assim, expressamos  $\mathbf{A}$  em termos das componentes das SVD's de  $\mathbf{A}_E$  e de  $\mathbf{Y}$ . No sentido de interpretar a expressão anterior como a SVD de  $\mathbf{A}$ , primeiramente devemos particionar a SVD de  $\mathbf{A}$  da mesma forma que particionamos a SVD de  $\mathbf{A}_E$  para permitir uma comparação. Desse modo, fazemos:

$$\begin{aligned} \mathbf{A} &= \mathbf{U}_A \cdot \Sigma_A \cdot \mathbf{V}_A^H \\ &= \left[ \mathbf{U}_{A1} \mid \mathbf{U}_{A2} \right] \cdot \left[ \begin{array}{c|c} \Sigma_{A1} & \mathbf{0} \\ \hline \mathbf{0} & \Sigma_{A2} \end{array} \right] \cdot \left[ \begin{array}{c} \mathbf{V}_{A1}^H \\ \mathbf{V}_{A2}^H \end{array} \right]. \end{aligned} \quad (5.16)$$

Então, para que (5.15) se constitua em uma decomposição em valores singulares para a matriz ruidosa  $\mathbf{A}$ , faz-se necessário que  $\mathbf{P}_1$  e  $\mathbf{P}_2$  sejam ortogonais, ou seja:

$$\mathbf{P}_1^H \mathbf{P}_2 = \mathbf{0} \quad (5.17)$$

Como  $\mathbf{P}_1$  forma uma base ortogonal para  $\mathcal{I}[(\mathbf{A}_E + \mathbf{Y})\mathbf{V}_{E1}]$  e  $\mathbf{P}_2$  para  $\mathcal{I}(\mathbf{Y}\mathbf{V}_{E2})$ , a ortogonalidade entre  $\mathbf{P}_1$  e  $\mathbf{P}_2$  implica em que

$$\mathbf{V}_{E1}^H (\mathbf{A}_E^H + \mathbf{Y}^H) \mathbf{Y} \mathbf{V}_{E2} = \mathbf{0} \quad (5.18)$$

Uma das possibilidades para que esta condição se verifique é:

$$\mathbf{V}_{E1}^H (\mathbf{A}_E^H \mathbf{Y}) \mathbf{V}_{E2} = \mathbf{0} \quad (5.19)$$

e

$$\mathbf{V}_{E1}^H (\mathbf{Y}^H \mathbf{Y}) \mathbf{V}_{E2} = \mathbf{0} \quad (5.20)$$

Para atender às duas condições acima, muitas considerações são admissíveis. Se assumirmos que o ruído é "branco" no contexto de  $\mathbf{Y}$ , isto é:

$$\mathbf{Y}^H \mathbf{Y} = \sigma^2 \mathbf{I} \quad (5.21)$$

a condição (5.20) é satisfeita. No caso de (5.19), poderíamos supor também  $\mathbf{A}_E^H \mathbf{Y} = \gamma^2 \mathbf{I}$ . Entretanto, considerando o sinal descorrelacionado do ruído, temos:

$$\mathbf{A}_E^H \mathbf{Y} = \mathbf{0} \quad (5.22)$$

Assim, as restrições (5.21) e (5.22) garantem as igualdades de (5.19) e (5.20) e podemos especificar as condições suficientes para que a expressão (5.15) represente a SVD de  $\mathbf{A}$ :

- Os dados exatos devem ser ortogonais ao ruído no sentido de que  $\mathbf{A}_E^H \mathbf{Y} = \mathbf{0}$ ;
- As matrizes  $\mathbf{V}_{E1}$  e  $\mathbf{V}_{E2}$  devem ser ortogonais entre si no produto interno gerado por  $\mathbf{Y}^H \mathbf{Y}$ ,  $\mathbf{V}_{E1}^H (\mathbf{Y}^H \mathbf{Y}) \mathbf{V}_{E2} = \mathbf{0}$ . Tal ocorre quando o ruído satisfaz (5.21).
- O menor valor singular em  $\Sigma_1$  deve ser maior do que o maior valor singular em  $\Sigma_2$  de modo a respeitarmos o teorema 5.1. Esta ordenação define os vetores de  $\mathbf{V}_{E1}$  em contraste com aqueles de  $\mathbf{V}_{E2}$ .

No final deste capítulo, apresentaremos as motivações de ordem prática para a opção pelas hipóteses acima, bem como os limites para a sua validade.

5.5.3 A SVD de  $\mathbf{A}$  em termos da SVD de  $\mathbf{A}_E$ 

Uma vez escolhidas as condições para se aplicar a decomposição em valores singulares no sentido de extrair informações do sinal exato, partindo das considerações (5.21) e (5.22), buscaremos agora explicitar uma expressão para a SVD de  $\mathbf{A}$  em termos da SVD de  $\mathbf{A}_E$ . Tomando a expressão (5.4) e, substituindo a SVD de  $\mathbf{A}_E$  de (5.7) e utilizando  $\mathbf{Y} = \mathbf{Y}\mathbf{V}_E\mathbf{V}_E^H$ , obtemos:

$$\mathbf{A} = \mathbf{U}_{E1}\Sigma_{E1}\mathbf{V}_{E1}^H + \mathbf{Y}\mathbf{V}_{E1}\mathbf{V}_{E1}^H + \mathbf{Y}\mathbf{V}_{E2}\mathbf{V}_{E2}^H \quad (5.23)$$

Reagrupando, fica:

$$\mathbf{A} = (\mathbf{U}_{E1}\Sigma_{E1} + \mathbf{Y}\mathbf{V}_{E1})\mathbf{V}_{E1}^H + (\mathbf{Y}\mathbf{V}_{E2})\mathbf{V}_{E2}^H \quad (5.24)$$

que pode ser expressa na forma matricial por:

$$\mathbf{A} = \left[ (\mathbf{U}_{E1}\Sigma_{E1} + \mathbf{Y}\mathbf{V}_{E1}) \mid \mathbf{Y}\mathbf{V}_{E2} \right] \cdot \begin{bmatrix} \mathbf{V}_{E1}^H \\ \mathbf{V}_{E2}^H \end{bmatrix} \quad (5.25)$$

Para podermos identificar a expressão acima com uma SVD, devemos inserir uma matriz diagonal  $\Sigma_{\mathbf{A}}$  entre os seus dois fatores. Tendo em consideração o desenvolvimento clássico da SVD como em [32], [41] ou [30], a forma básica dessa matriz é dada por:

$$\Sigma_{\mathbf{A}} = \left[ \begin{array}{c|c} (\Sigma_{E1}^2 + \sigma_y^2\mathbf{I})^{1/2} & \mathbf{0} \\ \hline \mathbf{0} & \sigma_y\mathbf{I}_{ret} \end{array} \right] \quad (5.26)$$

onde  $\mathbf{I}_{ret}$  é uma matriz retangular com dimensões  $[2(N-L) - M] \times (L-M)$ , cujos seus únicos elementos não nulos estão em sua diagonal principal e valem 1.

Para podermos inserir essa matriz, devemos modificar ligeiramente o primeiro fator, resultando em:

$$\mathbf{A} = \left[ (\mathbf{U}_{E1}\Sigma_{E1} + \mathbf{Y}\mathbf{V}_{E1})(\Sigma_{E1}^2 + \sigma_y^2\mathbf{I})^{-1/2} \mid \sigma_y^{-1}\mathbf{Y}\mathbf{V}_{E2} \right] \cdot \left[ \begin{array}{c|c} (\Sigma_{E1}^2 + \sigma_y^2\mathbf{I})^{1/2} & \mathbf{0} \\ \hline \mathbf{0} & \sigma_y\mathbf{I}_{ret} \end{array} \right] \cdot \begin{bmatrix} \mathbf{V}_{E1}^H \\ \mathbf{V}_{E2}^H \end{bmatrix} \quad (5.27)$$

que deve ser comparada com a SVD de  $\mathbf{A}$ :

$$\mathbf{A} = \left[ \mathbf{U}_{A1} \mid \mathbf{U}_{A2} \right] \cdot \left[ \begin{array}{c|c} \Sigma_{A1} & \mathbf{0} \\ \hline \mathbf{0} & \Sigma_{A2} \end{array} \right] \cdot \begin{bmatrix} \mathbf{V}_{E1}^H \\ \mathbf{V}_{E2}^H \end{bmatrix} \quad (5.28)$$

Podemos então, por comparação, identificar  $\mathbf{U}_{A1}$  e  $\mathbf{U}_{A2}$ , observar que os valores singulares de sinal são dados por  $(\Sigma_{E1}^2 + \sigma_y^2\mathbf{I})^{1/2}$  e que os valores singulares de ruído, dada a condição  $\mathbf{Y}^H\mathbf{Y} = \sigma_y^2\mathbf{I}$ , são iguais entre si e valem  $\sigma_y$ .

Logo, os valores singulares do subespaço de sinal de  $\mathbf{A}$  são aqueles correspondentes de  $\mathbf{A}_E$  contaminados pela variância do ruído. Por outro lado, os valores singulares do subespaço de ruído de  $\mathbf{A}$  são iguais ao próprio desvio padrão do ruído.

Podemos ainda constatar que, respeitadas as condições (5.21) e (5.22),  $\mathbf{V}_{A1} = \mathbf{V}_{E1}$  e  $\mathbf{V}_{A2} = \mathbf{V}_{E2}$ .

### 5.5.4 Limitações no Conhecimento de $\mathbf{A}_E$

Como partimos das condições (5.21) e (5.22), pela subseção 5.5.3 garantimos que se mantém a ortogonalidade entre  $\mathbf{U}_{A1}$  e  $\mathbf{U}_{A2}$ ,  $\mathbf{V}_{A1}$  e  $\mathbf{V}_{A2}$ . Como também garantimos que o menor valor singular de  $\Sigma_{A1}$  é maior do que o maior valor singular de  $\Sigma_{A2}$ , podemos afirmar que a expressão (5.27) é a SVD de  $\mathbf{A}$ .

Os  $\min[2(N - L), L] - M$  menores valores singulares de  $\Sigma_A$  são iguais entre si e permitem estimar a variância do ruído  $\sigma_y^2$  a qual, por sua vez, possibilita a obtenção dos valores singulares de  $\mathbf{A}_E$  através da seguinte expressão:

$$\Sigma_{E1} = (\Sigma_{A1}^2 - \sigma_y^2 \mathbf{I})^{1/2}. \quad (5.29)$$

Os vetores à direita de  $\mathbf{A}_E$  podem ser recuperados a partir da SVD de  $\mathbf{A}$ , permitindo-nos conhecer o espaço das linhas de  $\mathbf{A}_E$ ,  $\mathcal{D}(\mathbf{A}_E)$ , pois  $\mathbf{V}_{A1} = \mathbf{V}_{E1}$ , bem como podemos recuperar também o espaço nulo de  $\mathbf{A}_E$ ,  $\mathcal{N}(\mathbf{A}_E)$ , pois  $\mathbf{V}_{A2} = \mathbf{V}_{E2}$ .

Os vetores singulares de  $\mathbf{U}_{A1}$  guardam uma relação pitagórica com os de  $\mathbf{U}_{E1}$ , pois  $\mathbf{U}_{A1} = (\mathbf{U}_{E1} \Sigma_{E1} + \mathbf{YV}_{E1}) (\Sigma_{E1}^2 + \sigma_y^2 \mathbf{I})^{-1/2}$ , onde:

- $\mathbf{U}_{E1} \Sigma_{E1}$  é ortogonal a  $\mathbf{YV}_{E1}$  pois, tendo em conta que  $\mathbf{V}_{E1}^H \mathbf{V}_{E1} = \mathbf{I}$ , temos

$$(\mathbf{U}_{E1} \Sigma_{E1})^H \mathbf{YV}_{E1} = (\mathbf{U}_{E1} \Sigma_{E1} \mathbf{V}_{E1}^H \mathbf{V}_{E1})^H \mathbf{YV}_{E1} \quad (5.30)$$

Mas  $\mathbf{A}_E = \mathbf{U}_{E1} \Sigma_{E1} \mathbf{V}_{E1}^H$ , logo:

$$(\mathbf{U}_{E1} \Sigma_{E1})^H \mathbf{YV}_{E1} = (\mathbf{A}_E \mathbf{V}_{E1})^H \mathbf{YV}_{E1} \quad (5.31)$$

que podemos escrever como:

$$(\mathbf{U}_{E1} \Sigma_{E1})^H \mathbf{YV}_{E1} = \mathbf{V}_{E1}^H \mathbf{A}_E^H \mathbf{YV}_{E1} \quad (5.32)$$

Como  $\mathbf{A}_E^H \mathbf{Y} = \mathbf{0}$ ,

$$(\mathbf{U}_{E1} \Sigma_{E1})^H \mathbf{YV}_{E1} = \mathbf{0} \quad (5.33)$$

- $\mathbf{U}_{E1} \Sigma_{E1}$  consiste de vetores coluna cujas normas são dadas pelos elementos na diagonal de  $\Sigma_{E1}$ ;
- $\mathbf{YV}_{E1}$  consiste de vetores coluna cujas normas são dadas por  $\sigma_y$ ;
- $\mathbf{U}_{A1}$  consiste de vetores coluna cujas normas são dadas pelos elementos diagonais de  $(\Sigma_{E1}^2 + \sigma_y^2 \mathbf{I})^{1/2}$ .

Verificamos, entretanto, que não nos é possível conhecer os  $\mathbf{u}_{Ei}$  exatos a partir da SVD de  $\mathbf{A}$  somente, pois  $\mathbf{U}_{A1} \neq \mathbf{U}_{E1}$  e não conhecemos  $\mathbf{Y}$ . Dessa forma não conseguimos excluir de forma completa a influência do ruído na estimação de  $\mathbf{w}'$ .

Os vetores de  $\mathbf{U}_E$  definem  $\mathcal{I}(\mathbf{A}_E)$  que, com a adição de ruído conforme (5.4), sofre uma rotação que produz  $\mathcal{I}(\mathbf{A})$ , o qual é definido pelos vetores de  $\mathbf{U}_A$ . Embora não possamos determinar  $\mathbf{U}_{E1}$ , podemos calcular os ângulos canônicos entre os espaços das colunas de  $\mathbf{U}_{A1}$  e de  $\mathbf{U}_{E1}$ , ou seja, entre os subespaços de sinal de  $\mathbf{A}$  e  $\mathbf{A}_E$ , a partir dos valores singulares do produto [30]:

$$\mathbf{U}_{E1}^H \mathbf{U}_{A1} = \mathbf{U}_{E1}^H (\mathbf{U}_{E1} \boldsymbol{\Sigma}_{E1} + \mathbf{Y} \mathbf{V}_{E1}) (\boldsymbol{\Sigma}_{E1}^2 + \sigma_y^2 \mathbf{I})^{-1/2} \quad (5.34)$$

Como, a partir de (5.22),  $\mathbf{A}_E$  é ortogonal a  $\mathbf{Y}$ , então  $\mathcal{I}(\mathbf{A}_E)$  é ortogonal a  $\mathcal{I}(\mathbf{Y})$ . Logo:

$$\mathbf{U}_{E1}^H \mathbf{Y} = 0 \quad (5.35)$$

Usando este fato e lembrando que  $\mathbf{U}_{E1}^H \mathbf{U}_{E1} = \mathbf{I}$ , temos:

$$\begin{aligned} \mathbf{U}_{E1}^H \mathbf{U}_{A1} &= \boldsymbol{\Sigma}_{E1} (\boldsymbol{\Sigma}_{E1}^2 + \sigma_y^2 \mathbf{I})^{-1/2} \\ &= \boldsymbol{\Sigma}_{E1} \boldsymbol{\Sigma}_{A1}^{-1} = \mathbf{C} \end{aligned} \quad (5.36)$$

Os elementos na diagonal de  $\mathbf{C}$  são, respectivamente, os cossenos  $c_i$ 's dos ângulos canônicos  $\alpha_i$ 's entre os subespaços de sinal de  $\mathbf{A}_E$  e  $\mathbf{A}$ . Estes ângulos canônicos permitem-nos delimitar um conjunto de infinitas estimativas para  $\mathbf{A}_E$ . Este conjunto é conhecido como "bola de matrizes" e é melhor descrito em [53][19]. Portanto, o conhecimento desses ângulos não é suficiente para determinarmos os vetores singulares à esquerda de  $\mathbf{A}_E$ , dados pela matriz  $\mathbf{U}_{E1}$ . Conseqüentemente, não podemos excluir por completo a influência do ruído no cálculo de  $\mathbf{w}'$ .

## 5.6 Resumindo

Em resumo, nas seções anteriores verificamos que a partir da SVD de  $\mathbf{A}$  e da validade das considerações (5.4), (5.21) e 5.22), podemos conhecer:

- o espaço das linhas de  $\mathbf{A}_E$ , pois  $\mathbf{V}_{A1} = \mathbf{V}_{E1}$ ;
- o espaço nulo de  $\mathbf{A}_E$ , pois  $\mathbf{V}_{A2} = \mathbf{V}_{E2}$ ;
- os valores singulares de  $\mathbf{A}_E$ , calculados a partir de (5.29);
- os cossenos dos ângulos canônicos entre os subespaços de sinal de  $\mathbf{A}$  e de  $\mathbf{A}_E$ , obtidos a partir da expressão (5.36).

Entretanto, isto não é suficiente para determinarmos os vetores singulares à esquerda de  $\mathbf{A}_E$ , dados pela matriz  $\mathbf{U}_{E1}$ . Conseqüentemente, não pudemos excluir por completo a influência do ruído no cálculo de  $\mathbf{w}'$ .

## 5.7 Considerações Estatísticas

A análise até agora efetuada restringiu-se aos aspectos algébricos. Nesta seção vamos apresentar os aspectos estatísticos que envolvem a consistência da formulação apresentada anteriormente.

Inicialmente, mostraremos que as condições (5.21) e (5.22) para a extração das informações de  $\mathbf{A}_E$  a partir da SVD de  $\mathbf{A}$  têm apoio estatístico, sob algumas considerações adicionais. Analisemos, então, a expressão a seguir:

$$\mathbf{A}^H \mathbf{A} = \mathbf{A}_E^H \mathbf{A}_E + \mathbf{Y}^H \mathbf{Y} + \mathbf{A}_E^H \mathbf{Y} + \mathbf{Y}^H \mathbf{A}_E \quad (5.37)$$

Primeiramente, observamos os termos cruzados  $\mathbf{A}_E^H \mathbf{Y}$  e  $\mathbf{Y}^H \mathbf{A}_E$  sob o ponto de vista estatístico e efetuamos as seguintes considerações:

- Os elementos de  $\mathbf{Y}$  devem ter média nula. Dessa forma, como  $\mathbf{A}_E$  é determinística,  $E\{\mathbf{A}_E^H \mathbf{Y}\} = \mathbf{0}$  e  $E\{\mathbf{Y}^H \mathbf{A}_E\} = \mathbf{0}$ ;
- Os elementos de  $\mathbf{Y}$  devem ser independentes e identicamente distribuídos com variância  $\nu^2$ , possivelmente desconhecida. Dessa forma, temos  $E\{\mathbf{Y}^H \mathbf{Y}\} = 2(N - L)\nu^2 \mathbf{I} \triangleq \sigma_y^2 \mathbf{I}$

Com estas considerações, a esperança estatística de  $\mathbf{A}^H \mathbf{A}$  fica expressa como:

$$E\{\mathbf{A}^H \mathbf{A}\} = \mathbf{A}_E^H \mathbf{A}_E + \sigma_y^2 \mathbf{I} \quad (5.38)$$

Como  $\mathbf{A}_E^H \mathbf{A}_E$  tem apenas  $M$  autovalores não nulos, os  $M$  maiores autovalores de  $E\{\mathbf{A}^H \mathbf{A}\}$  correspondem aos autovalores de  $\mathbf{A}_E^H \mathbf{A}_E$  acrescidos de  $\sigma_y^2 \mathbf{I}$ , enquanto os demais autovalores são iguais entre si, valendo  $\sigma_y^2 \mathbf{I}$ .

As observações acima sugerem que quando calculamos a média de diversos experimentos com uma mesma matriz  $\mathbf{A}_E$  e várias realizações de ruído, como é realizado neste trabalho, as condições algébricas descritas anteriormente são satisfeitas assintoticamente com o número de experimentos considerados, pois:

- $E\{\mathbf{A}_E^H \mathbf{Y}\} = \mathbf{0}$ ;
- $\mathbf{V}_{E1}^H E\{\mathbf{Y}^H \mathbf{Y}\} \mathbf{V}_{E2} = \mathbf{0}$ ;
- em média, o menor valor singular de  $\Sigma_{\mathbf{A}1}$  é maior do que o maior valor singular de  $\Sigma_{\mathbf{A}2}$ .

Para situações onde a matriz  $\mathbf{A}_E$  não pode ser utilizada repetidamente com realizações distintas de  $\mathbf{Y}$ , De Moor [19] demonstra que as condições acima são atingidas assintoticamente com o aumento do número de amostras da matriz de dados ruidosos.

Estas considerações representam a motivação para a adoção das condições suficientes representadas por (5.21) e (5.22).

## 5.8 Considerações de Ordem Prática

Observa-se experimentalmente que os algoritmos baseados em SVD mantêm-se robustos mesmo com matrizes de dados compostas por poucas amostras, como será verificado nos próximos capítulos. Mesmo para casos onde  $\|\mathbf{A}_E^H \mathbf{Y}\|$  é somente pequena, a SVD de  $\mathbf{A}$  fornece boas aproximações para  $\mathcal{I}(\mathbf{V}_{E1})$  e  $\mathcal{I}(\mathbf{V}_{E2})$ , e quanto menor esta norma, melhores serão as aproximações.

Entretanto, a aplicação da análise efetuada ao caso finito torna-se razoavelmente complicada, pois boa parte da informação é perdida com o truncamento do conjunto de amostras. Desse modo, não mais se pode garantir que o ruído seja branco. Além disso, passa a existir uma certa correlação entre o sinal e o ruído. Esta correlação, mesmo desconhecida, danifica a ortogonalidade entre  $\mathbf{A}_E$  e  $\mathbf{Y}$ . Embora ainda não existam abordagens completamente satisfatórias, uma análise do caso finito pode ser encontrada em [74].

A seguir, usaremos as informações deste capítulo no sentido de melhorar o desempenho do método FBLP para a estimação de frequências.

# Capítulo 6

## FBLP Modificado

No Capítulo 4 apresentamos o método FBLP para a estimação de frequências. Verificamos que este método permite a estimação de exponenciais complexas não amortecidas embora seu desempenho seja vulnerável ao ruído, degradando-se rapidamente à medida que a SNR atinge valores pequenos. No Capítulo 5 estudamos a influência do ruído sobre a matriz de dados  $\mathbf{A}$ . Observamos que podemos separar as informações em um subespaço somente de ruído e outro de sinal. Porém, não podemos excluir por completo a atuação do ruído, pois o subespaço de sinal também sofre contaminação por este último. Essa limitação traduz-se na impossibilidade de determinar uma única matriz  $\mathbf{A}_E$  a partir da SVD de  $\mathbf{A}$ . Neste capítulo vamos determinar a estimativa de  $\mathbf{A}_E$  no sentido dos mínimos quadrados. Partindo daí, definiremos o método FBLP Modificado, discutindo as suas características principais.

### 6.1 Mínimos Quadrados com Posto Reduzido - LSr

Uma vez que, pela SVD somente, não podemos determinar uma única matriz  $\mathbf{A}_E$  que poderia ter gerado  $\mathbf{A}$ , faremos uma estimativa. A estimativa de  $\mathbf{A}_E$ , matriz de dados exata de posto incompleto, no sentido dos mínimos quadrados, dado o conhecimento de  $\mathbf{A}$ , é obtida pela minimização da norma:

$$\min_{\mathbf{A}_{LSr}} \|\mathbf{A} - \mathbf{A}_{LSr}\|_F^2 \quad (6.1)$$

onde  $\mathbf{A}_{LSr} \in \mathbb{C}^{2(N-L) \times L}$  e tem posto  $M$ , pois este é o posto de  $\mathbf{A}_E$ .

A solução desta otimização [30] fornece:

$$\mathbf{A}_{LSr} = \mathbf{U}_{\mathbf{A}1} \boldsymbol{\Sigma}_{\mathbf{A}1} \mathbf{V}_{\mathbf{A}1}^H \quad (6.2)$$

ou seja, uma estimativa baseada apenas nos elementos do subespaço de sinal de  $\mathbf{A}$ .

Como a estimativa de  $\mathbf{w}'$  no sentido dos mínimos quadrados é dada por  $\mathbf{w}' = \mathbf{A}^\# \mathbf{b}$ , substi-

tuindo  $\mathbf{A}$  pela estimativa LS de posto reduzido  $\mathbf{A}_{LSr}$ , encontramos

$$\mathbf{w}'_{LSr} = \mathbf{A}_{LSr}^\# \mathbf{b} = \sum_{k=1}^M \frac{1}{\sigma_{\mathbf{A}k}} \mathbf{v}_{\mathbf{A}k} \mathbf{u}_{\mathbf{A}k}^H \mathbf{b} \quad (6.3)$$

A estimativa obtida acima nada mais é do que aquela proposta por Tufts e Kumaresan [91], que será tratada em maiores detalhes na seção a seguir.

## 6.2 Melhorando o FBLP

O método FBLP Modificado foi proposto por Tufts e Kumaresan [91]. Nele, busca-se diminuir os efeitos dos valores e vetores singulares do subespaço de ruído de  $\mathbf{A}$ , os quais prejudicam desastrosamente o desempenho do método FBLP para a estimação de frequências próximas sob relações sinal-ruído pequenas. A solução  $\mathbf{w}'$  de norma mínima, no sentido dos mínimos quadrados, pode ser escrita como [3]:

$$\mathbf{w}'_{LS} = \mathbf{A}^\# \cdot \mathbf{b} = \sum_{k=1}^{\min[2(N-L), L]} \frac{1}{\sigma_{\mathbf{A}k}} \mathbf{v}_{\mathbf{A}k} \mathbf{u}_{\mathbf{A}k}^H \mathbf{b} \quad (6.4)$$

Como vimos no Capítulo 5, a SVD decompõe a matriz  $\mathbf{A}$  em partes de sinal e de ruído. O mesmo se aplica a  $\mathbf{A}^\#$  e, desse modo, podemos recompô-la com somente a parte do subespaço de sinal, “filtrando-a”, de certo modo. Truncando a somatório em (6.4) de forma a envolver apenas os  $M$  primeiros termos, ou seja, aqueles relativos ao subespaço de sinal, temos:

$$\mathbf{w}'_{Mod.} = \sum_{k=1}^M \frac{1}{\sigma_{\mathbf{A}k}} \mathbf{v}_{\mathbf{A}k} \mathbf{u}_{\mathbf{A}k}^H \mathbf{b} \quad (6.5)$$

que é idêntica à expressão (6.3).

Esta estimativa de  $\mathbf{w}'$  já não é a de norma mínima, porém sofre menor influência de ruído. Tufts e Kumaresan partiram daí, definindo uma função  $W(z)_{Mod.}$  utilizando este  $\mathbf{w}'_{Mod.}$ . Note que o polinômio  $W(z)_{Mod.}$  mantém sua ordem  $L$ . Tufts e Kumaresan demonstram que, para SNR's não muito pequenas, os valores e vetores singulares de sinal de  $\mathbf{A}_{LSr}$  permanecem próximos dos correspondentes valores e vetores singulares de  $\mathbf{A}_E$ . Agora, portanto, os coeficientes de  $\mathbf{w}_{Mod.}$  aproximam-se mais daqueles do caso sem ruído. Desse modo, os zeros de  $W(z)_{Mod.}$  apresentam um comportamento próximo ao do caso não ruidoso, oferecendo melhores condições à estimação precisa das frequências.

Definimos, assim, o método FBLP Modificado pelo subespaço de sinal, ou método dos componentes principais. Este resultado utiliza as informações obtidas no Capítulo 5, reduzindo a atuação ruidosa, porém ainda conserva a influência do ruído sobre os valores e vetores singulares de sinal.

Os resultados obtidos em [91] são melhor discutidos em [3], de onde extraímos as figuras 6.1, 6.2 e 6.3, para as mesmas amplitudes e frequências utilizadas na figura 4.3.

## 6.3 Desempenho

Comparando as figuras 6.1 e 6.2, verificamos que a restrição de subespaço reduziu bastante as flutuações dos zeros estranhos devido ao ruído em relação às suas posições ideais, aproximando-se mais do caso não ruidoso. Entretanto, podemos acrescentar que a manutenção da atuação do ruído sobre os valores e vetores de sinal continua prejudicando os zeros de sinal, os quais ainda apresentam espalhamento ao redor de suas posições ideais.

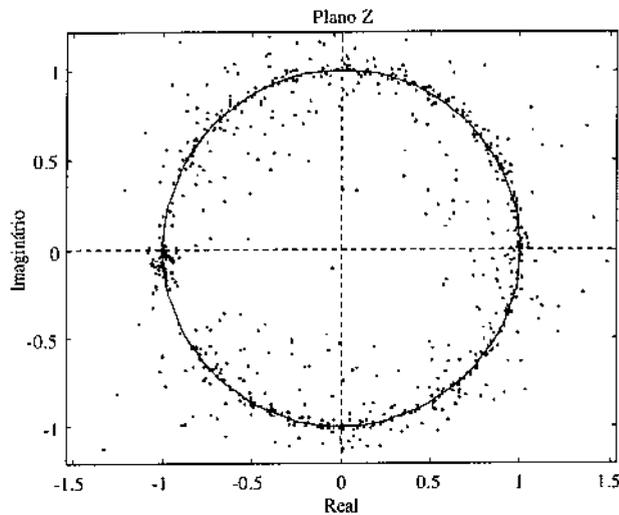


Figura 6.1: Zeros da função de transferência FBLP,  $N = 25$  amostras,  $L = 17$ , SNR = 10dB.

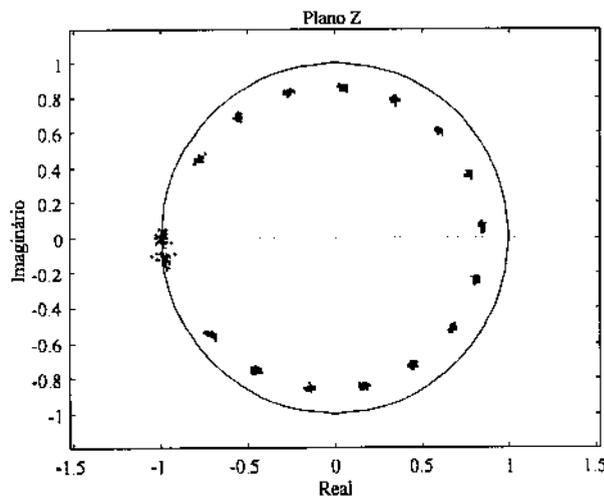


Figura 6.2: Zeros da função de transferência FBLP Modificado,  $N = 25$  amostras,  $L = 17$ , SNR = 10dB.

Na figura 6.3, podemos avaliar objetivamente os desempenhos dos métodos FBLP e FBLP Modificado para estimação de frequências. Observamos que a restrição ao subespaço de sinal

imprimiu significativo aumento na robustez ante o ruído, reduzindo a SNR de limiar em até 14 dB para  $L = 12$  e em 8 dB do melhor caso FBLP ( $L = 24$ ) para o melhor caso FBLP Modificado ( $L = 17$ ). O melhor desempenho foi alcançado para  $L = 17$  e SNR de limiar de 7dB, aproximando-se mais do desempenho dos métodos baseados em Máxima Verossimilhança.

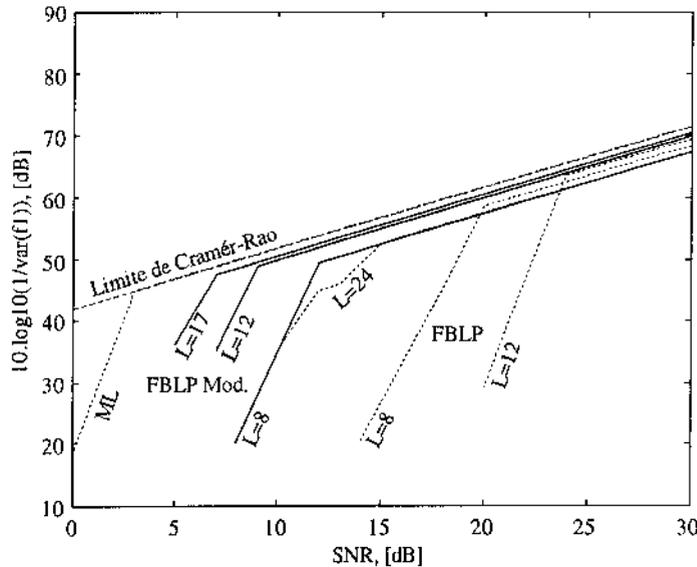


Figura 6.3: Comparação de desempenho entre os métodos ML, FBLP e FBLP Modificado.

Tufts e Kumaresan definiram experimentalmente o valor da ordem ótima para a estimação das frequências, ou seja, correspondente à menor SNR de limiar, como sendo  $L = (3/4)N$ . Para  $N = 25$ , a ordem ótima seria 18, porém verificamos que a mesma é na verdade 17. Acreditamos que Tufts e Kumaresan devam ter simulado apenas ordens pares, por isso não chegaram à nossa conclusão. No próximo capítulo apresentaremos simulações que esclarecerão de maneira conclusiva esta questão.

Em especial, destacamos o caso Kumaresan-Prony (KP), que ocorre quando  $L = N - (M/2)$ , ou seja,  $L = 24$  neste caso. Este valor de  $L$  faz com que a matriz  $\mathbf{A}$  tenha dimensões  $M \times L$ . Como o número de valores singulares não nulos de  $\mathbf{A}$  corresponde ao seu posto, que neste caso vale  $\min[M, L]$ ,  $\mathbf{A}$  tem somente  $M$  valores singulares correspondentes ao sinal, pois  $M \leq L$  por definição. Logo, a restrição da matriz  $\mathbf{A}$  ao subespaço de sinal ocorre automaticamente. Então, no caso KP, os métodos FBLP e FBLP Modificado se equivalem e  $\mathbf{w}'_{LS}$  não sofre contribuição do subespaço de ruído.

# Capítulo 7

## FBLP e Mínimos Quadrados Totais - TLS-FBLP

Nos capítulos anteriores utilizamos a otimização do filtro de erro de predição segundo o critério dos mínimos quadrados. Outro critério mais abrangente é o denominado Mínimos Quadrados Totais, ou TLS - do inglês *Total Least Squares*. Este critério permite otimizar de forma mais completa o vetor de coeficientes, admitindo erros tanto no vetor desejado como na matriz de dados. Neste capítulo propomos a aplicação do TLS à predição linear *forward-backward*, na tentativa de definir um método FBLP segundo o critério dos mínimos quadrados totais, TLS-FBLP, que supere o FBLP convencional. Neste sentido, detalhamos o desenvolvimento de soluções TLS pelos subespaços de sinal e de ruído. Propomos ainda uma formulação diferente para a aplicação do critério TLS, desenvolvendo um método que denominamos LS Equivalente ao TLS. Apresentamos também comparações, nas formas geométrica e analítica, entre o critério TLS e o mínimos quadrados convencional. Por fim, concluímos nossos esforços com uma comparação detalhada de desempenho entre o método proposto e o FBLP Modificado.

### 7.1 Mínimos Quadrados Totais - TLS

O método dos mínimos quadrados totais pode ser considerado como uma generalização da decomposição harmônica de Pisarenko, bem como um refinamento da abordagem dos mínimos quadrados, pois combate o efeito do ruído tanto na matriz de dados observados  $A$  quanto no vetor desejado  $b$ , através de ajustes em ambos quando da busca de um sistema homogêneo de equações para a otimização de  $w'$ . Intuitivamente, pode-se esperar um melhor comportamento do TLS em relação ao LS, que considera o efeito do ruído em um único termo da equação normal, como veremos a seguir.

### 7.1.1 Análise pela Teoria de Perturbações

Observando o problema de predição linear sob o ponto de vista da Teoria de Perturbações, podemos dizer que para obtermos a consistência do sistema de equações de predição linear:

$$\mathbf{A}\mathbf{w}' \approx \mathbf{b} \quad (7.1)$$

podemos inserir uma perturbação  $\Delta\mathbf{b}$  entre os dois termos, de modo a se estabelecer a igualdade. Desse modo podemos escrever:

$$\mathbf{A}\mathbf{w}' = (\mathbf{b} + \Delta\mathbf{b}) \quad (7.2)$$

A solução no sentido dos mínimos quadrados (LS) consiste em encontrar  $\mathbf{w}'$ , sujeito a (7.2), que minimize a norma quadrática de  $\Delta\mathbf{b}$ , isto é, que assegure:

$$\min \|\Delta\mathbf{b}\|_2, \quad \text{tal que } (\mathbf{b} + \Delta\mathbf{b}) \in \mathcal{I}(\mathbf{A}). \quad (7.3)$$

Neste caso  $(\mathbf{b} + \Delta\mathbf{b})$  poderá ser "predito" pelas colunas de  $\mathbf{A}$ , pois a expressão (7.2) implica em que, particionando  $\mathbf{A}$  em vetores coluna, tenhamos:

$$\left[ \mathbf{a}_1 \mid \mathbf{a}_2 \mid \cdots \mid \mathbf{a}_L \right] \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_L \end{bmatrix} = (\mathbf{b} + \Delta\mathbf{b}) \quad (7.4)$$

logo:

$$(\mathbf{b} + \Delta\mathbf{b}) = w_1\mathbf{a}_1 + w_2\mathbf{a}_2 + \cdots + w_L\mathbf{a}_L \quad (7.5)$$

ou seja,  $(\mathbf{b} + \Delta\mathbf{b})$  é uma combinação linear das colunas de  $\mathbf{A}$ .

No sentido de facilitar a compreensão da solução LS, poderíamos equivalentemente considerar a perturbação apenas na matriz  $\mathbf{A}$ , no sentido de obtermos a igualdade em (7.1):

$$(\mathbf{A} + \Delta\mathbf{A})\mathbf{w}' = \mathbf{b} \quad (7.6)$$

tal que a otimização recaia sobre uma matriz  $\tilde{\mathbf{A}} = \mathbf{A} + \Delta\mathbf{A}$ , que é uma aproximação da matriz  $\mathbf{A}$ , buscando  $\mathbf{w}'$  que minimize a norma de  $\Delta\mathbf{A}$  e assegure que:

$$\min \|\Delta\mathbf{A}\|_F, \quad \text{tal que } \mathbf{b} \in \mathcal{I}(\mathbf{A} + \Delta\mathbf{A}). \quad (7.7)$$

onde o índice  $_F$  indica norma de Frobenius. A solução genérica para este problema é obtida utilizando-se a pseudo-inversa da matriz aproximação  $\tilde{\mathbf{A}}$ :

$$\mathbf{w}' = \tilde{\mathbf{A}}\# \mathbf{b} \quad (7.8)$$

No caso LS, utilizaríamos simplesmente  $\tilde{\mathbf{A}} = \mathbf{A}$ , enquanto que, numa solução modificada como a de Tufts e Kumaresan, faríamos  $\tilde{\mathbf{A}} = \mathbf{A}_{LSr}$ , que tem posto reduzido. A solução (7.8) também satisfaz (7.2).

O critério TLS pressupõe a perturbação tanto em  $\mathbf{b}$  como em  $\mathbf{A}$ , para obtermos a igualdade em (7.1), conquistando assim um grau de liberdade maior para solucionar o nosso sistema. Portanto, escrevemos:

$$(\mathbf{A} + \Delta\mathbf{A}) \mathbf{w}' = (\mathbf{b} + \Delta\mathbf{b}) \quad (7.9)$$

onde  $\Delta\mathbf{A}$  é  $2(N - L) \times L$ , matriz de perturbação de  $\mathbf{A}$ . De outra forma:

$$\left( \left[ \begin{array}{c|c} \mathbf{b} & \mathbf{A} \end{array} \right] + \left[ \begin{array}{c|c} \Delta\mathbf{b} & \Delta\mathbf{A} \end{array} \right] \right) \begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix} = \mathbf{0} \quad (7.10)$$

que é uma equação homogênea como na decomposição harmônica de Pisarenko. Então, a solução TLS consiste em encontrar  $\mathbf{w}'$  que minimize a norma quadrática de  $\left[ \begin{array}{c|c} \Delta\mathbf{b} & \Delta\mathbf{A} \end{array} \right]$ , sujeito a (7.10).

### 7.1.2 A Solução dos Mínimos Quadrados Totais

Seja  $\mathbf{X} \triangleq \left[ \begin{array}{c|c} \mathbf{b} & \mathbf{A} \end{array} \right]$ ,  $\Delta\mathbf{X} \triangleq \left[ \begin{array}{c|c} \Delta\mathbf{b} & \Delta\mathbf{A} \end{array} \right]$  e  $\tilde{\mathbf{X}} = (\mathbf{X} + \Delta\mathbf{X})$  com dimensões  $2(N - L) \times (L + 1)$ . A solução TLS consiste em encontrar  $\mathbf{w}'$  que assegure:

$$\min \|\Delta\mathbf{X}\|_F^2, \text{ tal que } (\mathbf{b} + \Delta\mathbf{b}) \in \mathcal{I}(\mathbf{A} + \Delta\mathbf{A}) \quad (7.11)$$

Seguindo os passos desenvolvidos em [28], a estimativa TLS será obtida a partir da SVD de  $\mathbf{X}$ . Portanto, fazemos:

$$\mathbf{X} = \mathbf{U}_\mathbf{X} \Sigma_\mathbf{X} \mathbf{V}_\mathbf{X}^H \quad (7.12)$$

ou seja:

$$\mathbf{X} = \sum_{k=1}^{\min\{2(N-L), (L+1)\}} \sigma_{\mathbf{X}(k)} \mathbf{v}_{\mathbf{X}(k)} \mathbf{u}_{\mathbf{X}(k)}^H \quad (7.13)$$

onde:

- $\mathbf{U}_\mathbf{X} = [\mathbf{u}_{\mathbf{X}(1)} \cdots \mathbf{u}_{\mathbf{X}(2(N-L))}] \in \mathbb{C}^{2(N-L) \times 2(N-L)}$ ;
- $\mathbf{V}_\mathbf{X} = [\mathbf{v}_{\mathbf{X}(1)} \cdots \mathbf{v}_{\mathbf{X}(L+1)}] \in \mathbb{C}^{(L+1) \times (L+1)}$ ;
- $\Sigma_\mathbf{X} = \text{diag}[\sigma_{\mathbf{X}(1)}, \dots, \sigma_{\mathbf{X}(\min\{2(N-L), (L+1)\})}] \in \mathbb{R}^{2(N-L) \times (L+1)}$ ;
- $\sigma_{\mathbf{X}(1)} \geq \sigma_{\mathbf{X}(2)} \geq \dots \geq \sigma_{\mathbf{X}(M)} > \sigma_{\mathbf{X}(M+1)} = \dots = \sigma_{\mathbf{X}(\min\{2(N-L), (L+1)\})} \geq 0$ .

Como admitimos, a partir de (7.9), que  $(\mathbf{b} + \Delta\mathbf{b})$  é combinação linear das colunas de  $(\mathbf{A} + \Delta\mathbf{A})$ , podemos considerar que o posto de  $\left( \left[ \begin{array}{c|c} \mathbf{b} & \mathbf{A} \end{array} \right] + \left[ \begin{array}{c|c} \Delta\mathbf{b} & \Delta\mathbf{A} \end{array} \right] \right)$  é incompleto. Desse modo, pode ser demonstrado que [28]:

$$\min_{\text{posto}(\tilde{\mathbf{X}}) < \min\{2(N-L), (L+1)\}} \|\Delta\mathbf{X}\|_F = \sigma_{\mathbf{X}(\min\{2(N-L), (L+1)\})} \quad (7.14)$$

o qual é atingido se fizermos:

$$\Delta \mathbf{X} = -\mathbf{X}\mathbf{v}\mathbf{v}^H \quad (7.15)$$

e também:

$$\begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix} = -\frac{\mathbf{v}}{\alpha} \quad (7.16)$$

ou seja:

$$\mathbf{w}' = -\frac{\mathbf{z}}{\alpha} \quad (7.17)$$

onde:

$$\mathbf{v} = \begin{bmatrix} \alpha \\ \mathbf{z} \end{bmatrix} \quad (7.18)$$

é um vetor unitário no subespaço  $\mathcal{S}_w$  definido pela combinação linear de  $\{\mathbf{v}_{\mathbf{X}(M+1)}, \dots, \mathbf{v}_{\mathbf{X}(L+1)}\}$ , tal que  $\alpha \neq 0$  e  $\mathbf{z} \in \mathbb{C}^L$ .

Para verificarmos que esta escolha atende aos requisitos da solução procurada, vamos definir  $\Delta \mathbf{X}$  como em (7.15) e escrever:

$$\begin{aligned} (\mathbf{X} + \Delta \mathbf{X}) \begin{bmatrix} -1 \\ \mathbf{w} \end{bmatrix} &= (\mathbf{X} - \mathbf{X}\mathbf{v}\mathbf{v}^H) \begin{bmatrix} -1 \\ \mathbf{w} \end{bmatrix} \\ &= (\mathbf{X} - \mathbf{X}\mathbf{v}\mathbf{v}^H) \begin{bmatrix} -\mathbf{v} \\ \alpha \end{bmatrix} \\ &= -\frac{1}{\alpha} (\mathbf{X}\mathbf{v} - \mathbf{X}\mathbf{v}\mathbf{v}^H\mathbf{v}) \\ &= -\frac{1}{\alpha} (\mathbf{X}\mathbf{v} - \mathbf{X}\mathbf{v}) \\ &= \mathbf{0}. \end{aligned} \quad (7.19)$$

Portanto, o vetor  $\mathbf{w}'$  definido por (7.17) satisfaz a equação (7.10), assegurando que  $\|\Delta \mathbf{X}\|_F^2$  seja mínima. Logo, sob a luz das considerações efetuadas, segue-se que esse  $\mathbf{w}'$  resolve o problema TLS e depende apenas dos vetores singulares à direita correspondentes ao menor valor singular de  $\mathbf{X}$ . Entretanto, se não existir  $\mathbf{v}$  em  $\mathcal{S}_w$  tal que  $\alpha \neq 0$ , o sistema não tem solução segundo o critério TLS. Daqui em diante, denominaremos esse vetor ótimo do TLS como  $\mathbf{w}'_{TLS}$ .

No caso de um sistema subdeterminado, ou seja,  $2(N-L) < (L+1)$ , teríamos mais incógnitas do que equações em (7.10), o que, segundo [51], produz infinitas soluções com erro nulo. Para esta situação, a solução TLS é a mesma obtida pelo LS. Logo, podemos assumir  $\mathbf{w}'_{TLS} = \mathbf{A}^\# \mathbf{b}$ .

Para um sistema sobredeterminado,  $2(N-L) \geq (L+1)$  e  $\sigma_{\mathbf{X}(L+1)}$  é o menor valor singular de  $\mathbf{X}$ . Para o caso de  $\sigma_{\mathbf{X}(L+1)}$  ser um valor singular repetido de  $\mathbf{X}$ , temos infinitas soluções e, dentre elas, a solução de norma mínima.

Em particular, quando  $M = L$ , os valores singulares de  $\mathbf{X}$  são todos distintos e  $\sigma_{\mathbf{X}(L+1)}$  é único. Nesta situação é possível obter uma solução única dada por:

$$\mathbf{w}_{TLS} = -\frac{1}{v_{\mathbf{X}(1,(L+1))}} \begin{bmatrix} v_{\mathbf{X}(2,(L+1))} \\ \vdots \\ v_{\mathbf{X}[(L+1),(L+1)]} \end{bmatrix} \quad (7.20)$$

dado que  $v_{\mathbf{X}[1,(L+1)]} \neq 0$ , pois existe apenas um vetor  $\mathbf{v}$  em  $\mathbf{S}_w$ . Neste caso, a solução TLS reduz-se à mesma estrutura algébrica da solução de Pisarenko.

No caso de existirem infinitas soluções, isto é, para  $2(N - L) > (L + 1)$ , conforme Golub [28], podemos obter a solução TLS de norma mínima particionando a SVD de  $\mathbf{X}$  da seguinte forma:

$$\mathbf{X} = \left[ \mathbf{U}_{\mathbf{X}1} \mid \mathbf{U}_{\mathbf{X}2} \right] \left[ \begin{array}{c|c} \Sigma_{\mathbf{X}1} & \mathbf{0} \\ \hline \mathbf{0} & \Sigma_{\mathbf{X}2} \end{array} \right] \left[ \begin{array}{c} \mathbf{V}_{\mathbf{X}1}^H \\ \mathbf{V}_{\mathbf{X}2}^H \end{array} \right] \quad (7.21)$$

Submetendo-se  $\mathbf{V}_{\mathbf{X}2}$  a uma transformação  $\mathbf{H}$  de Householder tal que:

$$\mathbf{V}_{\mathbf{X}2}\mathbf{H} = \left[ \begin{array}{c|c} \mathbf{0} & \alpha \\ \hline \mathbf{z} & \mathbf{z} \end{array} \right] \quad (7.22)$$

obtemos o escalar  $\alpha$  e o vetor  $\mathbf{z}$ , que nos permitem escrever a solução TLS de norma mínima como:

$$\mathbf{w}'_{TLS} = -\frac{\mathbf{z}}{\alpha} \quad (7.23)$$

Esta solução atende aos casos de infinitas soluções para sistemas sobre e subdeterminados, igualando-se à solução única quando  $L = M$ . Na subsecção a seguir detalharemos o cálculo desta solução, apresentando-a em duas versões distintas: uma utilizando o subespaço de sinal e outra utilizando o subespaço de ruído.

### 7.1.3 As Soluções TLS pelos Subespaços de Sinal e de Ruído

Nesta subsecção reproduzimos a formulação empregada em Dowling e DeGroat [23] e em Lemos e Lopes [58][60], no sentido de obter uma expressão direta para o cálculo da solução TLS de norma mínima. Como a solução TLS clássica baseia-se no subespaço de ruído, de início trataremos deste caso e, posteriormente, estenderemos nossa análise ao subespaço de sinal. Dessa maneira, procedendo à seguinte partição de  $\mathbf{V}_{\mathbf{X}2}$ :

$$\mathbf{V}_{\mathbf{X}2} = \left[ \begin{array}{c} \mathbf{c}^T \\ \mathbf{V}'_{\mathbf{X}2} \end{array} \right] \quad (7.24)$$

onde  $\mathbf{c}^T$ ,  $1 \times (L + 1 - M)$ , é a primeira linha e  $\mathbf{V}'_{\mathbf{X}2}$ ,  $L \times (L + 1 - M)$ , são as outras linhas de  $\mathbf{V}_{\mathbf{X}2}$ , a partir de (7.22) concluímos que  $\mathbf{H}$  deve satisfazer:

$$\mathbf{c}^T\mathbf{H} = [0, \dots, 0, \alpha] \quad \text{ou seja,} \quad \mathbf{H}\mathbf{c}^* = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \alpha^* \end{bmatrix} \quad (7.25)$$

Definindo  $\mathbf{e}_{(L+1-M)}^T = [0, \dots, 0, 1]_{1 \times (L+1-M)}$ , a transformação de Householder pode ser escrita como [30]:

$$\mathbf{H} = \mathbf{I} - 2 \frac{\mathbf{s}\mathbf{s}^H}{\mathbf{s}^H\mathbf{s}} \quad (7.26)$$

na qual  $\mathbf{s} = \mathbf{c}^* - \alpha^* \mathbf{e}_{(L+1-M)}$  e  $\alpha^*$  é calculado como:

$$\alpha^* = \pm \frac{c_{(L+1-M)}}{|c_{(L+1-M)}|} \|\mathbf{c}\| \quad (7.27)$$

Para a solução TLS, utilizamos apenas a última coluna de  $\mathbf{V}_{\mathbf{X}_2}\mathbf{H} = \mathbf{V}_{\mathbf{X}_2} [\mathbf{h}_1 \mid \cdots \mid \mathbf{h}_{(L+1-M)}]$ , dada por:

$$\mathbf{V}_{\mathbf{X}_2}\mathbf{h}_{(L+1-M)} = \begin{bmatrix} \alpha \\ \mathbf{z} \end{bmatrix} \quad (7.28)$$

Como  $s_{(L+1-M)}^* = c_{(L+1-M)} - \alpha$  e  $\mathbf{s}^H\mathbf{s} = 2\alpha^* (\alpha - c_{(L+1-M)})$ , temos:

$$\begin{aligned} \mathbf{h}_{(L+1-M)} &= \mathbf{e}_{(L+1-M)} - 2 \frac{s_{(L+1-M)}^* \mathbf{s}}{\mathbf{s}^H\mathbf{s}} \\ &= \mathbf{e}_{(L+1-M)} - \frac{2c_{(L+1-M)} - \alpha}{2\alpha^* (\alpha - c_{(L+1-M)})} (\mathbf{c}^* - \alpha^* \mathbf{e}_{(L+1-M)}) \\ &= \frac{\mathbf{c}^*}{\alpha^*} \end{aligned} \quad (7.29)$$

Daí obtemos:

$$\mathbf{z} = \frac{\mathbf{V}'_{\mathbf{X}_2}\mathbf{c}^*}{\alpha^*} \quad (7.30)$$

e, dado (7.27), temos que:

$$\begin{aligned} \mathbf{w}'_{TLS} &= -\frac{\mathbf{z}}{\alpha} \\ &= -\frac{\mathbf{V}'_{\mathbf{X}_2}\mathbf{c}^*}{\alpha^*\alpha} \\ &= -\frac{\mathbf{V}'_{\mathbf{X}_2}\mathbf{c}^*}{\mathbf{c}^H\mathbf{c}} \end{aligned} \quad (7.31)$$

Esta é a solução TLS com base no subespaço de ruído. Podemos encontrar também sua versão com base no subespaço de sinal. Para tanto, fazemos a seguinte partição de  $\mathbf{V}_{\mathbf{X}}$ :

$$\begin{aligned} \mathbf{V}_{\mathbf{X}} &= \left[ \mathbf{V}_{\mathbf{X}_1} \mid \mathbf{V}_{\mathbf{X}_2} \right] \\ &= \left[ \begin{array}{c|c} \mathbf{g}^T & \mathbf{c}^T \\ \hline \mathbf{V}'_{\mathbf{X}_1} & \mathbf{V}'_{\mathbf{X}_2} \end{array} \right] \end{aligned} \quad (7.32)$$

Porém, a partir de  $\mathbf{V}_x \mathbf{V}_x^H = \mathbf{I}$ , temos:

$$\begin{aligned} & \left[ \begin{array}{c|c} \mathbf{g}^T & \mathbf{c}^T \\ \hline \mathbf{V}'_{x1} & \mathbf{V}'_{x2} \end{array} \right] \left[ \begin{array}{c|c} \mathbf{g}^* & \mathbf{V}'_{x1}{}^H \\ \hline \mathbf{c}^* & \mathbf{V}'_{x2}{}^H \end{array} \right] = \mathbf{I} \\ & \left[ \begin{array}{c|c} \mathbf{g}^T \mathbf{g}^* + \mathbf{c}^T \mathbf{c}^* & \mathbf{g}^T \mathbf{V}'_{x1}{}^H + \mathbf{c}^T \mathbf{V}'_{x2}{}^H \\ \hline \mathbf{V}'_{x1} \mathbf{g}^* + \mathbf{V}'_{x2} \mathbf{c}^* & \mathbf{V}'_{x1} \mathbf{V}'_{x1}{}^H + \mathbf{V}'_{x2} \mathbf{V}'_{x2}{}^H \end{array} \right] = \mathbf{I} \\ & \left[ \begin{array}{c|c} \mathbf{g}^H \mathbf{g} + \mathbf{c}^H \mathbf{c} & \mathbf{g}^T \mathbf{V}'_{x1}{}^H + \mathbf{c}^T \mathbf{V}'_{x2}{}^H \\ \hline \mathbf{V}'_{x1} \mathbf{g}^* + \mathbf{V}'_{x2} \mathbf{c}^* & \mathbf{V}'_{x1} \mathbf{V}'_{x1}{}^H + \mathbf{V}'_{x2} \mathbf{V}'_{x2}{}^H \end{array} \right] = \left[ \begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{I} \end{array} \right] \end{aligned} \quad (7.33)$$

o que nos permite, a partir da primeira linha e por comparação, afirmar que:

$$\mathbf{c}^H \mathbf{c} = 1 - \mathbf{g}^H \mathbf{g} \quad (7.34)$$

$$-\mathbf{V}'_{x2} \mathbf{c}^* = \mathbf{V}'_{x1} \mathbf{g}^* \quad (7.35)$$

Dessa forma, (7.31) pode finalmente ser escrita de maneira alternativa como:

$$\mathbf{w}'_{TLS} = \frac{\mathbf{V}'_{x1} \mathbf{g}^*}{1 - \mathbf{g}^H \mathbf{g}} \quad (7.36)$$

que constitui a versão com base no subespaço de sinal para a solução TLS.

As expressões (7.31) e (7.36) podem ser aplicadas de forma direta, pois a transformação de Householder está implícita em ambas. Embora estas duas expressões forneçam a mesma estimativa, conforme as dimensões dos subespaços aconselha-se o uso daquela que envolver o menor número de operações.

## 7.2 Comparação entre TLS e LS

Nesta seção buscamos estabelecer as diferenças entre os métodos LS e TLS. Inicialmente, interpretamos geometricamente os procedimentos de otimização empregados no LS e no TLS. Em seguida, comparamos analiticamente as duas soluções, verificando que a solução TLS corresponde na verdade a uma solução à maneira LS porém com a matriz de dados e o vetor desejado sofrendo restrição conjunta ao subespaço de sinal. Isto dá origem a uma forma peculiar de solução modificada para o subespaço de sinal, levando em conta as perturbações nos dois membros da equação normal. Entretanto, demonstramos que a solução clássica para o FBLP Modificado não corresponde à solução TLS-FBLP. Finalizando, confrontaremos experimentalmente os desempenhos dos métodos FBLP Modificado, TLS-FBLP com base no subespaço de sinal, e da solução de norma mínima com restrição conjunta da matriz de dados e do vetor desejado ao subespaço de sinal.

### 7.2.1 Interpretação Geométrica dos Mínimos Quadrados e Mínimos Quadrados Totais

Apresentamos agora uma interpretação geométrica para a diferença entre as minimizações LS e TLS. Minimizar o erro TLS equivale a minimizar as perturbações  $\|\Delta\mathbf{X}\|_2$  sujeito à condição (7.10). A norma  $\|\Delta\mathbf{X}\|_2$  é calculada como:

$$\|\Delta\mathbf{X}\|_2 = \max_{\mathbf{q} \neq \mathbf{0}} \frac{\|\Delta\mathbf{X}\mathbf{q}\|_2}{\|\mathbf{q}\|_2} \quad (7.37)$$

A matriz  $\Delta\mathbf{X}$ , condicionada por (7.10), apresentará norma mínima se satisfizer (7.15). Neste caso, segundo Golub [28] e Van Huffel [97], temos

$$\|\Delta\mathbf{X}\|_2 = \frac{\|\Delta\mathbf{X}\mathbf{v}\|_2}{\|\mathbf{v}\|_2} \quad (7.38)$$

para  $\mathbf{v} \in \mathbf{S}_w$ . Além disto, temos que, para  $2(N-L) \geq (L+1)$ :

$$\|\Delta\mathbf{X}\|_2 = \sigma_{\mathbf{X}(L+1)} \quad (7.39)$$

Portanto,  $\sigma_{\mathbf{X}(L+1)}^2$ , neste caso, é o valor do erro médio quadrático envolvido na predição linear.

Usando (7.16), substituímos  $\mathbf{v}$  e temos:

$$\frac{\left\| \Delta\mathbf{X} \begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix} \right\|_2}{\left\| \begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix} \right\|_2} = \sigma_{\mathbf{X}(L+1)} \quad (7.40)$$

Como, a partir de (7.10), podemos fazer:

$$\mathbf{X} \begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix} = -\Delta\mathbf{X} \begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix} \quad (7.41)$$

então:

$$\frac{\left\| \mathbf{X} \begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix} \right\|_2}{\left\| \begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix} \right\|_2} = \sigma_{\mathbf{X}(L+1)} \quad (7.42)$$

A interpretação geométrica do problema TLS torna-se visível quando calculamos:

$$\frac{\left\| \mathbf{X} \begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix} \right\|_2^2}{\left\| \begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix} \right\|_2^2} = \frac{\left\| \begin{bmatrix} \mathbf{b} & | & \mathbf{A} \end{bmatrix} \begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix} \right\|_2^2}{\left\| \begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix} \right\|_2^2} = \sum_{i=1}^{2(N-L)} \frac{|\mathbf{a}_i^H \mathbf{w}'_{TLS} - b_i|^2}{1 + \mathbf{w}'_{TLS}^H \mathbf{w}'_{TLS}} \quad (7.43)$$

onde  $\mathbf{a}_i$  é a  $i$ -ésima linha de  $\mathbf{A}$  e  $b_i$  é o  $i$ -ésimo elemento de  $\mathbf{b}$ . Vamos, agora, interpretar a quantidade

$$\mathbf{d}_i = \frac{|\mathbf{a}_i^H \mathbf{w}' - b_i|^2}{1 + \mathbf{w}'^H \mathbf{w}'} \quad (7.44)$$

No caso mais simples quando  $L = 1$ ,  $\mathbf{A}$  e  $\mathbf{b}$  são vetores  $2(N - 1) \times 1$ , enquanto  $\mathbf{w}'$  é um escalar. Assim, a relação  $\mathbf{A}\mathbf{w}' \approx \mathbf{b}$  fica expressa como:

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{2(N-1)} \end{bmatrix} w \approx \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{2(N-1)} \end{bmatrix} \quad (7.45)$$

Na figura 7.1 apresentamos uma representação no plano para as relações  $a_i w \approx b_i$ ,  $1 \leq i \leq 2(N - 1)$ . Dado um conjunto de pontos  $(a_i, b_i)$ , correspondentes às coordenadas de  $\mathbf{A}$  e  $\mathbf{b}$ , tentamos aproximá-los com uma reta de coeficiente angular  $w$ .

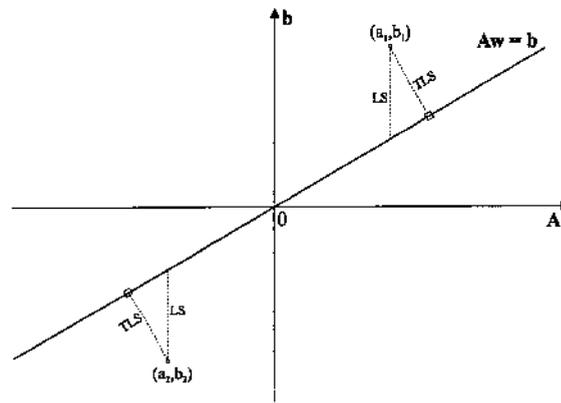


Figura 7.1: Diferenças no ajuste da reta  $w$  pelos critérios LS ou TLS.

Na otimização LS consideramos erros apenas em  $\mathbf{b}$  para aproximar o conjunto de pontos pela reta de coeficiente angular  $w'_{LS}$ . Na figura 7.2, tomando apenas o ponto  $(a_i, b_i)$ , o critério LS minimiza a distância do ponto  $(a_i, a_i w'_{LS})$  ao ponto dado, a qual tem o valor  $|a_i w'_{LS} - b_i|$ .

Por outro lado, na otimização TLS consideramos erros tanto em  $\mathbf{b}$  como  $\mathbf{a}$  de forma a aproximar os pontos pela reta de coeficiente angular  $w'_{TLS}$ . Na figura 7.3, verificamos que a reta  $w_{TLS}$  inclina-se de um ângulo  $\beta$ , tal que  $\cos(\beta) = 1/(w_{TLS}^2 + 1)^{1/2}$ . Logo, analisando a parcela da norma quadrática mínima do erro TLS dada por (7.44), verificamos que podemos escrever:

$$\frac{|a_i w_{TLS} - b_i|^2}{1 + w_{TLS}^2} = |a_i w_{TLS} - b_i|^2 \cos^2(\beta) \quad (7.46)$$

indicando que esse critério minimiza a distância perpendicular da reta ao ponto dado.

Assim, generalizando para outros valores de  $L$ , o critério TLS minimiza a soma das distâncias quadráticas de cada ponto  $\begin{bmatrix} \mathbf{a}_i \\ b_i \end{bmatrix} \in \mathbb{C}^{L+1}$  ao ponto mais próximo do subespaço  $\mathcal{P}_w$  definido por

$$\mathcal{P}_w = \left\{ \begin{bmatrix} \mathbf{a} \\ b \end{bmatrix} \mid \mathbf{a} \in \mathbb{C}^L, b \in \mathbb{C}, b = \mathbf{w}^H \mathbf{a} \right\} \quad (7.47)$$

Concluimos, portanto, que para o ajuste de uma curva entre os pontos observados, a abordagem LS atua na minimização das distâncias verticais, considerando o erro apenas no vetor  $\mathbf{b}$ , enquanto que o TLS minimiza as distâncias perpendiculares, ajustando tanto  $\mathbf{b}$  quanto  $\mathbf{A}$ .

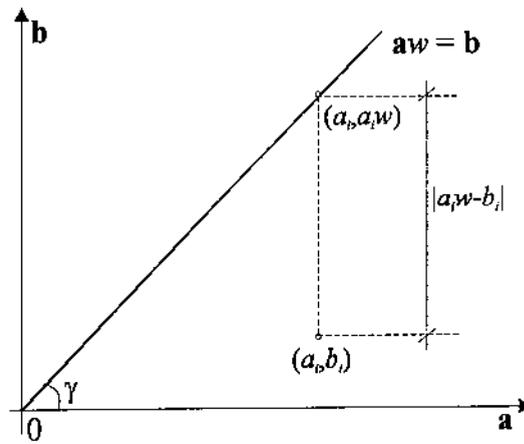


Figura 7.2: O ajuste LS.

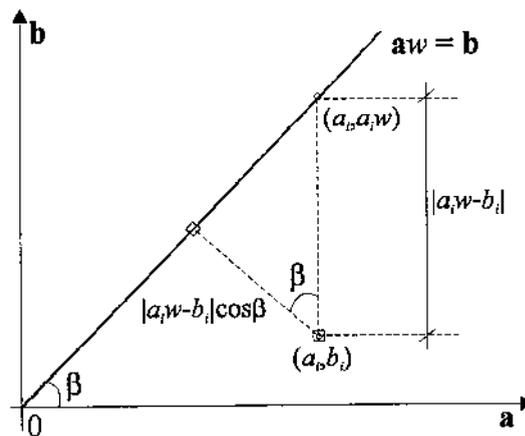


Figura 7.3: O ajuste TLS.

### 7.2.2 A Solução LS Equivalente

Tomando agora o caso LS, propomos uma nova solução de norma mínima modificada, porém efetuando a restrição sobre a matriz de dados estendida, ao invés de sobre a matriz de dados simplesmente. Lembrando (7.21), temos que:

$$\left[ \mathbf{b} \mid \mathbf{A} \right] = \mathbf{X} = \left[ \mathbf{U}_{\mathbf{X}_1} \mid \mathbf{U}_{\mathbf{X}_2} \right] \left[ \begin{array}{c|c} \Sigma_{\mathbf{X}_1} & \mathbf{0} \\ \hline \mathbf{0} & \Sigma_{\mathbf{X}_2} \end{array} \right] \left[ \begin{array}{c} \mathbf{V}_{\mathbf{X}_1}^H \\ \mathbf{V}_{\mathbf{X}_2}^H \end{array} \right]$$

Assim, a matriz estendida de posto reduzido, restrita ao subespaço de sinal, é expressa apenas por:

$$\mathbf{X}_{Mod} = \mathbf{U}_{\mathbf{X}_1} \Sigma_{\mathbf{X}_1} \mathbf{V}_{\mathbf{X}_1}^H \quad (7.48)$$

Se particionarmos  $\mathbf{V}_{\mathbf{X}_1}$  como em (7.32), podemos escrever (7.48) como:

$$\begin{aligned} \mathbf{X}_{Mod} &= \mathbf{U}_{\mathbf{X}_1} \Sigma_{\mathbf{X}_1} \left[ \frac{\mathbf{g}^T}{\mathbf{V}_{\mathbf{X}_1}^H} \right]^H \\ &= \mathbf{U}_{\mathbf{X}_1} \Sigma_{\mathbf{X}_1} \left[ \mathbf{g}^* \mid \mathbf{V}_{\mathbf{X}_1}^{iH} \right] \\ &= \left[ \mathbf{U}_{\mathbf{X}_1} \Sigma_{\mathbf{X}_1} \mathbf{g}^* \mid \mathbf{U}_{\mathbf{X}_1} \Sigma_{\mathbf{X}_1} \mathbf{V}_{\mathbf{X}_1}^{iH} \right] \end{aligned} \quad (7.49)$$

Desse modo, podemos definir:

$$\begin{aligned} \mathbf{X}_{Mod} &= \left[ \mathbf{b}_{Mod} \mid \mathbf{A}_{Mod} \right] \\ &= \left[ \mathbf{U}_{\mathbf{X}_1} \Sigma_{\mathbf{X}_1} \mathbf{g}^* \mid \mathbf{U}_{\mathbf{X}_1} \Sigma_{\mathbf{X}_1} \mathbf{V}_{\mathbf{X}_1}^{iH} \right] \end{aligned} \quad (7.50)$$

É importante alertar que a matriz  $\mathbf{A}_{Mod}$  não é a mesma matriz  $\mathbf{A}_{LSr}$  da solução modificada proposta por Tufts e Kumaresan, a qual é obtida pelo truncamento da SVD da matriz  $\mathbf{A}$  apenas, como expresso em (6.2).

Recordando, a solução LS de norma mínima é expressa por:

$$\mathbf{w}'_{LS} = \mathbf{A}^{\#} \mathbf{b}$$

Se trocarmos  $\mathbf{A}$  e  $\mathbf{b}$  por  $\mathbf{A}_{Mod}$  e  $\mathbf{b}_{Mod}$  resulta:

$$\begin{aligned} \mathbf{w}'_{LSEq} &= \mathbf{A}_{Mod}^{\#} \mathbf{b}_{Mod} \\ &= (\mathbf{U}_{\mathbf{X}_1} \Sigma_{\mathbf{X}_1} \mathbf{V}_{\mathbf{X}_1}^{iH})^{\#} \mathbf{U}_{\mathbf{X}_1} \Sigma_{\mathbf{X}_1} \mathbf{g}^* \end{aligned} \quad (7.51)$$

Entretanto, antes de aplicar a pseudo-inversa em (7.51), devemos observar que a expressão de  $\mathbf{A}_{Mod}$  não é a sua SVD. Porém, podemos calcular sua pseudo-inversa a partir de [53]:

$$\mathbf{B}^\# = \mathbf{B}^H (\mathbf{B}\mathbf{B}^H)^\#$$

ou seja,

$$\mathbf{A}_{Mod}^\# = \mathbf{A}_{Mod}^H (\mathbf{A}_{Mod}\mathbf{A}_{Mod}^H)^\# \quad (7.52)$$

Calculando, encontramos:

$$\begin{aligned} \mathbf{A}_{Mod}^\# &= \mathbf{V}'_{X1} \Sigma_{X1} \mathbf{U}_{X1}^H (\mathbf{U}_{X1} \Sigma_{X1} \mathbf{V}'_{X1}{}^H \mathbf{V}'_{X1} \Sigma_{X1} \mathbf{U}_{X1}^H)^\# \\ &= \mathbf{V}'_{X1} \Sigma_{X1} \mathbf{U}_{X1}^H \mathbf{U}_{X1} (\Sigma_{X1} \mathbf{V}'_{X1}{}^H \mathbf{V}'_{X1} \Sigma_{X1})^{-1} \mathbf{U}_{X1}^H \\ &= \mathbf{V}'_{X1} \Sigma_{X1} \mathbf{U}_{X1}^H \mathbf{U}_{X1} \Sigma_{X1}^{-1} (\mathbf{V}'_{X1}{}^H \mathbf{V}'_{X1})^{-1} \Sigma_{X1}^{-1} \mathbf{U}_{X1}^H \\ &= \mathbf{V}'_{X1} (\mathbf{V}'_{X1}{}^H \mathbf{V}'_{X1})^{-1} \Sigma_{X1}^{-1} \mathbf{U}_{X1}^H \end{aligned} \quad (7.53)$$

Porém, de  $\mathbf{V}'_{X1}{}^H \mathbf{V}'_{X1} = \mathbf{I}$ , temos que:

$$\mathbf{V}'_{X1}{}^H \mathbf{V}'_{X1} = \mathbf{I} - \mathbf{g}^* \mathbf{g}^T \quad (7.54)$$

Então, (7.53) fica expressa por:

$$\mathbf{A}_{Mod}^\# = \mathbf{V}'_{X1} (\mathbf{I} - \mathbf{g}^* \mathbf{g}^T)^{-1} \Sigma_{X1}^{-1} \mathbf{U}_{X1}^H \quad (7.55)$$

Substituindo em (7.51), encontramos:

$$\begin{aligned} \mathbf{w}'_{LSEq} &= \mathbf{V}'_{X1} (\mathbf{I} - \mathbf{g}^* \mathbf{g}^T)^{-1} \Sigma_{X1}^{-1} \mathbf{U}_{X1}^H \mathbf{U}_{X1} \Sigma_{X1} \mathbf{g}^* \\ &= \mathbf{V}'_{X1} (\mathbf{I} - \mathbf{g}^* \mathbf{g}^T)^{-1} \mathbf{g}^* \end{aligned} \quad (7.56)$$

Entretanto, para qualquer vetor  $\mathbf{t}$ , a decomposição em autovalores e autovetores de  $(\mathbf{I} - \mathbf{t}^* \mathbf{t}^T)$  fornece o autovalor  $(1 - \mathbf{t}^H \mathbf{t})$  e seu autovetor correspondente  $\mathbf{t}^*$ , tal que:

$$(\mathbf{I} - \mathbf{t}^* \mathbf{t}^T)^{-1} \mathbf{t}^* = \frac{\mathbf{t}^*}{(1 - \mathbf{t}^H \mathbf{t})} \quad (7.57)$$

Então, utilizando (7.57), é possível afirmar que (7.56) pode ser escrita como:

$$\mathbf{w}'_{LSEq} = \frac{\mathbf{V}'_{X1} \mathbf{g}^*}{(1 - \mathbf{g}^H \mathbf{g})} \quad (7.58)$$

Desse modo, finalmente concluímos que:

$$\mathbf{w}'_{LSEq} = \mathbf{w}'_{TLS} \quad (7.59)$$

Logo, podemos interpretar o TLS como um LS onde a matriz de dados e o vetor desejado sofrem restrição conjunta ao subespaço de sinal, como em (7.50). Este detalhamento é mais uma de nossas contribuições e foi publicado em [58] e [60].

### 7.2.3 Solução FBLP Modificado Expressa em termos do TLS

Para expressarmos uma solução do tipo FBLP Modificado em termos da solução TLS, faz-se necessário primeiramente relacionarmos as SVDs da matriz de dados e da matriz de dados estendida. Propomos encontrar esta relação a partir de [60]:

$$\left( \left[ \frac{\mathbf{b}^H}{\mathbf{A}^H} \right] \left[ \mathbf{b} \mid \mathbf{A} \right] \right) \mathbf{v}_{\mathbf{X}(k)} = \sigma_{\mathbf{X}(k)}^2 \mathbf{v}_{\mathbf{X}(k)}, \quad k = 1, \dots, M \quad (7.60)$$

que nos permite escrever:

$$\left[ \frac{\mathbf{b}^H \mathbf{b} \mid \mathbf{b}^H \mathbf{A}}{\mathbf{A}^H \mathbf{b} \mid \mathbf{A}^H \mathbf{A}} \right] \mathbf{V}_{\mathbf{X1}} = \left[ \sigma_{\mathbf{X}(1)}^2 \mathbf{v}_{\mathbf{X}(1)} \mid \sigma_{\mathbf{X}(2)}^2 \mathbf{v}_{\mathbf{X}(2)} \mid \dots \mid \sigma_{\mathbf{X}(M)}^2 \mathbf{v}_{\mathbf{X}(M)} \right] \quad (7.61)$$

Reorganizando o termo à direita, temos:

$$\mathbf{V}_{\mathbf{X1}} = \mathbf{V}_{\mathbf{X1}} \Sigma_{\mathbf{X1}}^2 \quad (7.62)$$

Substituindo  $\mathbf{V}_{\mathbf{X1}}$ , encontramos:

$$\left[ \frac{\mathbf{b}^H \mathbf{b} \mid \mathbf{b}^H \mathbf{A}}{\mathbf{A}^H \mathbf{b} \mid \mathbf{A}^H \mathbf{A}} \right] \left[ \frac{\mathbf{g}^T}{\mathbf{V}'_{\mathbf{X1}}} \right] = \left[ \frac{\mathbf{g}^T}{\mathbf{V}'_{\mathbf{X1}}} \right] \Sigma_{\mathbf{X1}}^2 \quad (7.63)$$

Tomando a segunda linha de (7.63), podemos escrever:

$$\mathbf{A}^H \mathbf{b} \mathbf{g}^T + \mathbf{A}^H \mathbf{A} \mathbf{V}'_{\mathbf{X1}} = \mathbf{V}'_{\mathbf{X1}} \Sigma_{\mathbf{X1}}^2 \quad (7.64)$$

Isolando a primeira parcela de (7.64), temos:

$$\mathbf{A}^H \mathbf{b} \mathbf{g}^T = \mathbf{V}'_{\mathbf{X1}} \Sigma_{\mathbf{X1}}^2 - \mathbf{A}^H \mathbf{A} \mathbf{V}'_{\mathbf{X1}} \quad (7.65)$$

Daí, multiplicando por  $\mathbf{g}^*$ , resulta:

$$\mathbf{A}^H \mathbf{b} \mathbf{g}^T \mathbf{g}^* = \mathbf{V}'_{\mathbf{X1}} \Sigma_{\mathbf{X1}}^2 \mathbf{g}^* - \mathbf{A}^H \mathbf{A} \mathbf{V}'_{\mathbf{X1}} \mathbf{g}^* \quad (7.66)$$

ou seja:

$$\mathbf{A}^H \mathbf{b} \mathbf{g}^H \mathbf{g} = \mathbf{V}'_{\mathbf{X1}} \Sigma_{\mathbf{X1}}^2 \mathbf{g}^* - \mathbf{A}^H \mathbf{A} \mathbf{V}'_{\mathbf{X1}} \mathbf{g}^* \quad (7.67)$$

Assim, temos:

$$\begin{aligned}
\mathbf{A}^H \mathbf{b} &= \frac{\mathbf{V}'_{\mathbf{x}_1} \Sigma_{\mathbf{x}_1}^2 \mathbf{g}^*}{\mathbf{g}^H \mathbf{g}} - \frac{\mathbf{A}^H \mathbf{A} \mathbf{V}'_{\mathbf{x}_1} \mathbf{g}^*}{\mathbf{g}^H \mathbf{g}} \\
&= \frac{1 - \mathbf{g}^H \mathbf{g}}{\mathbf{g}^H \mathbf{g}} \left( \frac{\mathbf{V}'_{\mathbf{x}_1} \Sigma_{\mathbf{x}_1}^2 \mathbf{g}^*}{1 - \mathbf{g}^H \mathbf{g}} - \mathbf{A}^H \mathbf{A} \frac{\mathbf{V}'_{\mathbf{x}_1} \mathbf{g}^*}{1 - \mathbf{g}^H \mathbf{g}} \right) \\
&= \frac{(1 - \mathbf{g}^H \mathbf{g})}{\mathbf{g}^H \mathbf{g}} \left( \frac{\mathbf{V}'_{\mathbf{x}_1} \Sigma_{\mathbf{x}_1}^2 \mathbf{g}^*}{1 - \mathbf{g}^H \mathbf{g}} - \mathbf{A}^H \mathbf{A} \mathbf{w}'_{TLS} \right) \tag{7.68}
\end{aligned}$$

Isto nos permite escrever a solução  $\mathbf{w}'_{LS} = (\mathbf{A}^H \mathbf{A})^\# \mathbf{A}^H \mathbf{b}$  na forma:

$$\mathbf{w}'_{LS} = \frac{1 - \mathbf{g}^H \mathbf{g}}{\mathbf{g}^H \mathbf{g}} \left\{ (\mathbf{A}^H \mathbf{A})^\# \frac{\mathbf{V}'_{\mathbf{x}_1} \Sigma_{\mathbf{x}_1}^2 \mathbf{g}^*}{1 - \mathbf{g}^H \mathbf{g}} - \mathbf{w}'_{TLS} \right\} \tag{7.69}$$

ou seja:

$$\mathbf{w}'_{LS} = \frac{\mathbf{g}^H \mathbf{g} - 1}{\mathbf{g}^H \mathbf{g}} \left\{ \mathbf{w}'_{TLS} - (\mathbf{A}^H \mathbf{A})^\# \frac{\mathbf{V}'_{\mathbf{x}_1} \Sigma_{\mathbf{x}_1}^2 \mathbf{g}^*}{1 - \mathbf{g}^H \mathbf{g}} \right\} \tag{7.70}$$

Analisando agora a solução do tipo FBLP Modificado, na qual truncamos a SVD apenas da matriz de dados  $\mathbf{A}$ , verificamos que ela pode ser escrita como [53]:

$$\begin{aligned}
\mathbf{w}'_{LSr} &= \mathbf{A}_{LSr}^\# \mathbf{b} \\
&= (\mathbf{V}_{\mathbf{A}_1} \Sigma_{\mathbf{A}_1}^{-1} \mathbf{U}_{\mathbf{A}_1}^H) \mathbf{b} \\
&= (\mathbf{A}_{LSr}^H \mathbf{A}_{LSr})^\# \mathbf{A}_{LSr}^H \mathbf{b} \tag{7.71}
\end{aligned}$$

Porém, desenvolvendo a forma alternativa a seguir:

$$\begin{aligned}
\mathbf{w}'_{Alt.} &= (\mathbf{A}_{LSr}^H \mathbf{A}_{LSr})^\# \mathbf{A}^H \mathbf{b} \\
&= (\mathbf{V}_{\mathbf{A}_1} \Sigma_{\mathbf{A}_1} \mathbf{U}_{\mathbf{A}_1}^H \mathbf{U}_{\mathbf{A}_1} \Sigma_{\mathbf{A}_1} \mathbf{V}_{\mathbf{A}_1}^H)^\# \mathbf{A}^H \mathbf{b} \\
&= (\mathbf{V}_{\mathbf{A}_1} \Sigma_{\mathbf{A}_1}^2 \mathbf{V}_{\mathbf{A}_1}^H)^\# \mathbf{A}^H \mathbf{b} \\
&= \left( \left[ \mathbf{V}_{\mathbf{A}_1} \mid \mathbf{0} \right] \begin{bmatrix} \Sigma_{\mathbf{A}_1}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{\mathbf{A}_1}^H \\ \mathbf{0} \end{bmatrix} \right)^\# \left[ \mathbf{V}_{\mathbf{A}_1} \mid \mathbf{V}_{\mathbf{A}_2} \right] \begin{bmatrix} \Sigma_{\mathbf{A}_1} & \mathbf{0} \\ \mathbf{0} & \Sigma_{\mathbf{A}_2} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{\mathbf{A}_1}^H \\ \mathbf{U}_{\mathbf{A}_2}^H \end{bmatrix} \mathbf{b} \\
&= \left[ \mathbf{V}_{\mathbf{A}_1} \mid \mathbf{0} \right] \begin{bmatrix} \Sigma_{\mathbf{A}_1}^{-2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{\mathbf{A}_1}^H \\ \mathbf{0} \end{bmatrix} \left[ \mathbf{V}_{\mathbf{A}_1} \mid \mathbf{V}_{\mathbf{A}_2} \right] \begin{bmatrix} \Sigma_{\mathbf{A}_1} & \mathbf{0} \\ \mathbf{0} & \Sigma_{\mathbf{A}_2} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{\mathbf{A}_1}^H \\ \mathbf{U}_{\mathbf{A}_2}^H \end{bmatrix} \mathbf{b} \\
&= \left[ \mathbf{V}_{\mathbf{A}_1} \mid \mathbf{0} \right] \begin{bmatrix} \Sigma_{\mathbf{A}_1}^{-2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{\mathbf{A}_1}^H \mathbf{V}_{\mathbf{A}_1} & \mathbf{V}_{\mathbf{A}_1}^H \mathbf{V}_{\mathbf{A}_2} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Sigma_{\mathbf{A}_1} & \mathbf{0} \\ \mathbf{0} & \Sigma_{\mathbf{A}_2} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{\mathbf{A}_1}^H \\ \mathbf{U}_{\mathbf{A}_2}^H \end{bmatrix} \mathbf{b} \\
&= \left[ \mathbf{V}_{\mathbf{A}_1} \mid \mathbf{0} \right] \begin{bmatrix} \Sigma_{\mathbf{A}_1}^{-2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Sigma_{\mathbf{A}_1} & \mathbf{0} \\ \mathbf{0} & \Sigma_{\mathbf{A}_2} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{\mathbf{A}_1}^H \\ \mathbf{U}_{\mathbf{A}_2}^H \end{bmatrix} \mathbf{b}
\end{aligned}$$

$$\begin{aligned}
&= \left[ \mathbf{V}_{A1} \mid \mathbf{0} \right] \left[ \begin{array}{c|c} \Sigma_{A1}^{-2} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array} \right] \left[ \begin{array}{c|c} \Sigma_{A1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array} \right] \left[ \begin{array}{c} \mathbf{U}_{A1}^H \\ \hline \mathbf{U}_{A2}^H \end{array} \right] \mathbf{b} \\
&= \left[ \mathbf{V}_{A1} \mid \mathbf{0} \right] \left[ \begin{array}{c|c} \Sigma_{A1}^{-1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array} \right] \left[ \begin{array}{c} \mathbf{U}_{A1}^H \\ \hline \mathbf{U}_{A2}^H \end{array} \right] \mathbf{b} \\
&= \left[ \mathbf{V}_{A1} \mid \mathbf{0} \right] \left[ \begin{array}{c} \Sigma_{A1}^{-1} \mathbf{U}_{A1}^H \\ \hline \mathbf{0} \end{array} \right] \mathbf{b} \\
&= \left( \mathbf{V}_{A1} \Sigma_{A1}^{-1} \mathbf{U}_{A1}^H \right) \mathbf{b} \\
&= \mathbf{w}'_{LSr}
\end{aligned} \tag{7.72}$$

Logo, podemos escrever  $\mathbf{w}'_{LSr}$  alternativamente como:

$$\mathbf{w}'_{LSr} = \left( \hat{\mathbf{A}}_{LSr}^H \hat{\mathbf{A}}_{LSr} \right)^\# \mathbf{A}^H \mathbf{b} \tag{7.73}$$

Assim, calculamos uma expressão para  $\mathbf{w}'_{LSr}$  em termos de  $\mathbf{w}'_{TLS}$  a partir de (7.68):

$$\begin{aligned}
\mathbf{w}'_{LSr} &= \frac{\mathbf{g}^H \mathbf{g} - 1}{\mathbf{g}^H \mathbf{g}} \left\{ \left[ \left( \hat{\mathbf{A}}_{LSr}^H \hat{\mathbf{A}}_{LSr} \right)^\# \mathbf{A}^H \right] \mathbf{A} \mathbf{w}'_{TLS} - \left( \hat{\mathbf{A}}_{LSr}^H \hat{\mathbf{A}}_{LSr} \right)^\# \frac{\mathbf{V}'_{X1} \Sigma_{X1}^2 \mathbf{g}^*}{1 - \mathbf{g}^H \mathbf{g}} \right\} \\
&= \frac{\mathbf{g}^H \mathbf{g} - 1}{\mathbf{g}^H \mathbf{g}} \left\{ \left( \mathbf{V}_{A1} \Sigma_{A1}^{-1} \mathbf{U}_{A1}^H \right) \mathbf{A} \mathbf{w}'_{TLS} - \left( \hat{\mathbf{A}}_{LSr}^H \hat{\mathbf{A}}_{LSr} \right)^\# \frac{\mathbf{V}'_{X1} \Sigma_{X1}^2 \mathbf{g}^*}{1 - \mathbf{g}^H \mathbf{g}} \right\} \\
&= \frac{\mathbf{g}^H \mathbf{g} - 1}{\mathbf{g}^H \mathbf{g}} \left\{ \mathbf{V}_{A1} \Sigma_{A1}^{-1} \mathbf{U}_{A1}^H \left[ \mathbf{U}_{A1} \mid \mathbf{U}_{A2} \right] \left[ \begin{array}{c|c} \Sigma_{A1} & \mathbf{0} \\ \hline \mathbf{0} & \Sigma_{A2} \end{array} \right] \left[ \begin{array}{c} \mathbf{V}_{A1}^H \\ \hline \mathbf{V}_{A2}^H \end{array} \right] \mathbf{w}'_{TLS} \right. \\
&\quad \left. - \left( \hat{\mathbf{A}}_{LSr}^H \hat{\mathbf{A}}_{LSr} \right)^\# \frac{\mathbf{V}'_{X1} \Sigma_{X1}^2 \mathbf{g}^*}{1 - \mathbf{g}^H \mathbf{g}} \right\}
\end{aligned} \tag{7.74}$$

que, conforme o desenvolvimento em (7.72), pode ser escrita como:

$$\begin{aligned}
\mathbf{w}'_{LSr} &= \frac{\mathbf{g}^H \mathbf{g} - 1}{\mathbf{g}^H \mathbf{g}} \left\{ \mathbf{V}_{A1} \Sigma_{A1}^{-1} \mathbf{U}_{A1}^H \mathbf{U}_{A1} \Sigma_{A1} \mathbf{V}_{A1}^H \mathbf{w}'_{TLS} - \left( \hat{\mathbf{A}}_{LSr}^H \hat{\mathbf{A}}_{LSr} \right)^\# \frac{\mathbf{V}'_{X1} \Sigma_{X1}^2 \mathbf{g}^*}{1 - \mathbf{g}^H \mathbf{g}} \right\} \\
&= \frac{\mathbf{g}^H \mathbf{g} - 1}{\mathbf{g}^H \mathbf{g}} \left\{ \left( \mathbf{V}_{A1} \mathbf{V}_{A1}^H \right) \mathbf{w}'_{TLS} - \left( \mathbf{V}_{A1}^H \Sigma_{A1}^{-2} \mathbf{V}_{A1} \right) \frac{\mathbf{V}'_{X1} \Sigma_{X1}^2 \mathbf{g}^*}{1 - \mathbf{g}^H \mathbf{g}} \right\}
\end{aligned} \tag{7.75}$$

Portanto, obtivemos uma nova forma de demonstrar que as estimativas dos métodos TLS-FBLP e FBLP Modificado são diferentes [60].

### 7.3 Desempenhos em Perspectiva

Visando comparar o desempenho dos métodos TLS-FBLP e FBLP Modificado, efetuamos o cálculo das curvas de desempenho para todas as ordens de predição cabíveis, utilizando

as expressões (6.5) e (7.36). Apresentamos estas curvas de forma simultânea, através das superfícies de desempenho mostradas nas figuras 7.4 e 7.5, as quais permitem acompanhar a variação da performance dos métodos em função da ordem de predição e da SNR. Destacamos, em ambas, através de linhas descendentes, os níveis de desempenho para cada valor de ordem de predição  $L$ , desde  $L = 2$  até  $L = 24$ . As linhas transversais foram traçadas almejando, tão somente, melhorar o aspecto das figuras. Estas superfícies permitem verificar onde ocorre o melhor desempenho, o pior ou então acompanhar a variação das SNR's de limiar pelo formato da borda da superfície a partir da qual o desempenho sofre maior degradação.

Através das figuras 7.4 e 7.5, podemos observar que os desempenhos gerais do TLS-FBLP e do FBLP Modificado, para grandes valores de SNR, oscilam conforme a ordem de predição, sendo melhores para as ordens ímpares e piores para as pares. Devemos observar que no caso Kumaresan-Prony (KP)[91],  $L = 24$ , tanto o FBLP Modificado quanto o TLS-FBLP, fornecem a mesma curva de desempenho. Na verdade, as curvas se assemelham a partir da ordem 20, embora com algumas diferenças sutis. Esta semelhança se deve ao fato de que para  $(L + 1) > 2(N - L)$ , os critérios TLS e LS fornecem os mesmos vetores de coeficientes, como apontado na Subseção 7.1.2. Entretanto, a restrição ao subespaço de sinal, empregada no FBLP Modificado, altera a matriz de dados, produzindo as diferenças observadas.

Nas figuras 7.6 e 7.7 traçamos também as curvas de nível para as superfícies apresentadas, gerando alternativas para a visualização do contorno de limiar. Nas regiões onde as superfícies são mais íngremes, as linhas ficam mais próximas entre si, definindo o limiar de desempenho e, conseqüentemente, as respectivas SNR's de limiar para cada ordem de predição.

No TLS-FBLP verificamos que o desempenho acima do limiar varia mais suavemente com a ordem e que o contorno de limiar é mais suave, ou seja, a variação dos valores de SNR de limiar é muito pequena para ordens  $L$  de 8 a 18. Neste caso, o menor valor de SNR de limiar encontrado foi 8 dB e ocorreu em  $L = 16$ , conforme pode ser observado na figura 7.6. Observando o FBLP Modificado, é fácil constatar que o contorno de limiar é mais agudo, bem como o melhor desempenho ocorre na região de  $L = 14$  a  $L = 18$ , piorando bastante fora deste intervalo. Para ordens pequenas verificamos vales profundos mesmo para grandes SNR's. Pela figura 7.7 percebemos que a menor SNR de limiar para o FBLP Modificado tem o valor de 7dB e verifica-se para  $L = 17$ , diferente da ordem ótima proposta por Tufts e Kumaresan,  $L = (3/4)N$ , que equivale a  $L = 18$  neste caso. Como dito antes, acreditamos que Tufts e Kumaresan simularam apenas ordens pares.

Observamos que o melhor desempenho, tanto para o TLS-FBLP quanto para o FBLP Modificado, ocorreu quando a ordem de predição empregada foi tal que as respectivas matrizes de dados tornaram-se aproximadamente quadradas, ou seja, para  $L = 16$  temos  $\mathbf{X}$  de dimensões  $18 \times 17$ , enquanto que  $\mathbf{A}$  tem dimensões  $16 \times 17$  para  $L = 17$ . Assim, propomos a medida analítica que afirma que a ordem ótima é aquela que mais aproxima a matriz de dados da forma quadrada, em substituição à medida empírica assinalada por Tufts e Kumaresan para a ordem de menor SNR de limiar do FBLP Modificado. Para os casos FLP e BLP, a ordem ótima corresponde àquela mais próxima da metade do número de amostras, arredondando-se para a imediatamente superior no caso de  $N$  ser ímpar.

Na figura 7.8 destacamos as curvas correspondentes aos menores valores de SNR de limiar para os métodos TLS-FBLP e FBLP Modificado, ou seja, relativas aos melhores desempenhos.

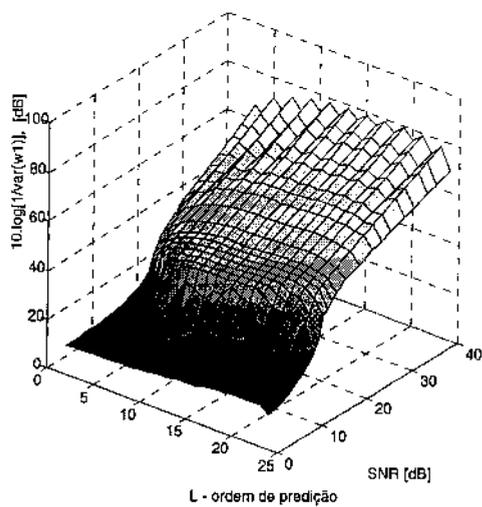


Figura 7.4: Superfície de desempenho para o método TLS-FBLP.

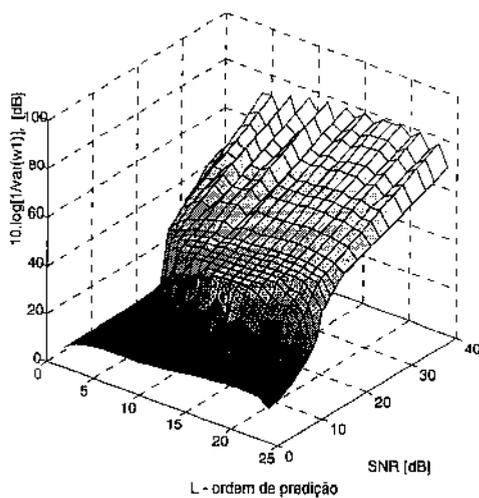


Figura 7.5: Superfície de desempenho para o método FBLP Modificado.

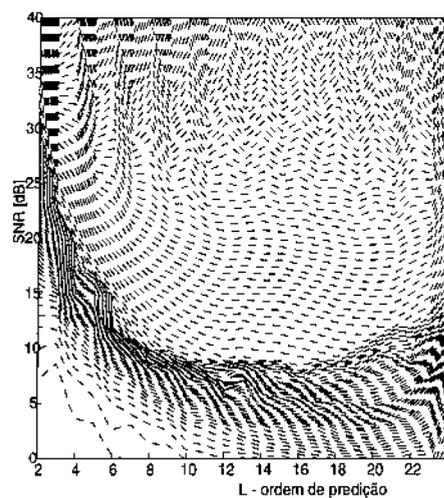


Figura 7.6: Linhas de nível para a superfície de desempenho TLS-FBLP (vista de topo).

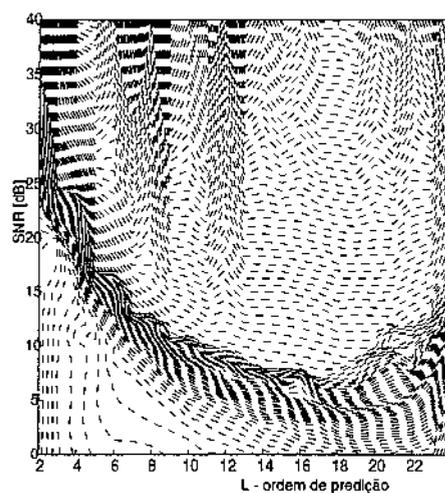


Figura 7.7: Linhas de nível para a superfície de desempenho FBLP Modificado(vista de topo).

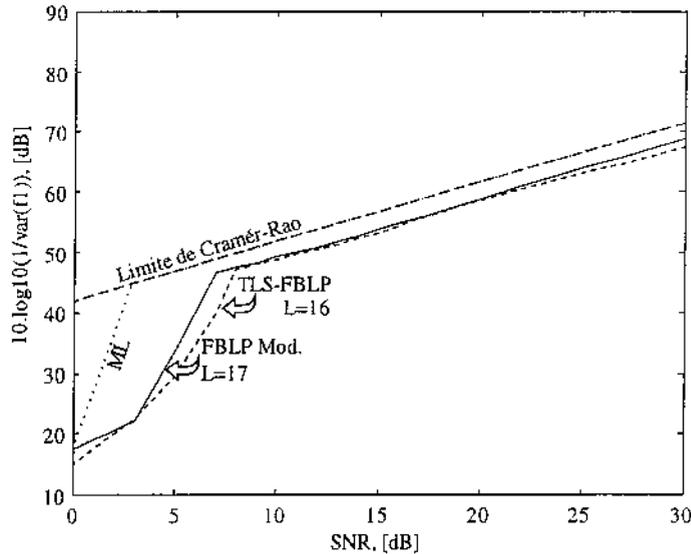


Figura 7.8: Comparação entre os melhores desempenhos TLS-FBLP e FBLP Modificado.

Portanto, o melhor desempenho absoluto foi verificado no método FBLP Modificado de Tufts e Kumaresan. Entretanto, pelas figuras 7.6, 7.7 e 7.8, fica patente que a diferença é pequena e que o desempenho do método TLS-FBLP é mais robusto à variação da ordem de predição  $L$ . Estes fatos permitem afirmar que o desempenho do TLS-FBLP é significativamente superior ao do FBLP convencional, sendo, sim, comparável ao do FBLP Modificado.

Comparando qualitativamente a complexidade computacional dos métodos TLS-FBLP e FBLP Modificado, verificamos que a aplicação da SVD sobre a matriz  $\mathbf{X}$  sempre envolve mais cálculos do que a SVD de  $\mathbf{A}$ , pois  $\mathbf{X}$  tem uma coluna a mais do que  $\mathbf{A}$ . Diferentemente, a obtenção da solução  $\mathbf{w}'$  depende das dimensões dos subespaços de sinal e de ruído. Para  $M < \min(2(N-L), L)$ , com a forma direta de cálculo das soluções TLS, podemos utilizar o subespaço de sinal tanto para obter  $\mathbf{w}'_{LSr}$  quanto  $\mathbf{w}'_{TLS}$ . Então, podemos dizer que, neste caso, as complexidades computacionais dos métodos TLS-FBLP e FBLP Modificado para a obtenção de  $\mathbf{w}'$  são equivalentes. Por outro lado, para  $M > \min(2(N-L), L)$ , o subespaço de ruído é menor do que o de sinal. Neste caso, podemos utilizar a fórmula direta da solução TLS pelo subespaço de ruído, que envolve menos cálculos do que a solução FBLP Modificado, restrita ao subespaço de sinal. Portanto, acreditamos que a complexidade computacional do TLS-FBLP é equivalente à do FBLP Modificado.

Estas análises e comparações nos permitem afirmar que o método TLS-FBLP revelou-se competitivo com o FBLP Modificado para a estimação de frequências. Entretanto, podemos observar que os desempenhos destes métodos ainda estão muito distantes da Máxima Verossimilhança (ML), necessitando reduzir a SNR de limiar em pelo menos 4 dB. Observamos também que, contrariando as nossas expectativas, o desempenho do TLS-FBLP mostrou-se ligeiramente inferior ao do FBLP Modificado, mesmo dispondo de maior grau de liberdade para ajuste da

otimização. Estas observações podem ser esclarecidas a partir do fato que as soluções LS e TLS não preservam a estrutura, como a inserida pelo filtro de predição linear na construção da matriz de dados estendida.

Tanto o LS quanto o TLS pressupõem que as componentes de ruído distribuem-se idêntica e independentemente pelas matrizes  $\Delta\mathbf{A}$  e  $\Delta\mathbf{X}$ , respectivamente. Isto quer dizer que se supõe que estas matrizes não são estruturadas. A prova disso é que as aproximações de posto reduzido  $\mathbf{A}_{LSr}$  e  $\mathbf{X}_{Mod}$ , obtidas pelo truncamento da SVD, não apresentam qualquer estrutura, mesmo que  $\mathbf{A}$  e  $\mathbf{X}$  a possuam. Isto ocorre justamente porque o truncamento da SVD de uma matriz não preserva sua estrutura. Por essa razão, os desempenhos desses métodos distanciam-se daquela da Máxima Verossimilhança.

O método TLS-FBLP é mais fortemente afetado pela existência de estrutura, pois altera tanto  $\mathbf{A}$  quanto  $\mathbf{b}$  no cálculo de sua solução, enquanto que o FBLP Modificado altera somente a matriz  $\mathbf{A}$ , preservando a estrutura de  $\mathbf{b}$ .

## 7.4 Conclusões

Neste capítulo propusemos o método TLS-FBLP, formulado pelos subespaços de sinal ou de ruído. Demostramos que o TLS pode ser interpretado como um LS modificado onde a matriz de dados e o vetor desejado sofrem restrição conjunta ao subespaço de sinal. Também relacionamos analiticamente as expressões de solução dos métodos FBLP Modificado e TLS-FBLP. Tínhamos esperanças de superar o desempenho do método FBLP Modificado, devido à maior abrangência do ajuste TLS em relação ao LS. Porém, verificamos que o desempenho do TLS-FBLP, embora competitivo com o do FBLP Modificado, revelou-se ligeiramente inferior a este último. Além disso, tanto o TLS-FBLP quanto o FBLP Modificado apresentaram SNRs de limiar cerca de 4 dB acima à da Máxima Verossimilhança. Isto se deve ao truncamento da SVD não preservar a informação de estrutura da matriz de dados. No capítulo seguinte discutiremos avanços recentes na formulação TLS que permitem levar em conta esta estrutura no processo de otimização, alcançando o desempenho ML.

# Capítulo 8

## Métodos TLS com Desempenho ML

Nos capítulos anteriores, verificamos que métodos de alta resolução como o FBLP Modificado e o TLS-FBLP, provêem excelentes desempenhos em relações sinal-ruído (SNR) grandes, acompanhando o limite de Cramér-Rao, mesmo com poucas amostras disponíveis de dados. Entretanto, seus desempenhos sofrem severas degradações quando a SNR assume valores abaixo de 7 dB, ainda significativamente superiores à SNR de limiar em 3 dB, apresentada pela estimação ótima de Máxima Verossimilhança (ML). Embora o critério de ML forneça o melhor desempenho, a maximização direta da função de verossimilhança requer uma busca multidimensional de custo computacional proibitivo.

Neste capítulo, apresentamos a comparação entre alguns métodos recentes de estimação de parâmetros, como o método Máxima Verossimilhança Quadrática Iterativa (IQML) [6], o método Mínimos Quadrados Totais Restrito (CTLS) [2] e o método Mínimos Quadrados Totais Branqueado (WTLS) [39]. O IQML propõe uma expressão para o critério de máxima verossimilhança em termos do polinômio de predição do sinal isento de ruído [41]. Os métodos CTLS e WTLS consistem em melhorias à solução clássica do TLS para o caso de existirem dependências lineares entre as linhas da matriz de dados estendida.

Diferentemente dos desempenhos do FBLP Modificado e do TLS-FBLP, os desempenhos do CTLS e do WTLS acompanham o da ML mesmo em SNR pequenas. No sentido de esclarecer esse comportamento, decidimos comparar analiticamente estes três métodos. Dessa forma, ao longo do capítulo, demonstraremos que o CTLS, o WTLS e o IQML minimizam a norma do mesmo erro, produzindo soluções equivalentes. Esta equivalência não é descrita na literatura nem com a abrangência, nem com o detalhamento e didática que procuramos empregar neste capítulo. Assim, consideramos esta mais uma contribuição à compreensão do poder desses métodos para a estimação de parâmetros.

### 8.1 Máxima Verossimilhança Quadrática Iterativa - IQML

Recordando o a expressão (2.21), as estimativas ML das frequências  $\omega_i$  do sinal  $x[n]$  em (4.27), são dadas pela solução do seguinte problema não-linear de mínimos quadrados:

$$\min_{\mathbf{E}} \|\mathbf{x} - \mathbf{E}\hat{\mathbf{a}}_c\|_2^2 = \min_{\mathbf{E}} \left\{ \mathbf{x}^H \left[ \mathbf{I} - \mathbf{E} (\mathbf{E}^H \mathbf{E})^{-1} \mathbf{E}^H \right] \mathbf{x} \right\} \quad (8.1)$$

onde  $\mathbf{x}$  é um vetor  $N \times 1$  de dados corrompido por um vetor  $\mathbf{y}$ ,  $N \times 1$ , de ruído branco gaussiano complexo com partes real e imaginária descorrelacionadas, cada uma com variância  $\sigma_y^2/2$ ;  $\mathbf{E}$  é uma matriz  $N \times M$  de Vandermonde sobre  $\exp[j\omega_i n]$  e  $\hat{\mathbf{a}}_c$  é um vetor  $M \times 1$  de amplitudes complexas

Entretanto, como (8.1) é altamente não-linear nas frequências  $\{\omega_1, \dots, \omega_M\}$ , minimizá-la exigiria uma busca sobre um espaço  $M$ -dimensional, computacionalmente dispendiosa para valores grandes de  $M$ .

Numa tentativa de substituir o critério de ML (8.1) por expressões mais tratáveis, muitos pesquisadores propuseram transformar o espaço de parâmetros, de frequências para parâmetros AR. Isto é possível a partir da observação de que, como uma soma de exponenciais complexas pode ser modelada como a resposta ao impulso de um filtro causal de pólos e zeros contendo pólos na circunferência de raio unitário (CRU), podemos equivalentemente determinar as frequências a partir da posição angular desses pólos. A localização dos pólos, por sua vez, é função dos parâmetros AR. Para compreender esta transformação, recordemos o polinômio associado ao filtro de predição linear:

$$W(z) = w_0 + w_1 z^{-1} + \dots + w_L z^{-L}, \quad \text{para } M \leq L \leq (N - M) \quad (8.2)$$

Na ausência de ruído, conforme o teorema 4.1 do Capítulo 4,  $W(z)$  tem raízes na CRU em  $z_i = \exp[j\omega_i]$ , para  $i = 1, \dots, M$ , isto é:

$$W(\exp[j\omega_i]) = \sum_{k=0}^L w_k \exp[-j\omega_i k] = 0, \quad i = 1, \dots, M \quad (8.3)$$

Multiplicando por  $\exp[j\omega_i L]$  e realizando uma mudança na variável de índice, encontramos:

$$W(\exp[j\omega_i]) = \sum_{k=0}^L w_{L-k} \exp[j\omega_i k] = 0, \quad i = 1, \dots, M \quad (8.4)$$

Então, definindo a matriz  $(N - L) \times N$ :

$$\mathbf{W}^H = \begin{bmatrix} w_L & w_{L-1} & \dots & w_0 & 0 & \dots & 0 \\ 0 & w_L & w_{L-1} & \dots & w_0 & \dots & 0 \\ \vdots & & \ddots & \ddots & & \ddots & \vdots \\ 0 & 0 & \dots & w_L & w_{L-1} & \dots & w_0 \end{bmatrix} \quad (8.5)$$

temos que (8.4) dá lugar a:

$$\mathbf{W}^H \mathbf{E} = \mathbf{0} \quad (8.6)$$

Pela teoria das projeções, podemos verificar que o operador  $\mathbf{E} (\mathbf{E}^H \mathbf{E})^{-1} \mathbf{E}^H$  projeta um vetor qualquer no espaço gerado pelas colunas de  $\mathbf{E}$ , enquanto que  $\mathbf{W} (\mathbf{W}^H \mathbf{W})^{-1} \mathbf{W}^H$  projeta um vetor no espaço gerado pelas colunas de  $\mathbf{W}$ . Como, a partir de (8.6), podemos afirmar que

o espaço das colunas de  $\mathbf{W}$  é o complemento ortogonal do espaço das colunas de  $\mathbf{E}$ , concluímos que, para  $L = M$ , a soma desses dois operadores produz o operador identidade [41]:

$$\mathbf{E} (\mathbf{E}^H \mathbf{E})^{-1} \mathbf{E}^H + \mathbf{W} (\mathbf{W}^H \mathbf{W})^{-1} \mathbf{W}^H = \mathbf{I} \quad (8.7)$$

Utilizando (8.7) em (8.1) resulta que:

$$\min_{\mathbf{E}} \|\mathbf{x} - \mathbf{E} \hat{\mathbf{a}}_c\|_2^2 = \min_{\mathbf{W}} \left\{ \mathbf{x}^H \mathbf{W} (\mathbf{W}^H \mathbf{W})^{-1} \mathbf{W}^H \mathbf{x} \right\} \quad (8.8)$$

Restringindo-nos ao caso *forward*, de (4.13) e (4.15) recordamos a matriz de dados estendida  $\mathbf{X} = \mathbf{X}_f$ :

$$\mathbf{X} = \begin{bmatrix} x[L] & x[L-1] & \cdots & x[0] \\ x[L+1] & x[L] & \cdots & x[1] \\ \vdots & \vdots & \ddots & \vdots \\ x[N-1] & x[N-2] & \cdots & x[N-L-1] \end{bmatrix} \quad (8.9)$$

e definimos o vetor completo de coeficientes de predição  $\mathbf{w} = [w_0 \ w_1 \ \cdots \ w_L]^T$ , que inclui também  $w_0$ . A partir da comutatividade da operação de convolução, podemos afirmar que:

$$\mathbf{W}^H \mathbf{x} = \mathbf{X} \mathbf{w} \quad (8.10)$$

Então, substituindo (8.10) em (8.8), finalmente obtemos:

$$\min_{\mathbf{E}} \|\mathbf{x} - \mathbf{E} \hat{\mathbf{a}}_c\|_2^2 = \min_{\mathbf{w}} \left\{ \mathbf{w}^H \mathbf{X}^H (\mathbf{W}^H \mathbf{W})^{-1} \mathbf{X} \mathbf{w} \right\} \quad (8.11)$$

Este problema quadrático deve ser minimizado sob um conjunto de restrições de forma que os pólos se posicionem sobre a CRU para produzir as estimativas ML das frequências. Infelizmente, esta minimização com restrições é difícil de ser efetuada diretamente. Porém, observando que  $\mathbf{W} (\mathbf{W}^H \mathbf{W})^{-1} \mathbf{W}^H$  é Hermitiana e idempotente [41], podemos escrever (8.8) como:

$$\min_{\mathbf{E}} \|\mathbf{x} - \mathbf{E} \hat{\mathbf{a}}_c\|_2^2 = \min_{\mathbf{w}} \left\| \mathbf{W} (\mathbf{W}^H \mathbf{W})^{-1} \mathbf{W}^H \mathbf{x} \right\|_2^2 \quad (8.12)$$

que, substituindo (8.10), fica expressa por:

$$\begin{aligned} \min_{\mathbf{E}} \|\mathbf{x} - \mathbf{E} \hat{\mathbf{a}}_c\|_2^2 &= \min_{\mathbf{w}} \left\| \mathbf{W} (\mathbf{W}^H \mathbf{W})^{-1} \mathbf{X} \mathbf{w} \right\|_2^2 \\ &= \min_{\mathbf{w}} \left\| (\mathbf{W}^H)^{\#} \mathbf{X} \mathbf{w} \right\|_2^2 \end{aligned} \quad (8.13)$$

Esta formulação sugere que o problema seja resolvido de maneira iterativa através de:

$$\min_{\mathbf{w}^{(k)}} \left\| (\mathbf{W}^{(k)H})^{\#} \mathbf{X} \mathbf{w}^{(k+1)} \right\|_2^2 \quad (8.14)$$

Esta expressão é utilizada em métodos iterativos como o de Evans e Fischl [25], o método KiSS [49], o IFA [41] e o IQML [6]. As diferenças que distinguem estes métodos entre si, surgem da aplicação de diferentes restrições sobre  $\mathbf{w}$  bem como da utilização de algoritmos distintos para assegurar a estabilidade dos mesmos. Propomos minimizar a expressão anterior iterando a solução TLS para  $\mathbf{w}^{(k+1)}$  seguida da atualização de  $\mathbf{W}^{(k+1)H}$  até a convergência de  $\mathbf{w}$ . Concentrando-nos no método IQML, além da restrição de não-trivialidade,  $\{\mathbf{w} : \mathbf{w} \neq \mathbf{0}\}$  (dada por  $w_0 = 1$ ,  $w_0 = -1$  ou  $\mathbf{w}^H \mathbf{w} = 1$ ), que assegura a invertibilidade de  $(\mathbf{W}^H \mathbf{W})$ , utiliza-se  $L = M$  por construção e aplicam-se sobre  $\mathbf{w}$  as seguintes restrições conforme as características dos parâmetros a serem estimados:

1. Restrição de estabilidade, para o caso de exponenciais complexas amortecidas:

$$\{\mathbf{w} : \mathbf{w} \text{ tem raízes dentro ou sobre a CRU}\};$$

2. Restrição para o caso de exponenciais complexas não-amortecidas:

$$\{\mathbf{w} : w_i = w_{(L-i)}^*, \quad i = 0, \dots, L\};$$

3. Restrição para o caso de senóides reais amortecidas:

$$\{\mathbf{w} : \text{Im}(\mathbf{w}) = \mathbf{0}\};$$

4. Restrição para o caso de senóides reais não-amortecidas:

$$\{\mathbf{w} : \text{Im}(\mathbf{w}) = \mathbf{0} \text{ e } w_i = w_{(L-i)}, \quad i = 0, \dots, L\}.$$

Devemos observar que a restrição 2 não pode ser atendida se impusermos  $w_0 = 1$ . Assim, adotaremos  $\mathbf{w}^H \mathbf{w} = 1$  no caso de exponenciais complexas não-amortecidas. A restrição 2, de simetria dos coeficientes do polinômio  $W(z)$ , pode ser implementada explicitamente como uma restrição quadrática sobre  $\mathbf{w}$  ou, de forma implícita, sobre  $\mathbf{X}$ . A implementação implícita da restrição de simetria converte o problema atual em outro sem restrições e reduz seu tamanho pela metade. A formulação da restrição 2 pode seguir por dois caminhos, conforme a paridade de  $L$ , que aqui serão apresentados separadamente.

Para  $L$  ímpar, seja  $L = 2q + 1$ . Particione  $\mathbf{w}$  e  $\mathbf{X}$  de forma que:

$$\mathbf{w} = \begin{bmatrix} w_0 & \cdots & w_q & w_{q+1} & \cdots & w_{2q+1} \end{bmatrix}^T = \begin{bmatrix} \mathbf{w}_1^T & | & \mathbf{w}_2^T \end{bmatrix}^T \quad (8.15)$$

e

$$\mathbf{X}\mathbf{w} = \begin{bmatrix} \mathbf{X}_1 & | & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix} = \mathbf{X}_1 \mathbf{w}_1 + \mathbf{X}_2 \mathbf{w}_2 \quad (8.16)$$

Utilizando a matriz  $\tilde{\mathbf{I}}$  de reflexão lateral dada por:

$$\tilde{\mathbf{I}} = \begin{bmatrix} 0 & 0 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \end{bmatrix} \quad (8.17)$$

a relação de simetria sobre  $\mathbf{w}$  pode ser representada por  $\mathbf{w} = \tilde{\mathbf{I}} \mathbf{w}^*$ , ou seja:

$$\mathbf{w}_2 = \tilde{\mathbf{I}} \mathbf{w}_1^* \quad (8.18)$$

que nos permite escrever:

$$\begin{aligned} \mathbf{X}\mathbf{w} &= \mathbf{X}_1 \mathbf{w}_1 + \mathbf{X}_2 \tilde{\mathbf{I}} \mathbf{w}_1^* \\ &= \mathbf{X}_1 \{ \text{Re}(\mathbf{w}_1) + j \text{Im}(\mathbf{w}_1) \} + \mathbf{X}_2 \tilde{\mathbf{I}} \{ \text{Re}(\mathbf{w}_1) - j \text{Im}(\mathbf{w}_1) \} \\ &= \{ \mathbf{X}_1 + \mathbf{X}_2 \tilde{\mathbf{I}} \} \text{Re}(\mathbf{w}_1) + j \{ \mathbf{X}_1 - \mathbf{X}_2 \tilde{\mathbf{I}} \} \text{Im}(\mathbf{w}_1) \\ &= \left[ \begin{array}{c} (\mathbf{X}_1 + \mathbf{X}_2 \tilde{\mathbf{I}}) \\ j (\mathbf{X}_1 - \mathbf{X}_2 \tilde{\mathbf{I}}) \end{array} \right] \left[ \begin{array}{c} \text{Re}(\mathbf{w}_1) \\ \text{Im}(\mathbf{w}_1) \end{array} \right] \\ &= \check{\mathbf{X}} \mathbf{c} \end{aligned} \quad (8.19)$$

Dessa forma, a minimização:

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{X}^H (\mathbf{W}^H \mathbf{W})^{-1} \mathbf{X} \mathbf{w} \quad (8.20)$$

tal que  $\mathbf{w}^H \mathbf{w} = 1$  e  $\mathbf{w} = \tilde{\mathbf{I}} \mathbf{w}^*$ , substituindo  $\mathbf{X}\mathbf{w} = \check{\mathbf{X}}\mathbf{c}$ , equivale à minimização:

$$\min_{\mathbf{c}} \mathbf{c}^T \check{\mathbf{X}}^H (\mathbf{W}^H \mathbf{W})^{-1} \check{\mathbf{X}} \mathbf{c} \quad (8.21)$$

onde  $c_0 = 1$  para se evitar a solução trivial  $\mathbf{c} = \mathbf{0}$ . No sentido de garantirmos que  $\mathbf{c}$  seja real, devemos alterar a minimização acima, fazendo:

$$\min_{\mathbf{c}} \mathbf{c}^T \text{Re} \left\{ \check{\mathbf{X}}^H (\mathbf{W}^H \mathbf{W})^{-1} \check{\mathbf{X}} \right\} \mathbf{c} \quad (8.22)$$

Este procedimento evita que resíduos da parte imaginária do produto de matrizes danifique o cálculo de  $\mathbf{c}$ . Então, definindo:

$$\check{\mathbf{C}}_X^{(k)} \triangleq \text{Re} \left\{ \check{\mathbf{X}}^H (\mathbf{W}_{(k)}^H \mathbf{W}_{(k)})^{-1} \check{\mathbf{X}} \right\} \quad (8.23)$$

podemos, para o caso de exponenciais complexas não-amortecidas, substituir o problema quadrático do IQML pelo seguinte problema quadrático de solução iterativa:

$$\min_{\mathbf{c}_{(k+1)}} \mathbf{c}_{(k+1)}^T \check{\mathbf{C}}_X^{(k)} \mathbf{c}_{(k+1)} \quad (8.24)$$

onde  $c_0 = 1$ . Originalmente, Bresler e Makovski afirmam em [6] que a solução dessa minimização é a mesma do sistema de equações:

$$\check{\mathbf{C}}_X^{(k)} \mathbf{c}_{(k+1)} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (8.25)$$

resolvendo este sistema a cada iteração até a convergência. Porém, não foi possível obtermos convergência para valores satisfatórios de  $\mathbf{c}$  e  $\mathbf{w}$  quando a SNR assumia valores abaixo de 30 dB.

Alternativamente, propomos transformar (8.24) em um problema TLS. Isto pode ser feito pois, diferenciando (8.24) em relação a  $\mathbf{c}_{(k+1)}$  e igualando o resultado a zero, encontramos:

$$\check{\mathbf{C}}_X^{(k)} \mathbf{c}_{(k+1)} = \mathbf{0} \quad (8.26)$$

A solução desta equação pode ser obtida, no sentido do critério TLS, a partir da SVD de  $\check{\mathbf{C}}_X^{(k)}$ :

$$\check{\mathbf{C}}_X^{(k)} = \mathbf{U}_{C_x}^{(k)} \boldsymbol{\Sigma}_{C_x}^{(k)} \mathbf{V}_{C_x}^{(k)H} \quad (8.27)$$

sendo dada por:

$$\mathbf{c}_{(k+1)} = \frac{\mathbf{v}_{C_x(L+1)}^{(k)}}{v_{C_x(L+1,1)}^{(k)}} \quad (8.28)$$

onde  $\mathbf{v}_{C_x(L+1)}^{(k)}$  é a última coluna de  $\mathbf{V}_{C_x}^{(k)}$  e  $v_{C_x(L+1,1)}^{(k)}$  é seu primeiro elemento, conforme descrito no Capítulo 7. Esta proposta mostrou-se mais estável que a solução empregada originalmente no IQML, possibilitando convergência para valores satisfatórios de  $\mathbf{c}$  e  $\mathbf{w}$  em valores de SNR de até 3 dB, como mostraremos adiante em nossas simulações.

Encontrado  $\mathbf{c}$ , o vetor  $\mathbf{w}_1$  é reconstruído e, a partir dele, o vetor de coeficientes  $\mathbf{w}$ . Portanto, a restrição de simetria é implementada formando-se  $\check{\mathbf{C}}_X^{(k)}$  conforme (8.23) e se resolvendo a equação (8.26) até a convergência.

Como é de nosso interesse, procuramos desenvolver uma versão da abordagem anterior para o caso de  $L$  ser **par**. Neste caso, seja  $L = 2q$ . Assim,  $\mathbf{w}$  que tem dimensões  $(L + 1) \times 1$ , possui um número ímpar de elementos. Dessa forma, para que a simetria  $\mathbf{w} = \tilde{\mathbf{I}} \mathbf{w}^*$  se verifique, é necessário que o elemento central de  $\mathbf{w}$  seja real, valendo a relação  $w_i = w_{L-i}^*$ , para  $i = 0, \dots, L$ . Daí, particionando  $\mathbf{w}$  e  $\mathbf{X}$  de forma que:

$$\mathbf{w} = \left[ w_0 \quad \dots \quad w_{q-1} \quad w_q \quad w_{q+1} \quad \dots \quad w_{2q} \right]^T = \left[ \mathbf{w}_1^T \mid w_q \mid \mathbf{w}_2^T \right]^T \quad (8.29)$$

e

$$\mathbf{X} \mathbf{w} = \left[ \mathbf{X}_1 \mid \mathbf{x}_q \mid \mathbf{X}_2 \right] \begin{bmatrix} \mathbf{w}_1 \\ w_q \\ \mathbf{w}_2 \end{bmatrix} = \mathbf{X}_1 \mathbf{w}_1 + \mathbf{x}_q w_q + \mathbf{X}_2 \mathbf{w}_2 \quad (8.30)$$

Da mesma forma que antes, utilizando a matriz de reflexão lateral  $\tilde{\mathbf{I}}$ , a relação de simetria sobre  $\mathbf{w}$  pode ser representada por:

$$\mathbf{w}_2 = \tilde{\mathbf{I}} \mathbf{w}_1^* \quad (8.31)$$

o que nos permite agora escrever:

$$\begin{aligned}
\mathbf{X}\mathbf{w} &= \mathbf{X}_1 \mathbf{w}_1 + \mathbf{x}_q w_q + \mathbf{X}_2 \tilde{\mathbf{I}} \mathbf{w}_1^* \\
&= \mathbf{X}_1 \{ \text{Re}(\mathbf{w}_1) + j \text{Im}(\mathbf{w}_1) \} + \mathbf{X}_2 \tilde{\mathbf{I}} \{ \text{Re}(\mathbf{w}_1) - j \text{Im}(\mathbf{w}_1) \} + \mathbf{x}_q w_q \\
&= \{ \mathbf{X}_1 + \mathbf{X}_2 \tilde{\mathbf{I}} \} \text{Re}(\mathbf{w}_1) + j \{ \mathbf{X}_1 - \mathbf{X}_2 \tilde{\mathbf{I}} \} \text{Im}(\mathbf{w}_1) + \mathbf{x}_q w_q \\
&= \left[ \left( \mathbf{X}_1 + \mathbf{X}_2 \tilde{\mathbf{I}} \right) \mid \mathbf{x}_q \mid j \left( \mathbf{X}_1 - \mathbf{X}_2 \tilde{\mathbf{I}} \right) \right] \begin{bmatrix} \text{Re}(\mathbf{w}_1) \\ w_q \\ \text{Im}(\mathbf{w}_1) \end{bmatrix} \\
&= \check{\mathbf{X}} \mathbf{c}
\end{aligned} \tag{8.32}$$

A partir daqui a minimização segue os passos descritos de (8.20) a (8.28), exceto quanto à reconstrução de  $\mathbf{w}$  a partir de  $\mathbf{w}_1$  e de  $\mathbf{c}$ .

### 8.1.1 Convergência

Embora a matriz resultante de  $\mathbf{W}^H \mathbf{W}$  tenha posto completo, ela é mal condicionada devido à forma como é construída. Assim, na inversão da mesma podem surgir imprecisões nos cálculos que, acumuladas a cada iteração, comprometem a estimação de  $\mathbf{c}$  e, conseqüentemente, de  $\mathbf{w}$ . Procurando obter  $(\mathbf{W}^H \mathbf{W})^{-1}$  sem calcular diretamente a inversa, Kumaresan, Scharf e Shaw [49] propuseram aumentar  $\mathbf{W}^H \mathbf{W}$  de forma a gerar uma matriz circulante. A inversa da matriz circulante é então calculada e a identidade de Woodbury é aplicada à mesma para se obter  $(\mathbf{W}^H \mathbf{W})^{-1}$ . Entretanto, sob certas condições, esta matriz circulante pode ser singular. No sentido de corrigir esta deficiência, Clark e Scharf [12] propuseram alterar os autovalores da matriz circulante de forma a garantir seu bom condicionamento. A seguir detalhamos analiticamente os procedimentos anteriormente descritos.

Definamos então a matrix  $\mathbf{T}^H$  como:

$$\mathbf{T}^H = \begin{bmatrix} w_0 & 0 & \cdots & 0 & w_L & \cdots & w_2 & w_1 \\ w_1 & w_0 & 0 & \cdots & 0 & w_L & \cdots & w_2 \\ \vdots & \ddots & \ddots & \ddots & & \ddots & \ddots & \vdots \\ w_{L-1} & \cdots & w_1 & w_0 & 0 & \cdots & 0 & w_L \end{bmatrix} \tag{8.33}$$

e formemos a matrix circulante  $N \times N$ :

$$\Theta^H = \begin{bmatrix} \mathbf{T}^H \\ \mathbf{W}^H \end{bmatrix} \tag{8.34}$$

Tomando a decomposição em autovalores de  $\Theta$  como:

$$\Theta^H = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H \tag{8.35}$$

podemos verificar que  $\mathbf{W}^H \mathbf{W}$  é o bloco inferior direito da matrix circulante dada por:

$$\mathbf{Q} = \Theta^H \Theta = \mathbf{U} |\Lambda|^2 \mathbf{U}^H = \left[ \begin{array}{c|c} \mathbf{T}^H \mathbf{T} & \mathbf{T}^H \mathbf{W} \\ \hline \mathbf{W}^H \mathbf{T} & \mathbf{W}^H \mathbf{W} \end{array} \right] \quad (8.36)$$

Entretanto, enquanto  $\mathbf{W}^H \mathbf{W}$  é necessariamente definida positiva,  $\mathbf{Q}$  será singular se qualquer raiz de  $W(z)$  corresponder à  $N$ -ésima raiz da unidade. Portanto, se desejamos inverter  $\mathbf{Q}$ , precisamos garantir que ela seja tanto definida positiva quanto bem condicionada. A matriz  $\mathbf{Q}$  será bem condicionada se e somente se  $|\Lambda_{i,i}|^2 \geq \epsilon$ , para  $i = 0, \dots, N-1$ , onde  $\epsilon$  é alguma tolerância não-negativa. Denominando por  $\mathbf{p}$  um vetor contendo os índices  $i \in \{0, \dots, (N-1)\}$  de quaisquer autovalores  $\Lambda_{i,i}$  menores do que  $\epsilon$ , e por  $P$  o número de elementos de  $\mathbf{p}$ , definimos  $\hat{\Lambda}$  como a matriz diagonal contendo "1" nas posições dadas pelos elementos de  $\mathbf{p}$  e "0" nas demais posições. Assim, podemos construir uma matriz  $\hat{\mathbf{Q}}$  definida positiva e bem condicionada, dada por:

$$\hat{\mathbf{Q}} = \mathbf{U} (|\Lambda|^2 + \hat{\Lambda}) \mathbf{U}^H = \mathbf{Q} + \sum_{i=0}^P \mathbf{U}_{p_i} \mathbf{U}_{p_i}^H \quad (8.37)$$

onde  $\mathbf{U}_{p_i}$  denota a coluna de  $\mathbf{U}$  indicada pelo  $i$ -ésimo elemento do vetor  $\mathbf{p}$  de índices. No sentido de simplificar esta equação, definimos a matriz:

$$\mathbf{G}^H = \left[ \begin{array}{cccc} \mathbf{U}_{p_0} & \mathbf{U}_{p_1} & \dots & \mathbf{U}_{p_{(P-1)}} \end{array} \right] \quad (8.38)$$

e particionamos:

$$\mathbf{G}^H = \left[ \begin{array}{c} \mathbf{G}_1^H \\ \mathbf{G}_2^H \end{array} \right] \quad (8.39)$$

onde  $\mathbf{G}_1^H$  é  $L \times P$  e  $\mathbf{G}_2^H$  é  $(N-L) \times P$ . Dessa maneira, podemos reescrever (8.37) como:

$$\hat{\mathbf{Q}} = \mathbf{Q} + \mathbf{G}^H \mathbf{G} = \left[ \begin{array}{c|c} \mathbf{T}^H \mathbf{T} + \mathbf{G}_1^H \mathbf{G}_1 & \mathbf{T}^H \mathbf{W} + \mathbf{G}_1^H \mathbf{G}_2 \\ \hline \mathbf{W}^H \mathbf{T} + \mathbf{G}_2^H \mathbf{G}_1 & \mathbf{W}^H \mathbf{W} + \mathbf{G}_2^H \mathbf{G}_2 \end{array} \right] \quad (8.40)$$

A inversa da matriz circulante  $\hat{\mathbf{Q}}$  é facilmente calculada a partir de:

$$\hat{\mathbf{Q}}^{-1} = \mathbf{U} (|\Lambda|^2 + \hat{\Lambda})^{-1} \mathbf{U}^H \quad (8.41)$$

Efetuada a seguinte partição:

$$\hat{\mathbf{Q}}^{-1} = \left[ \begin{array}{c|c} \boldsymbol{\alpha} & \boldsymbol{\beta} \\ \hline \boldsymbol{\beta}^H & \boldsymbol{\gamma} \end{array} \right] \quad (8.42)$$

onde  $\boldsymbol{\alpha}$  é  $L \times L$ ,  $\boldsymbol{\beta}$  é  $L \times (N-L)$  e  $\boldsymbol{\gamma}$  é  $(N-L) \times (N-L)$ , como  $\hat{\mathbf{Q}}^{-1}$  também é uma matriz circulante, temos que sua inversa, a matriz  $\hat{\mathbf{Q}}$ , pode ser escrita como:

$$\hat{\mathbf{Q}} = \left[ \begin{array}{c|c} \ddots & \vdots \\ \hline \dots & (\boldsymbol{\gamma} - \boldsymbol{\beta}^H \boldsymbol{\alpha}^{-1} \boldsymbol{\beta})^{-1} \end{array} \right] \quad (8.43)$$

aqui apresentado somente seu bloco inferior direito, que é o de nosso interesse. Comparando (8.43) e (8.40), podemos afirmar que:

$$\left(\mathbf{W}^H \mathbf{W} + \mathbf{G}_2^H \mathbf{G}_2\right)^{-1} = \gamma - \beta^H \alpha^{-1} \beta \quad (8.44)$$

Caso  $P = 0$ , podemos obter diretamente a inversa desejada fazendo simplesmente:

$$\left(\mathbf{W}^H \mathbf{W}\right)^{-1} = \gamma - \beta^H \alpha^{-1} \beta \quad (8.45)$$

Caso  $P > 0$ , alguns procedimentos adicionais são necessários. Assim, denominamos:

$$\tilde{\mathbf{Q}} = \mathbf{W}^H \mathbf{W} + \mathbf{G}_2^H \mathbf{G}_2 \quad (8.46)$$

de forma que:

$$\tilde{\mathbf{Q}}^{-1} = \gamma - \beta^H \alpha^{-1} \beta \quad (8.47)$$

Como dispomos tanto de  $\tilde{\mathbf{Q}}^{-1}$  quanto de  $\mathbf{G}_2^H$ , podemos utilizar a indentidade de Woodbury para finalmente determinar  $\left(\mathbf{W}^H \mathbf{W}\right)^{-1}$ :

$$\begin{aligned} \left(\mathbf{W}^H \mathbf{W}\right)^{-1} &= \left(\tilde{\mathbf{Q}} - \mathbf{G}_2^H \mathbf{G}_2\right)^{-1} \\ &= \tilde{\mathbf{Q}}^{-1} + \tilde{\mathbf{Q}}^{-1} \mathbf{G}_2^H \left(\mathbf{I} - \mathbf{G}_2 \tilde{\mathbf{Q}}^{-1} \mathbf{G}_2^H\right) \mathbf{G}_2 \tilde{\mathbf{Q}}^{-1} \end{aligned} \quad (8.48)$$

Estes procedimentos não somente garantem a estabilidade do método IQML como também permitem reduzir sua complexidade computacional, pois a maior inversa exigida passa a ser a da matriz Toeplitz  $\alpha$ , que tem dimensões  $L \times L$ .

### 8.1.2 Desempenho

Procurando verificar o desempenho do método IQML quanto à estimação de freqüências, aplicamos o IQML ao exemplo de simulação utilizado anteriormente para avaliar o desempenho dos métodos FBLP Modificado e TLS-FBLP. Na figura 8.1 apresentamos a curva de desempenho do método IQML com  $L = M = 2$ , obtida para 20 experimentos, conforme efetuado no artigo que propõe o método [6]. Podemos observar que seu desempenho (linha contínua) reproduz aproximadamente o desempenho teórico da Máxima Verossimilhança (linha pontilhada), apresentando limiar apenas abaixo de 3 dB de SNR. A convergência ocorreu normalmente em não mais que 20 iterações, sendo em média 5 iterações para SNR de 30 dB, 9 para 10 dB e 19 para 3 dB.

Esclarecemos que, como o exemplo escolhido consiste em uma soma de exponenciais complexas sem amortecimento imersa em ruído, devemos impor a restrição 2, que foi implementada de maneira implícita sobre  $\mathbf{X}$ . Para evitar imprecisões na inversão de  $\left(\mathbf{W}^H \mathbf{W}\right)$ , devido ao seu mal condicionamento, utilizamos a proposta de Clark e Scharf [12]. Adotamos também nossa

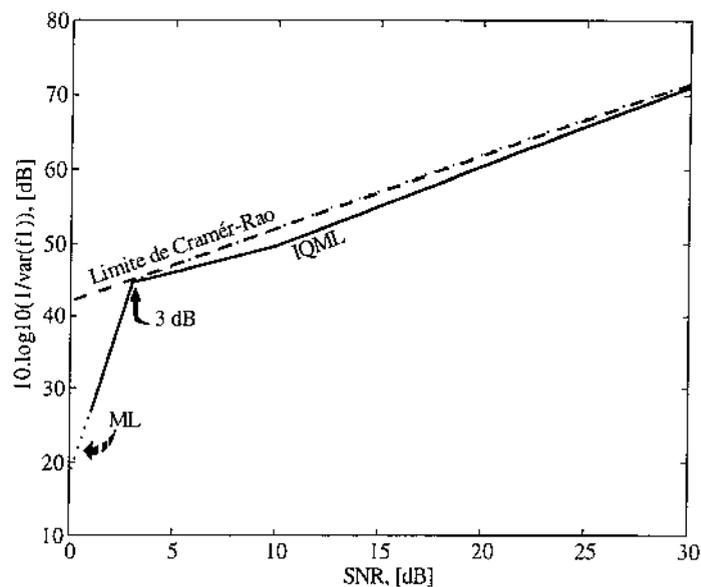


Figura 8.1: Desempenho do método IQML,  $N = 25$ ,  $L = 2$ ,  $M = 2$ , 20 experimentos.

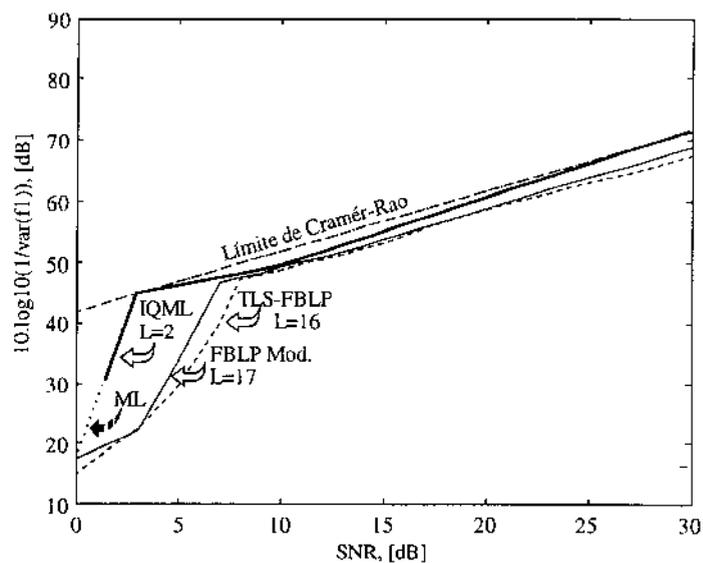


Figura 8.2: Desempenhos dos métodos IQML, FBLP Modificado, TLS-FBLP e ML.

proposta de minimização da função objetivo através de uma solução TLS, ressaltando que foi a única que possibilitou convergência.

No sentido de comparar o desempenho do IQML com o do FBLP Modificado e do TLS-FBLP, na figura 8.2 traçamos suas curvas de desempenho em um mesmo gráfico. Podemos observar que o IQML mantém seu desempenho mais próximo do Limite de Cramér-Rao do que os desempenhos do TLS-FBLP e do FBLP Modificado, revelando-se promissor para a substituição dos métodos mais tradicionais de estimação de Máxima Verossimilhança. Entretanto, infelizmente, a complexidade computacional do IQML é muito superior à do FBLP Modificado, uma vez que emprega pelo menos 5 iterações, cada uma com complexidade equivalente à solução do FBLP Modificado, para cada estimativa de  $w$ .

## 8.2 Mínimos Quadrados Totais Restrito - CTLS

Nesta seção apresentamos uma extensão do método TLS para o caso de existirem dependências lineares entre os componentes de ruído da matriz de dados estendida. Esta extensão foi proposta por Abatzoglou e Mendel [2], originando o Método Mínimos Quadrados Totais Restrito (CTLS). No sentido de esclarecer a motivação de sua formulação recordamos do Capítulo 7 que, no caso ruidoso, a predição linear pode ser descrita por:

$$\left[ \mathbf{b} \mid \mathbf{A} \right] \mathbf{w} \approx \mathbf{0} \quad (8.49)$$

onde

$$\mathbf{w} = \begin{bmatrix} w_0 \\ \mathbf{w}' \end{bmatrix} = \begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix} \quad (8.50)$$

De forma a estabelecer a igualdade em (8.49), a otimização TLS de  $w$  leva em consideração perturbações tanto em  $\mathbf{A}$  quanto em  $\mathbf{b}$ , tal que:

$$\left[ (\mathbf{b} + \Delta\mathbf{b}) \mid (\mathbf{A} + \Delta\mathbf{A}) \right] \mathbf{w} = \mathbf{0} \quad (8.51)$$

ou seja,

$$\left( \left[ \mathbf{b} \mid \mathbf{A} \right] + \left[ \Delta\mathbf{b} \mid \Delta\mathbf{A} \right] \right) \mathbf{w} = \mathbf{0} \quad (8.52)$$

Como definimos anteriormente  $\mathbf{X} = \left[ \mathbf{b} \mid \mathbf{A} \right]$ , definindo agora  $\Delta\mathbf{X} = \left[ \Delta\mathbf{b} \mid \Delta\mathbf{A} \right]$ , podemos reescrever (8.52) como:

$$(\mathbf{X} + \Delta\mathbf{X}) \mathbf{w} = \mathbf{0} \quad (8.53)$$

A solução TLS clássica otimiza  $w$  sujeito a (8.53), pela minimização do produto  $\Delta\mathbf{X}\mathbf{w}$  sem qualquer consideração adicional. Quando a matriz de dados estendida é estruturada, ou seja, quando ela é Toeplitz (FLP), Hankel (BLP), Hankel-Toeplitz (FBLP) ou esparsa, o

procedimento padrão de solução do TLS não fornece um vetor de parâmetros estatisticamente ótimo. Isto ocorre porque a solução do TLS é obtida a partir da SVD de uma matriz de posto incompleto que é uma aproximação, em norma de Frobenius, da matriz de dados estendida original. Esta matriz de posto incompleto é resultante do truncamento da SVD da matriz de dados estendida, não preservando a estrutura da mesma. O método CTLS surge como uma melhoria da abordagem TLS, levando em conta a estrutura da matriz de dados estendida  $\mathbf{X}$  no processo de otimização. Isto é feito observando que, como as perturbações em uma amostra qualquer são submetidas ao filtro de predição juntamente com a própria amostra,  $\Delta\mathbf{X}$  é estruturada da mesma forma que  $\mathbf{X}$ :

$$\Delta\mathbf{X} = \begin{bmatrix} \Delta x[L] & \Delta x[L-1] & \cdots & \Delta x[0] \\ \Delta x[L+1] & \Delta x[L] & \cdots & \Delta x[1] \\ \vdots & \vdots & \ddots & \vdots \\ \Delta x[N-1] & \Delta x[N-2] & \cdots & \Delta x[N-L-1] \end{bmatrix} \quad (8.54)$$

No sentido de considerarmos esta estrutura na otimização, particionamos  $\Delta\mathbf{X}$  em termos de suas colunas:

$$\Delta\mathbf{X} = \left[ \Delta\mathbf{x}_0 \mid \Delta\mathbf{x}_1 \mid \cdots \mid \Delta\mathbf{x}_L \right] \quad (8.55)$$

A partir daí, podemos observar que a informação de estrutura de  $\Delta\mathbf{X}$  pode ser fatorada em um vetor de perturbações  $\Delta\mathbf{x}$  e um conjunto de matrizes  $\{\mathbf{F}_k\}$ ,  $k = 0, \dots, L$ , tal que  $\Delta\mathbf{x}_k = \mathbf{F}_k \Delta\mathbf{x}$ , ou seja:

$$\Delta\mathbf{X} = \left[ \mathbf{F}_0 \Delta\mathbf{x} \mid \mathbf{F}_1 \Delta\mathbf{x} \mid \cdots \mid \mathbf{F}_L \Delta\mathbf{x} \right] \quad (8.56)$$

O vetor  $\Delta\mathbf{x}$  é constituído pelos diferentes valores  $\Delta x[i]$ ,  $i = 0, \dots, K-1$ , onde  $K \leq N$ , que são suficientes para construir  $\Delta\mathbf{X}$ . As matrizes  $\mathbf{F}_k$  são  $(N-L) \times K$  e possuem a mesma estrutura de  $\Delta\mathbf{X}$ . Tomando  $\Delta\mathbf{x}_L$ , podemos verificar que:

$$\begin{aligned} \Delta\mathbf{x}_L &= \begin{bmatrix} \Delta x[0] \\ \Delta x[1] \\ \vdots \\ \Delta x[N-L-1] \end{bmatrix} \\ &= \underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \end{bmatrix}}_{(N-L)} \cdot \begin{bmatrix} \Delta x[0] \\ \Delta x[1] \\ \vdots \\ \Delta x[K-1] \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \end{bmatrix}_{(N-L) \times K} \Delta\mathbf{x} \end{aligned} \quad (8.57)$$

ou seja:

$$\mathbf{F}_L = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 \end{bmatrix} \quad (8.58)$$

Repetindo os últimos procedimentos para  $\Delta \mathbf{x}_{L-1}$  encontramos a matriz:

$$\mathbf{F}_{L-1} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \end{bmatrix} \quad (8.59)$$

Da mesma forma, como os vetores  $\Delta \mathbf{x}_k$  resultam do deslocamento de uma janela de  $(N - L)$  posições sobre as respectivas perturbações nas  $K$  amostras, chegamos até:

$$\mathbf{F}_0 = \begin{bmatrix} 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \quad (8.60)$$

Desse modo, consideramos a estrutura de  $\Delta \mathbf{X}$  em (8.53) fazendo:

$$\left( \mathbf{X} + \left[ \mathbf{F}_0 \Delta \mathbf{x} \mid \mathbf{F}_1 \Delta \mathbf{x} \mid \cdots \mid \mathbf{F}_L \Delta \mathbf{x} \right] \right) \mathbf{w} = 0 \quad (8.61)$$

que pode ser reescrita como:

$$\mathbf{X} \mathbf{w} + \left( \sum_{k=0}^L \mathbf{F}_k w_k \right) \Delta \mathbf{x} = 0 \quad (8.62)$$

Substituindo as matrizes  $\{\mathbf{F}_k\}$  por seus valores de (8.58) a (8.60), observamos que:

$$\sum_{k=0}^L \mathbf{F}_k w_k = \mathbf{W}^H \quad (8.63)$$

Assim, (8.62) fica expressa por:

$$\mathbf{X} \mathbf{w} + \mathbf{W}^H \Delta \mathbf{x} = 0 \quad (8.64)$$

Portanto, minimizar  $\|\Delta \mathbf{X}\|_F$  sujeito a (8.53) é equivalente a minimizar  $\|\Delta \mathbf{x}\|_2$  sujeito a (8.64). O método CTLS consiste, então, em determinar um vetor  $\mathbf{w}$  e a mínima perturbação  $\Delta \mathbf{x}$  tal que (8.64) seja satisfeita. Explicitando  $\Delta \mathbf{x}$  encontramos:

$$\Delta \mathbf{x} = - \left( \mathbf{W}^H \right)^{\#} \mathbf{X} \mathbf{w} \quad (8.65)$$

A norma mínima de  $\Delta \mathbf{x}$  é então dada por:

$$\min \|\Delta \mathbf{x}\|_2^2 = \min (\Delta \mathbf{x}^H \Delta \mathbf{x}) = \min_{\mathbf{w}} \left[ \mathbf{w}^H \mathbf{X}^H \mathbf{W}^\# (\mathbf{W}^H)^\# \mathbf{X} \mathbf{w} \right] \quad (8.66)$$

Como  $\mathbf{W}^H$  tem posto completo, (8.66) pode ser escrita como:

$$\min \|\Delta \mathbf{x}\|_2^2 = \min_{\mathbf{w}} \left[ \mathbf{w}^H \mathbf{X}^H (\mathbf{W}^H \mathbf{W})^{-1} \mathbf{X} \mathbf{w} \right] \quad (8.67)$$

Portanto, a solução CTLS é dada pelo vetor  $\mathbf{w}$  que minimiza (8.67). Devemos observar que (8.67) tem a mesma forma que (8.11), exceto que a formulação TLS utiliza tipicamente  $w_0 = -1$ , enquanto no IQML não se impõe esta condição. Perguntamo-nos então se esta restrição leva os métodos CTLS e IQML a soluções distintas ou se eles são equivalentes. Para responder a esta questão, lançamos mão do lema seguinte, onde demonstramos que a minimização não é afetada por esta restrição [52].

**Lema 8.1:** *Seja  $\mathbf{A} \in \mathbb{C}^{g \times h}$ ,  $\mathbf{z} \in \mathbb{C}^{h \times 1}$ ,  $\mathbf{Q}_z \in \mathbb{C}^{g \times g}$  onde os elementos de  $\mathbf{Q}_z$  são funções quadráticas dos componentes de  $\mathbf{z}$ . Seja  $f$  uma função multivariável de  $\mathbf{z}$  tal que:  $f(\mathbf{z}) = \mathbf{z}^H \mathbf{A}^H \mathbf{Q}_z^{-1} \mathbf{A} \mathbf{z}$ . Então  $f(\alpha \mathbf{z}) = f(\mathbf{z})$ ,  $\forall \alpha \neq 0$ .*

**Prova:** Da definição da função  $f$ , temos  $f(\alpha \mathbf{z}) = (\alpha \mathbf{z})^H \mathbf{A}^H \mathbf{Q}_{\alpha \mathbf{z}}^{-1} \mathbf{A} (\alpha \mathbf{z})$ . Como os elementos de  $\mathbf{Q}_{\alpha \mathbf{z}}$  são funções quadráticas das componentes de  $(\alpha \mathbf{z})$ , temos  $\mathbf{Q}_{\alpha \mathbf{z}} = (\alpha^* \alpha) \mathbf{Q}_z$ . Então, a função  $f(\alpha \mathbf{z}) = \alpha^* \mathbf{z}^H \mathbf{A}^H (\alpha^* \alpha)^{-1} \mathbf{Q}_z^{-1} \mathbf{A} \alpha \mathbf{z} = \mathbf{z}^H \mathbf{A}^H \mathbf{Q}_z^{-1} \mathbf{A} \mathbf{z} = f(\mathbf{z})$ ,  $\forall \alpha \neq 0$ . ■

Então, existe um escalar  $\alpha = -w_0^{-1}$  para a ponderação de  $\mathbf{w}$  de forma a obtermos  $\alpha w_0 = -1$ , sem afetar a minimização conseguida. Isto implica que o IQML e o CTLS atingem o mínimo no mesmo vetor  $\mathbf{w}$  e (8.67) é realmente equivalente a (8.11). Portanto, o CTLS e o IQML minimizam o mesmo erro e fornecem a mesma solução. Assim, o desempenho do CTLS é, conseqüentemente, o apresentado pelo IQML. A equivalência entre o CTLS e o IQML também pode ser observada comparando-se (8.13) com (8.65), de onde podemos concluir que o IQML procura pelo vetor  $\mathbf{w}$  de norma mínima no espaço “nulo” da matriz  $(\mathbf{W}^H)^\# \mathbf{X}$ . Assim, a minimização quadrática no IQML pode ser interpretada como uma otimização TLS. Esta equivalência não é descrita na literatura, classificando-se como mais uma contribuição à compreensão das diversas propostas mais modernas para a estimação de parâmetros.

### 8.3 Mínimos Quadrados Totais Branqueado - WTLS

Uma outra maneira de tratar as dependências lineares entre as componentes de ruído da matriz de dados estendida foi proposta por Hua e Sarkar [37] e denominada Mínimos Quadrados Totais Branqueado (WTLS). Para analisar os efeitos do ruído sobre o critério TLS, lembramos que devido à atuação ruidosa, em geral, não podemos ter uma igualdade em (8.49). A presença de ruído acaba por gerar um erro de predição dado por:

$$\mathbf{X} \mathbf{w} = \mathbf{e} \quad (8.68)$$

Entretanto, utilizando (8.10), (8.68) pode também ser escrita como:

$$\mathbf{W}^H \mathbf{x} = \mathbf{e} \quad (8.69)$$

Substituindo o valor de  $\mathbf{x}$  a partir de (2.15), obtemos:

$$\mathbf{W}^H (\mathbf{E}\mathbf{a}_c + \mathbf{y}) = \mathbf{e} \quad (8.70)$$

Recordando de (8.6) que  $\mathbf{W}^H \mathbf{E} = \mathbf{0}$ , encontramos:

$$\mathbf{W}^H \mathbf{y} = \mathbf{e} \quad (8.71)$$

Como  $\mathbf{y}$  tem média nula e variância  $\sigma_y^2$ , calculamos a variância deste vetor de erro de predição fazendo:

$$\begin{aligned} E \{ \mathbf{e}\mathbf{e}^H \} &= E \{ \mathbf{W}^H \mathbf{y}\mathbf{y}^H \mathbf{W} \} \\ &= \mathbf{W}^H E \{ \mathbf{y}\mathbf{y}^H \} \mathbf{W} \\ &= \sigma_y^2 \mathbf{W}^H \mathbf{W} \end{aligned} \quad (8.72)$$

Assim, devido à estrutura da matriz de dados estendida (expressa pela matriz  $\mathbf{W}$ ), verificamos que o vetor  $\mathbf{e}$  não é branco. Como o vetor de erro de predição ótimo deve ser necessariamente branco, a solução segundo o critério TLS, para o caso de  $\mathbf{X}$  ser estruturada, jamais será ótima a menos que o erro de predição seja branqueado durante o processo de otimização. Dessa forma, para produzirmos o erro de estimação  $\mathbf{e}_w$  ótimo, utilizamos  $(\mathbf{W}^H \mathbf{W})^{-1/2}$  para branquear o vetor  $\mathbf{e}$ :

$$(\mathbf{W}^H \mathbf{W})^{-1/2} \mathbf{W}^H \mathbf{x} = \mathbf{e}_w \quad (8.73)$$

ou

$$(\mathbf{W}^H \mathbf{W})^{-1/2} \mathbf{X}\mathbf{w} = \mathbf{e}_w \quad (8.74)$$

Agora, minimizarmos a norma do erro  $\mathbf{e}_w$  corresponde a fazermos:

$$\begin{aligned} \min (\mathbf{e}_w^H \mathbf{e}_w) &= \min_{\mathbf{w}} \left\{ \mathbf{w}^H \mathbf{X}^H [(\mathbf{W}^H \mathbf{W})^{-1/2}]^H (\mathbf{W}^H \mathbf{W})^{-1/2} \mathbf{X}\mathbf{w} \right\} \\ &= \min_{\mathbf{w}} \left\{ \mathbf{w}^H \mathbf{X}^H [(\mathbf{W}^H \mathbf{W})^H]^{-1/2} (\mathbf{W}^H \mathbf{W})^{-1/2} \mathbf{X}\mathbf{w} \right\} \\ &= \min_{\mathbf{w}} \left\{ \mathbf{w}^H \mathbf{X}^H (\mathbf{W}^H \mathbf{W})^{-1/2} (\mathbf{W}^H \mathbf{W})^{-1/2} \mathbf{X}\mathbf{w} \right\} \\ &= \min_{\mathbf{w}} \left\{ \mathbf{w}^H \mathbf{X}^H (\mathbf{W}^H \mathbf{W})^{-1} \mathbf{X}\mathbf{w} \right\} \end{aligned} \quad (8.75)$$

Assim, a solução WTLS é dada pelo vetor  $\mathbf{w}$  que minimiza (8.75). Novamente, (8.75) é equivalente a (8.67) e a (8.11) e, portanto, o IQML, o CTLS e o WTLS minimizam a norma do mesmo erro e fornecem a mesma solução. A solução WTLS é obtida iterativamente, otimizando o vetor  $\mathbf{w}^{(k+1)}$  do espaço “quasi” nulo da matriz  $\mathbf{R}_w^{(k)} = \mathbf{X}^H (\mathbf{W}_{(k)}^H \mathbf{W}_{(k)})^{-1} \mathbf{X}$ , ou seja:

$$\mathbf{R}_w^{(k)} \mathbf{w}^{(k+1)} = \mathbf{0} \quad (8.76)$$

cuja solução é obtida no sentido TLS a partir da SVD de  $\mathbf{R}_w^{(k)}$ , como descrito no Capítulo 7, até a convergência. Para o caso de exponenciais complexas não-amortecidas, conforme [37], uma versão FBLP para o método WTLS pode ser obtida fazendo-se simplesmente:

$$\mathbf{R}_w^{(k)} = \mathbf{X}^H (\mathbf{W}_{(k)}^H \mathbf{W}_{(k)})^{-1} \mathbf{X} + \tilde{\mathbf{I}} \left[ \mathbf{X}^H (\mathbf{W}_{(k)}^H \mathbf{W}_{(k)})^{-1} \mathbf{X} \right]^* \tilde{\mathbf{I}} \quad (8.77)$$

Esta expressão é obtida a partir das relações de simetria entre  $\mathbf{X}_f$  e  $\mathbf{X}_b$ .

### 8.3.1 Desempenho

No sentido de verificar o desempenho do método WTLS para a estimação de frequências, aplicamos o mesmo ao exemplo de simulação utilizado anteriormente. Na figura 8.3 apresentamos as curvas de desempenho do método WTLS para  $L = M = 2$  e  $L = 12$ , obtidas para 20 experimentos, conforme efetuado no IQML. Podemos observar que seu desempenho para  $L = 12$  (linha contínua) mostrou-se melhor do que com  $L = 2$  (linha tracejada), embora ambos tenham apresentado limiar apenas abaixo de 3 dB de SNR. Assim como no IQML, a convergência ocorreu normalmente em não mais do que 20 iterações. Para evitar imprecisões no cálculo da inversa de  $\mathbf{W}_{(k)}^H \mathbf{W}_{(k)}$ , adaptamos os procedimentos de Clark e Scharf [12] ao WTLS.

Buscando comparar o desempenho do WTLS com o do IQML, na figura 8.4 traçamos suas curvas de desempenho em um mesmo gráfico. Podemos observar que o IQML (linha tracejada) mantém seu desempenho mais próximo do Limite de Cramér-Rao do que o WTLS (linha contínua), embora apresentem limiar apenas aos 3 dB de SNR. Portanto, esses métodos são igualmente robustos à variação da SNR, embora tenham apresentado ligeira diferença na precisão das estimativas. Esta distinção deve-se às diferentes precisões dos algoritmos empregados para a otimização de (8.75), não prejudicando a equivalência entre os métodos, a qual já foi demonstrada analiticamente, pois minimizam a mesma função objetivo, impondo as mesmas restrições.

## 8.4 Conclusões

Neste capítulo discutimos os métodos recentes IQML, CTLS e WTLS que, levando em conta a informação de estrutura da matriz de dados estendida, buscam atingir o mesmo desempenho da ML. Demonstramos analiticamente que estes métodos minimizam a norma do mesmo erro, fornecendo soluções equivalentes. Conseqüentemente, o problema de minimização quadrático no IQML pode também ser interpretado como um problema TLS. Os desempenhos destes métodos superam os do FBLP Modificado e do TLS-FBLP, porém a um custo computacional maior, pois

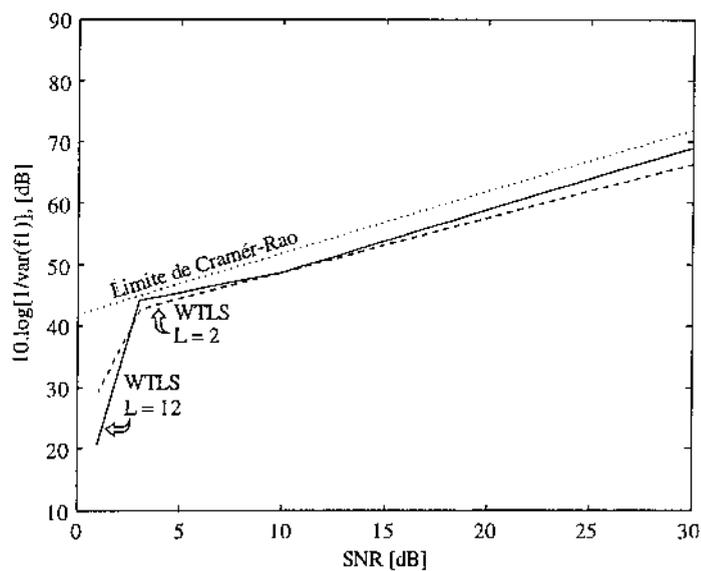


Figura 8.3: Desempenho do método WTLS, na versão FBLP, obtido para 20 experimentos.

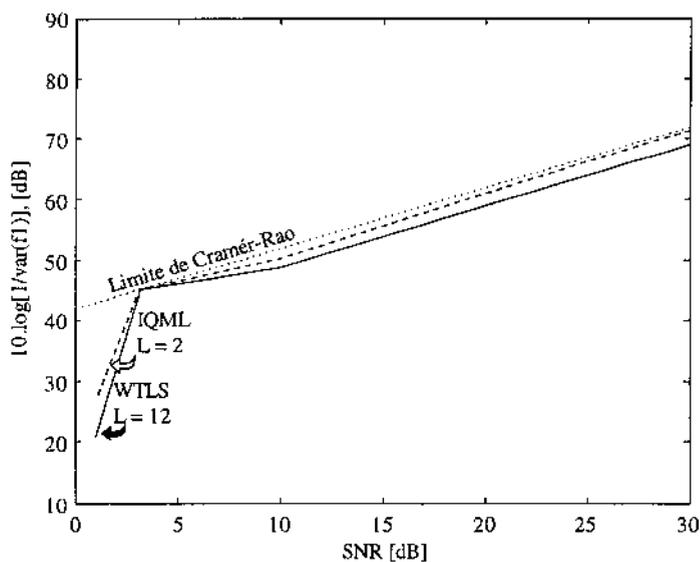


Figura 8.4: Desempenhos dos métodos WTLS e IQML, para 20 experimentos.

utilizam pelo menos 5 iterações, de custo computacional equivalente ao do TLS-FBLP, para cada estimativa do vetor de coeficientes. No capítulo seguinte discutiremos uma outra proposta de correção da deficiência do critério TLS em tratar problemas com estruturação, denominada TLS Estruturado, ou STLS. Esta é a proposta mais recente, sendo por isso mesmo, tratada em destaque.

## Capítulo 9

# Mínimos Quadrados Totais Estruturado - STLS

Como observado no capítulo anterior, o procedimento padrão de solução do TLS não fornece um vetor de parâmetros estatisticamente ótimo quando a matriz de dados estendida é estruturada. Isto ocorre porque a solução do TLS é obtida a partir do truncamento da SVD da matriz de dados estendida, o qual não preserva a estrutura da mesma. Buscando corrigir este problema, De Moor [18], propôs a utilização de uma SVD diferente, denominada SVD Riemanniana, que permite levar em conta a informação de estrutura.

Recordando, pela análise de perturbações, o critério TLS pode ser descrito a partir de:

$$\left[ (\mathbf{b} + \Delta\mathbf{b}) \mid (\mathbf{A} + \Delta\mathbf{A}) \right] \mathbf{w} = \mathbf{0} \quad (9.1)$$

ou seja,

$$\left( \left[ \mathbf{b} \mid \mathbf{A} \right] + \left[ \Delta\mathbf{b} \mid \Delta\mathbf{A} \right] \right) \mathbf{w} = \mathbf{0} \quad (9.2)$$

onde  $\Delta\mathbf{b}$  e  $\Delta\mathbf{A}$  são vetor e matriz de perturbações. No Capítulo 8, verificamos que o método CTLS, no sentido de considerar as informações de estrutura na solução TLS, impõe sobre  $\Delta\mathbf{X} = \left[ \Delta\mathbf{b} \mid \Delta\mathbf{A} \right]$  a mesma estrutura da matriz de dados  $\mathbf{X} = \left[ \mathbf{b} \mid \mathbf{A} \right]$ . Em contrapartida, o método Mínimos Quadrados Totais Estruturado (STLS), proposto por De Moor [18], impõe essa estrutura sobre a matriz  $\left[ (\mathbf{b} + \Delta\mathbf{b}) \mid (\mathbf{A} + \Delta\mathbf{A}) \right]$ . Para facilitar a compreensão desta formulação, seguindo os passos empregados por De Moor, trataremos inicialmente o problema TLS não-estruturado, através da mesma formulação matemática a ser empregada no método STLS.

### 9.1 Revisando o TLS

Nesta seção retomamos o critério TLS, tratando-o através da formulação a ser empregada no método STLS, de forma a chegar à sua já conhecida solução. No critério TLS, como visto anteriormente, obtemos a solução através do truncamento da SVD da matriz  $\mathbf{X}$  de dados estendida.

Isto equivale a utilizarmos, em verdade, a SVD de uma matriz  $\tilde{\mathbf{X}}$  de posto incompleto e que é uma aproximação de  $\mathbf{X}$ . Assim, podemos encarar o TLS como a aproximação, em norma de Frobenius, de uma dada matriz  $\mathbf{X} = \begin{bmatrix} \mathbf{b} & | & \mathbf{A} \end{bmatrix}$  por uma matriz  $\tilde{\mathbf{X}} = \begin{bmatrix} (\mathbf{b} + \Delta\mathbf{b}) & | & (\mathbf{A} + \Delta\mathbf{A}) \end{bmatrix}$  de posto incompleto. Isto pode ser formulado como:

$$\min_{\tilde{\mathbf{X}} \in \mathbb{C}^{(N-L) \times L}, \mathbf{w} \in \mathbb{C}^{L \times 1}} \|\mathbf{X} - \tilde{\mathbf{X}}\|_F^2 \quad \text{sujeito a} \quad \tilde{\mathbf{X}}\mathbf{w} = 0 \quad \text{e} \quad \mathbf{w}^H\mathbf{w} = 1 \quad (9.3)$$

Para efetuar esta minimização podemos utilizar a teoria de Multiplicadores de Lagrange. Neste sentido, definimos  $\lambda \in \mathbb{C}$  como um multiplicador de Lagrange e  $\mathbf{l} \in \mathbb{C}^{(N-L) \times 1}$  como um vetor de multiplicadores de Lagrange, de forma que o Lagrangiano de (9.3) pode ser escrito como:

$$\mathcal{L}(\tilde{\mathbf{X}}, \mathbf{w}, \mathbf{l}, \lambda) = \sum_{i=1}^{(N-L)} \sum_{k=1}^{(L+1)} (x_{ik} - \tilde{x}_{ik})^* (x_{ik} - \tilde{x}_{ik}) + \mathbf{l}^H \tilde{\mathbf{X}}\mathbf{w} + \lambda(1 - \mathbf{w}^H\mathbf{w}) \quad (9.4)$$

Vamos então mostrar como podemos chegar à bem conhecida solução em termos da SVD da matriz  $\mathbf{X}$ .

### 9.1.1 Derivadas

Para determinar os valores dos parâmetros que conduzem  $\mathcal{L}(\tilde{\mathbf{X}}, \mathbf{w}, \mathbf{l}, \lambda)$  ao seu mínimo, fazemos as derivadas de (9.4) em relação a  $\tilde{x}_{ik}$ ,  $\mathbf{w}$ ,  $\mathbf{l}$  e  $\lambda$  iguais a zero. Conforme as propriedades descritas no Apêndice A, obtemos o seguinte conjunto de equações:

$$(\mathbf{X} - \tilde{\mathbf{X}}) = \mathbf{l}\mathbf{w}^H \quad \tilde{\mathbf{X}}^H\mathbf{l} = \mathbf{w}\lambda^* \quad \tilde{\mathbf{X}}\mathbf{w} = 0 \quad \mathbf{w}^H\mathbf{w} = 1 \quad (9.5)$$

Observemos que a perturbação  $\Delta\mathbf{X} = -(\mathbf{X} - \tilde{\mathbf{X}})$  é uma matriz de posto unitário, pois é gerada pelo produto externo dos vetores  $\mathbf{l}$  e  $\mathbf{w}$ . Além disso, temos que:

$$\mathbf{l}^H \tilde{\mathbf{X}}\mathbf{w} = 0 \quad (9.6)$$

Logo,

$$(\tilde{\mathbf{X}}^H\mathbf{l})^H \mathbf{w} = 0 \quad (9.7)$$

ou seja,

$$(\mathbf{w}\lambda^*)^H \mathbf{w} = \lambda\mathbf{w}^H\mathbf{w} = 0 \quad (9.8)$$

e como  $\mathbf{w}^H\mathbf{w} = 1$ , temos:

$$\lambda = 0 \quad (9.9)$$

### 9.1.2 Eliminação de $\tilde{\mathbf{X}}$

Partindo de (9.5) e pós-multiplicando  $(\mathbf{X} - \tilde{\mathbf{X}})$  por  $\mathbf{w}$  e  $(\mathbf{X} - \tilde{\mathbf{X}})^H$  por  $\mathbf{l}$ , utilizando as relações anteriores, encontramos:

$$\mathbf{X}\mathbf{w} = \mathbf{l} \quad \text{e} \quad \mathbf{X}^H\mathbf{l} = \mathbf{w}(\mathbf{l}^H\mathbf{l}) \quad (9.10)$$

### 9.1.3 Normalização

A seguir normalizamos  $\mathbf{l}$  como  $\mathbf{u} = \mathbf{l}/\|\mathbf{l}\|$ , denominamos  $\sigma = \|\mathbf{l}\|$  e rebatizamos  $\mathbf{w}$  como  $\mathbf{v}$ , encontrando:

$$\begin{aligned} \mathbf{X}\mathbf{v} &= \mathbf{u}\sigma & \text{tal que} & \quad \mathbf{u}^H\mathbf{u} = 1 \\ \mathbf{X}^H\mathbf{u} &= \mathbf{v}\sigma & \text{tal que} & \quad \mathbf{v}^H\mathbf{v} = 1 \end{aligned} \quad (9.11)$$

Daí temos que:

$$\begin{aligned} \mathbf{u}^H\mathbf{X}\mathbf{v} &= \mathbf{u}^H\mathbf{u}\sigma \\ &= \sigma \end{aligned}$$

o que implica em que  $\sigma$  seja valor singular de  $\mathbf{X}$  e  $\mathbf{u}$  e  $\mathbf{v}$  seus respectivos vetores singulares. Observe agora que de (9.5) temos:

$$\begin{aligned} \|\mathbf{X} - \tilde{\mathbf{X}}\|_F^2 &= \text{Tr} \left[ (\mathbf{X} - \tilde{\mathbf{X}})^H (\mathbf{X} - \tilde{\mathbf{X}}) \right] \\ &= \text{Tr} \left[ (\mathbf{l}\mathbf{w}^H)^H (\mathbf{l}\mathbf{w}^H) \right] \\ &= \text{Tr} \left[ \mathbf{w}\mathbf{l}^H\mathbf{l}\mathbf{w}^H \right] \end{aligned} \quad (9.12)$$

Mas  $\mathbf{l}^H\mathbf{l} = \sigma^2$ , logo:

$$\begin{aligned} \|\mathbf{X} - \tilde{\mathbf{X}}\|_F^2 &= \sigma^2 \text{Tr} \left[ \mathbf{w}\mathbf{w}^H \right] \\ &= \sigma^2 \mathbf{w}^H\mathbf{w} \\ &= \sigma^2 \end{aligned} \quad (9.13)$$

Assim, a minimização desejada consiste em buscarmos o trio singular correspondente ao menor valor singular. Além disso, a solução do TLS é dada a partir do vetor  $\mathbf{v}$  (ou  $\mathbf{w}$ ) correspondente ao menor valor singular de  $\mathbf{X}$ .

## 9.2 Mínimos Quadrados Totais Estruturado (STLS)

O método de Mínimos Quadrados Totais Estruturado (STLS) foi proposto por De Moor [18] em 1993, como uma extensão do TLS convencional para os casos em que a matriz de dados estendida apresenta uma certa estrutura, por exemplo, Toeplitz, Hankel, Hankel-Toeplitz, matrizes

esparsas, etc. Embora já existissem outros métodos com a mesma finalidade, como o CTLS de Abatzoglou e Mendel [2], a proposta de De Moor apresenta uma formulação mais abrangente e original, incorporando a informação de estrutura à otimização de maneira conceitualmente diferente. No CTLS impõe-se sobre uma matriz de perturbações a mesma estrutura da matriz de dados estendida, enquanto que no STLS esta estrutura é imposta sobre uma matriz que aproxima, em norma de Frobenius, a matriz de dados estendida.

Seguindo a formulação original de De Moor, a partir daqui vamos desenvolver o método STLS. Consideremos então que a matriz  $\mathbf{X} \in \mathbb{C}^{(N-L) \times L}$  possui uma certa estrutura como, por exemplo, Toeplitz, que é típica da abordagem FLP:

$$\mathbf{X} = \begin{bmatrix} x[L] & x[L-1] & \cdots & x[0] \\ x[L+1] & x[L] & \cdots & x[1] \\ \vdots & \vdots & \ddots & \vdots \\ x[N-1] & x[N-2] & \cdots & x[N-L-1] \end{bmatrix} \quad (9.14)$$

Isto significa genericamente que embora  $\mathbf{X}$  seja formada por  $(N-L)(L+1)$  elementos, apenas  $K$  elementos são distintos dentre eles. No caso de  $\mathbf{X}$  ser Toeplitz, temos  $K = N$ . Assim, utilizamos apenas  $K$  elementos do vetor de dados  $\mathbf{x}$  para formar a matriz  $\mathbf{X}$  de dados estendida. Escolhendo um conjunto apropriado de matrizes  $\mathbf{X}_i \in \mathbb{C}^{(N-L) \times L}$ ,  $i = 0, \dots, K$ , podemos representar a informação de estrutura de  $\mathbf{X}$ , decompondo-a na forma de uma função afim dos elementos de  $\mathbf{x}$ :

$$\mathbf{X} = \sum_{i=0}^{K-1} x_i \mathbf{X}_i + \mathbf{X}_K \quad (9.15)$$

onde  $x_i$  é o  $i$ -ésimo elemento de  $\mathbf{x}$  e  $\mathbf{X}_K$  atua como uma polarização. Para  $\mathbf{X}$  Toeplitz, um conjunto possível e trivial de matrizes  $\mathbf{X}_i$  seria dado por  $\mathbf{X}_K = \mathbf{0}$  e  $\mathbf{X}_i$ ,  $i = 0, \dots, K-1$ , matrizes cujos elementos são iguais a "1" nas posições que na matriz  $\mathbf{X}$  correspondem aos elementos  $x[i]$ , e "0" nas demais posições. Por exemplo:

$$\mathbf{X}_{L+1} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix} \quad (9.16)$$

Impomos a estrutura de  $\mathbf{X}$  sobre  $\tilde{\mathbf{X}}$  fazendo:

$$\tilde{\mathbf{X}} = \sum_{i=0}^{K-1} \tilde{x}_i \mathbf{X}_i + \mathbf{X}_K \quad (9.17)$$

onde  $\tilde{x}_i$  é o  $i$ -ésimo elemento de  $\tilde{\mathbf{x}} = \mathbf{x} + \Delta\mathbf{x}$ . Dessa maneira, partindo de (9.3), o problema STLS pode ser descrito por:

$$\min_{\tilde{\mathbf{x}}, \mathbf{w}} \sum_{i=0}^{K-1} (x_i - \tilde{x}_i)^* (x_i - \tilde{x}_i) \text{ tal que } \tilde{\mathbf{X}}\mathbf{w} = \mathbf{0}, \mathbf{w}^H \mathbf{w} = 1 \text{ e } \tilde{\mathbf{X}} \text{ satisfaz (9.17)} \quad (9.18)$$

Em nosso caso,  $K = N$ , ou seja, todas as amostras disponíveis são consideradas na construção de  $\mathbf{X}$  e  $\tilde{\mathbf{X}}$ .

No sentido de resolver a minimização (9.18), tomamos seu Lagrangiano:

$$\mathcal{L}(\tilde{\mathbf{x}}, \mathbf{w}, \mathbf{l}, \lambda) = \sum_{i=0}^{N-1} (x_i - \tilde{x}_i)^* (x_i - \tilde{x}_i) + \mathbf{l}^H \tilde{\mathbf{X}}\mathbf{w} + \lambda(1 - \mathbf{w}^H \mathbf{w}) \quad (9.19)$$

### 9.2.1 Derivadas

Utilizando as propriedades descritas no Apêndice A, fazemos todas as derivadas do Lagrangiano iguais a zero, encontrando:

$$\frac{\partial \mathcal{L}(\tilde{\mathbf{x}}, \mathbf{w}, \mathbf{l}, \lambda)}{\partial \mathbf{l}} = 0 \Rightarrow \tilde{\mathbf{X}}\mathbf{w} = \mathbf{0} \quad (9.20)$$

$$\frac{\partial \mathcal{L}(\tilde{\mathbf{x}}, \mathbf{w}, \mathbf{l}, \lambda)}{\partial \mathbf{w}} = 0 \Rightarrow \tilde{\mathbf{X}}^H \mathbf{l} = \mathbf{w} \lambda^* \quad (9.21)$$

$$\frac{\partial \mathcal{L}(\tilde{\mathbf{x}}, \mathbf{w}, \mathbf{l}, \lambda)}{\partial \lambda} = 0 \Rightarrow \mathbf{w}^H \mathbf{w} = 1 \quad (9.22)$$

$$\frac{\partial \mathcal{L}(\tilde{\mathbf{x}}, \mathbf{w}, \mathbf{l}, \lambda)}{\partial \tilde{\mathbf{x}}} = 0 \Rightarrow (x_i - \tilde{x}_i) = (\mathbf{l}^H \mathbf{X}_i \mathbf{w})^*, \quad i = 0, \dots, N-1 \quad (9.23)$$

De (9.20) temos que  $\mathbf{l}^H \tilde{\mathbf{X}}\mathbf{w} = 0$ . Utilizando este resultado juntamente com (9.21) e (9.22), obtemos  $\lambda = 0$ , analogamente aos cálculos da subseção 9.1.1.

### 9.2.2 Eliminação de $\tilde{\mathbf{X}}$

A partir de (9.23) podemos escrever:

$$\begin{aligned} (\mathbf{X} - \tilde{\mathbf{X}}) \mathbf{w} &= \left( \sum_{i=0}^{N-1} x_i \mathbf{X}_i + \mathbf{X}_K - \sum_{i=0}^{N-1} \tilde{x}_i \mathbf{X}_i - \mathbf{X}_K \right) \mathbf{w} \\ &= \sum_{i=0}^{N-1} (x_i - \tilde{x}_i) \mathbf{X}_i \mathbf{w} \\ &= \sum_{i=0}^{N-1} (\mathbf{l}^H \mathbf{X}_i \mathbf{w})^* \mathbf{X}_i \mathbf{w} \end{aligned} \quad (9.24)$$

e utilizando (9.20), temos que  $(\mathbf{X} - \tilde{\mathbf{X}}) \mathbf{w} = \mathbf{X}\mathbf{w} - \tilde{\mathbf{X}}\mathbf{w} = \mathbf{X}\mathbf{w}$ . Logo, utilizando este resultado em (9.24), obtemos:

$$\begin{aligned}
\mathbf{X}\mathbf{w} &= \sum_{i=0}^{N-1} \mathbf{X}_i\mathbf{w} \left( \mathbf{1}^H \mathbf{X}_i\mathbf{w} \right)^* \\
&= \sum_{i=0}^{N-1} \mathbf{X}_i\mathbf{w} \left[ \left( \mathbf{X}_i\mathbf{w} \right)^T \mathbf{1}^* \right]^* \\
&= \sum_{i=0}^{N-1} \mathbf{X}_i\mathbf{w} \left( \mathbf{X}_i\mathbf{w} \right)^H \mathbf{1} \\
&= \mathbf{D}_w \mathbf{1}
\end{aligned} \tag{9.25}$$

Similarmente, utilizando (9.21) em  $(\mathbf{X} - \bar{\mathbf{X}})^H \mathbf{1}$ , podemos encontrar:

$$\begin{aligned}
\mathbf{X}^H \mathbf{1} &= \sum_{i=0}^{N-1} \left( \mathbf{X}_i^H \mathbf{1} \right) \left( \mathbf{X}_i^H \mathbf{1} \right)^H \mathbf{w} \\
&= \mathbf{D}_l \mathbf{w}
\end{aligned} \tag{9.26}$$

### 9.2.3 Normalização

Definindo  $\mathbf{z} = \mathbf{1} / \|\mathbf{1}\|$  e denominando  $\tau = \|\mathbf{1}\|$ , uma matriz  $\mathbf{D}_z$  pode ser definida da mesma forma que  $\mathbf{D}_l$ , pela simples substituição de cada componente de  $\mathbf{1}$  em  $\mathbf{D}_l$  por seu correspondente em  $\mathbf{z}$ . Como os elementos de  $\mathbf{D}_l$  são funções quadráticas dos componentes de  $\mathbf{1}$ , temos que  $\mathbf{D}_l = \mathbf{D}_z \tau^2$ . Assim, podemos escrever as seguintes relações:

$$\begin{aligned}
\mathbf{X}\mathbf{w} &= \mathbf{D}_w \mathbf{z} \tau \quad \text{tal que} \quad \mathbf{z}^H \mathbf{z} = 1 \\
\mathbf{X}^H \mathbf{z} &= \mathbf{D}_z \mathbf{w} \tau \quad \text{tal que} \quad \mathbf{w}^H \mathbf{w} = 1
\end{aligned} \tag{9.27}$$

Neste caso,  $\{\mathbf{z}, \tau, \mathbf{w}\}$  não é um trio singular de  $\mathbf{X}$ . Isto pode ser verificado fazendo-se:

$$\mathbf{z}^H \mathbf{X}\mathbf{w} = \mathbf{z}^H \mathbf{D}_w \mathbf{z} \tau \tag{9.28}$$

pois, como  $\mathbf{z}^H \mathbf{z} = 1$ , o produto  $\mathbf{z}^H \mathbf{D}_w \mathbf{z}$  não se iguala necessariamente a 1 e não pode ser eliminado da expressão anterior. Na seção seguinte, manipularemos (9.27) da forma a reescrevê-la como uma SVD.

### 9.2.4 STLS na forma de uma SVD Riemanniana

Nesta seção mostraremos que o problema STLS pode ser representado, não por uma SVD convencional, mas sim numa forma denominada SVD Riemanniana [20]. Inicialmente, vamos escrever as variáveis de (9.27) de uma maneira mais apropriada. Neste sentido, definimos  $\mathbf{w} = \mathbf{v} / \|\mathbf{v}\|$ ,  $\mathbf{z} = \mathbf{u} / \|\mathbf{u}\|$  e  $\tau = \sigma \|\mathbf{u}\| \|\mathbf{v}\|$ , tal que  $\mathbf{D}_w = \mathbf{D}_v / \|\mathbf{v}\|^2$  e  $\mathbf{D}_z = \mathbf{D}_u / \|\mathbf{u}\|^2$ . Procedendo à mudança de variáveis em  $\mathbf{X}\mathbf{w} = \mathbf{D}_w \mathbf{z} \tau$ , obtemos:

$$\mathbf{X} \frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{\mathbf{D}_v}{\|\mathbf{v}\|^2} \frac{\mathbf{u}}{\|\mathbf{u}\|} \sigma \|\mathbf{u}\| \|\mathbf{v}\|$$

ou seja:

$$\mathbf{X}\mathbf{v} = \mathbf{D}_v\mathbf{u}\sigma \quad (9.29)$$

Da mesma maneira, substituindo em  $\mathbf{X}^H\mathbf{z} = \mathbf{D}_z\mathbf{w}\tau$ , encontramos:

$$\mathbf{X}^H \frac{\mathbf{u}}{\|\mathbf{u}\|} = \frac{\mathbf{D}_u}{\|\mathbf{u}\|^2} \frac{\mathbf{v}}{\|\mathbf{v}\|} \sigma \|\mathbf{u}\| \|\mathbf{v}\|$$

logo:

$$\mathbf{X}^H\mathbf{u} = \mathbf{D}_u\mathbf{v}\sigma \quad (9.30)$$

Para que  $\{\mathbf{u}, \sigma, \mathbf{v}\}$  correspondesse a um trio singular de  $\mathbf{X}$ , deveríamos ter  $\mathbf{u}^H\mathbf{X}\mathbf{v} = \sigma$ . Porém, de (9.29) temos:

$$\mathbf{u}^H\mathbf{X}\mathbf{v} = \mathbf{u}^H\mathbf{D}_v\mathbf{u}\sigma \quad (9.31)$$

Desse modo, para que (9.31) seja uma SVD, faz-se necessário que  $\mathbf{u}^H\mathbf{D}_v\mathbf{u} = 1$ . Da mesma forma, devemos ter  $\mathbf{v}^H\mathbf{D}_u\mathbf{v} = 1$ . Assim, podemos reescrever (9.27) como a SVD Riemanniana de  $\mathbf{X}$ :

$$\begin{aligned} \mathbf{X}\mathbf{v} &= \mathbf{D}_v\mathbf{u}\sigma \quad \text{tal que} \quad \mathbf{u}^H\mathbf{D}_v\mathbf{u} = 1 \\ \mathbf{X}^H\mathbf{u} &= \mathbf{D}_u\mathbf{v}\sigma \quad \text{tal que} \quad \mathbf{v}^H\mathbf{D}_u\mathbf{v} = 1 \end{aligned} \quad (9.32)$$

A matriz  $\mathbf{D}_u$  é dada por  $\mathbf{D}_u = \sum_{i=0}^{N-1} (\mathbf{X}_i^H\mathbf{u})(\mathbf{X}_i^H\mathbf{u})^H$  e, portanto, é definida semipositiva ou não-negativa, e seus elementos são funções quadráticas das componentes de  $\mathbf{u}$ . A matriz  $\mathbf{D}_v$  é dada por  $\mathbf{D}_v = \sum_{i=0}^{N-1} (\mathbf{X}_i\mathbf{v})(\mathbf{X}_i\mathbf{v})^H$ , é definida semipositiva e seus elementos são funções quadráticas das componentes de  $\mathbf{v}$ . A relação (9.32) é conhecida também por SVD Restrita, ou SVD do quociente de três matrizes  $\{\mathbf{X}, \mathbf{D}_v^{1/2}, \mathbf{D}_u^{1/2}\}$ , sendo descrita em [15][17]. Comparando (9.32) a (9.11), verificamos que a SVD convencional exige que os vetores singulares sejam unitários, enquanto a SVD Riemanniana introduz matrizes de ponderação  $\mathbf{D}_u$  e  $\mathbf{D}_v$ , de forma que não mais os vetores singulares são unitários, mas sim os produtos  $\mathbf{u}^H\mathbf{D}_v\mathbf{u}$  e  $\mathbf{v}^H\mathbf{D}_u\mathbf{v}$  é que se igualam a 1.

Vamos agora retornar a (9.18) buscando expressar a função objetivo do problema STLS em termos de  $\{\mathbf{u}, \sigma, \mathbf{v}\}$ . Partindo de (9.23) podemos escrever:

$$\sum_{i=0}^{N-1} (x_i - \bar{x}_i)(x_i - \bar{x}_i)^* = \sum_{i=0}^{N-1} (\mathbf{1}^H\mathbf{X}_i\mathbf{w})(\mathbf{1}^H\mathbf{X}_i\mathbf{w})^*$$

$$\begin{aligned}
&= \sum_{i=0}^{N-1} (\mathbf{z}^H \mathbf{X}_i \mathbf{w}) (\mathbf{z}^H \mathbf{X}_i \mathbf{w})^* \tau^2 \\
&= \mathbf{z}^H \sum_{i=0}^{N-1} (\mathbf{X}_i \mathbf{w}) (\mathbf{z}^H \mathbf{X}_i \mathbf{w})^H \tau^2 \\
&= \mathbf{z}^H \sum_{i=0}^{N-1} (\mathbf{X}_i \mathbf{w}) (\mathbf{X}_i \mathbf{w})^H \mathbf{z} \tau^2
\end{aligned} \tag{9.33}$$

que, utilizando a definição de  $\mathbf{D}_w$  dada em (9.25), pode ser escrita na forma:

$$\sum_{i=0}^{N-1} (x_i - \tilde{x}_i)^* (x_i - \tilde{x}_i) = \mathbf{z}^H \mathbf{D}_w \mathbf{z} \tau^2 \tag{9.34}$$

Substituindo os valores de  $\mathbf{w}$ ,  $\mathbf{z}$  e  $\tau$ , encontramos:

$$\begin{aligned}
\sum_{i=0}^{N-1} (x_i - \tilde{x}_i)^* (x_i - \tilde{x}_i) &= \left( \frac{\mathbf{u}}{\|\mathbf{u}\|} \right)^H \frac{\mathbf{D}_v}{\|\mathbf{v}\|^2} \frac{\mathbf{u}}{\|\mathbf{u}\|} \sigma^2 \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 \\
&= \mathbf{u}^H \mathbf{D}_v \mathbf{u} \sigma^2 \\
&= \sigma^2
\end{aligned} \tag{9.35}$$

Portanto, o problema STLS dado pela minimização (9.18) corresponde a encontrar os vetores singulares  $\mathbf{u}$  e  $\mathbf{v}$ , correspondentes ao mínimo valor singular  $\sigma$  que satisfaz a SVD Riemanniana de  $\mathbf{X}$ , dada por (9.32), onde o vetor  $\mathbf{w}$  ótimo é calculado a partir de  $\mathbf{w} = \mathbf{v} / \|\mathbf{v}\|$ .

O cálculo dos vetores singulares  $\mathbf{u}$  e  $\mathbf{v}$  é efetuado recursivamente a partir das equações em (9.32). Partindo de valores iniciais quaisquer  $\mathbf{u}^{(0)}$ ,  $\mathbf{v}^{(0)}$  e  $\sigma^{(0)}$ , primeiramente calcula-se  $\mathbf{u}^{(k+1)} = (\mathbf{D}_v^{(k)})^{-1} \mathbf{X} \mathbf{v}^{(k)} / \sigma^{(k)}$  e depois  $\mathbf{v}^{(k+1)} = \mathbf{X}^\# \mathbf{D}_u^{(k)} \mathbf{u}^{(k+1)} \sigma^{(k)}$ , impondo também as restrições  $\mathbf{u}^H \mathbf{D}_v \mathbf{u} = \mathbf{v}^H \mathbf{D}_u \mathbf{v} = 1$  sobre  $\mathbf{u}$  e  $\mathbf{v}$  até sua convergência.

Assim, através da inserção das matrizes de ponderação  $\mathbf{D}_u$  e  $\mathbf{D}_v$ , a SVD Riemanniana permite incorporar a informação de estrutura de  $\mathbf{X}$  na otimização do vetor  $\mathbf{w}$  de coeficientes. Isto corrige a deficiência do critério TLS em tratar problemas estruturados e define o método STLS.

### 9.3 Equivalência entre STLS e CTLS

Como observado até aqui, muitos métodos distintos, como o CTLS e o STLS, foram propostos como extensões do TLS para encontrar a solução ótima no caso de a matriz de dados apresentar uma certa estrutura. Verificaremos, a seguir, que quando se impõe sobre  $\tilde{\mathbf{X}}$  a mesma estrutura de  $\mathbf{X}$ , os problemas STLS e CTLS, bem como suas soluções, são equivalentes. Este resultado pode ser encontrado em [52], tendo sido concomitantemente desenvolvido também por nós ao longo de nosso trabalho de pesquisa. Para procedermos à demonstração desta equivalência, supomos  $\mathbf{D}_w$  definida positiva e reescrevemos (9.25) como:

$$\mathbf{l} = \mathbf{D}_w^{-1} \mathbf{X} \mathbf{w} \quad (9.36)$$

Partindo de (9.23) e utilizando (9.36), podemos reescrever (9.18) como:

$$\begin{aligned} \min_{\tilde{\mathbf{x}}, \mathbf{w}} \sum_{i=0}^{N-1} (x_i - \tilde{x}_i)^* (x_i - \tilde{x}_i) &= \min_{\mathbf{w}} \mathbf{l}^H \sum_{i=0}^{N-1} (\mathbf{1}^H \mathbf{X}_i \mathbf{w})^* \mathbf{X}_i \mathbf{w} \\ &= \min_{\mathbf{w}} \mathbf{l}^H \mathbf{D}_w \mathbf{l} \\ &= \min_{\mathbf{w}} \mathbf{w}^H \mathbf{X}^H (\mathbf{D}_w^{-1})^H \mathbf{D}_w (\mathbf{D}_w^{-1} \mathbf{X} \mathbf{w}) \\ &= \min_{\mathbf{w}} \mathbf{w}^H \mathbf{X}^H (\mathbf{D}_w^{-1})^H \mathbf{X} \mathbf{w} \end{aligned} \quad (9.37)$$

Como  $\mathbf{D}_w$  é Hermitiana por construção, temos que  $(\mathbf{D}_w^{-1})^H = \mathbf{D}_w^{-1}$ . Desta maneira, a função objetivo do STLS fica expressa por:

$$\min_{\tilde{\mathbf{x}}} \sum_{i=0}^{N-1} (x_i - \tilde{x}_i)^* (x_i - \tilde{x}_i) = \min_{\mathbf{w}} \mathbf{w}^H \mathbf{X}^H \mathbf{D}_w^{-1} \mathbf{X} \mathbf{w} \quad (9.38)$$

Vamos agora analisar a diferença  $(\mathbf{X} - \tilde{\mathbf{X}}) = -\Delta \mathbf{X}$  à luz da formulação empregada na seção referente ao CTLS. Neste método, a informação de estrutura de  $\Delta \mathbf{X}$  é expressa em termos de um conjunto de matrizes  $\{\mathbf{F}_k\}$ ,  $k = 0, \dots, L$ , e de um vetor  $\Delta \mathbf{x}$  de perturbações, tal que:

$$\begin{aligned} (\mathbf{X} - \tilde{\mathbf{X}}) &= -\Delta \mathbf{X} \\ &= - \left[ \mathbf{F}_0 \Delta \mathbf{x} \mid \mathbf{F}_1 \Delta \mathbf{x} \mid \dots \mid \mathbf{F}_L \Delta \mathbf{x} \right] \end{aligned} \quad (9.39)$$

Mas, se considerarmos que  $\mathbf{X}$  e  $\tilde{\mathbf{X}}$  são estruturados da mesma forma,  $\Delta \mathbf{X}$  também apresentará a mesma estrutura e, segundo a formulação empregada no STLS, podemos escrever:

$$\begin{aligned} (\mathbf{X} - \tilde{\mathbf{X}}) &= \sum_{i=0}^{N-1} (x_i - \tilde{x}_i) \mathbf{X}_i \\ &= - \sum_{i=0}^{N-1} \Delta x_i \mathbf{X}_i \end{aligned} \quad (9.40)$$

Portanto, as representações da estrutura de  $\mathbf{X}$  utilizadas pelos métodos CTLS e STLS podem ser relacionadas através de:

$$\left[ \mathbf{F}_0 \Delta \mathbf{x} \mid \mathbf{F}_1 \Delta \mathbf{x} \mid \dots \mid \mathbf{F}_L \Delta \mathbf{x} \right] = \sum_{i=0}^{N-1} \Delta x_i \mathbf{X}_i \quad (9.41)$$

Para esclarecer a relação entre os conjuntos de matrizes  $\{\mathbf{F}_k\}$ ,  $k = 0, \dots, L$ , e  $\{\mathbf{X}_i\}$ ,  $i = 0, \dots, (N-1)$ , seguimos a notação empregada em Matlab<sup>®</sup>, na qual a  $k$ -ésima coluna de  $\mathbf{X}_i$

é representada por  $\mathbf{X}_i(:, k)$ . Neste equacionamento admitimos que a primeira coluna recebe o índice "0". Podemos então escrever cada coluna de  $\Delta \mathbf{X}$  como:

$$\mathbf{F}_k \Delta \mathbf{x} = \sum_{i=0}^{N-1} \Delta x_i \mathbf{X}_i(:, k) \quad (9.42)$$

que na forma matricial fica expressa por:

$$\mathbf{F}_k \Delta \mathbf{x} = \left[ \mathbf{X}_0(:, k) \mid \mathbf{X}_1(:, k) \mid \cdots \mid \mathbf{X}_{(N-1)}(:, k) \right] \Delta \mathbf{x} \quad (9.43)$$

que é válida para todo  $\Delta \mathbf{x}$  se e somente se:

$$\mathbf{F}_k = \left[ \mathbf{X}_0(:, k) \mid \mathbf{X}_1(:, k) \mid \cdots \mid \mathbf{X}_{(N-1)}(:, k) \right] \quad (9.44)$$

Assim, esclarecemos a relação entre as matrizes de estrutura dos métodos CTLS e STLS, concluindo que a  $i$ -ésima coluna de uma matriz  $\mathbf{F}_k$  corresponde à  $k$ -ésima coluna de uma matriz  $\mathbf{X}_i$ .

Retornando ao STLS, a partir de (9.25), podemos escrever  $\mathbf{D}_w$  na forma matricial como:

$$\begin{aligned} \mathbf{D}_w &= \sum_{i=0}^{N-1} (\mathbf{X}_i \mathbf{w}) (\mathbf{X}_i \mathbf{w})^H \\ &= \left[ \mathbf{X}_0 \mathbf{w} \mid \mathbf{X}_1 \mathbf{w} \mid \cdots \mid \mathbf{X}_{(N-1)} \mathbf{w} \right] \left[ \mathbf{X}_0 \mathbf{w} \mid \mathbf{X}_1 \mathbf{w} \mid \cdots \mid \mathbf{X}_{(N-1)} \mathbf{w} \right]^H \end{aligned} \quad (9.45)$$

Explicitando os produtos em cada coluna de  $\left[ \mathbf{X}_0 \mathbf{w} \mid \mathbf{X}_1 \mathbf{w} \mid \cdots \mid \mathbf{X}_{(N-1)} \mathbf{w} \right]$ , obtemos:

$$\begin{aligned} \left[ \mathbf{X}_0 \mathbf{w} \mid \cdots \mid \mathbf{X}_{(N-1)} \mathbf{w} \right] &= \left[ \sum_{k=0}^L \mathbf{X}_0(:, k) w_k \mid \cdots \mid \sum_{k=0}^L \mathbf{X}_{(N-1)}(:, k) w_k \right] \\ &= \sum_{k=0}^L \left[ \mathbf{X}_0(:, k) \mid \cdots \mid \mathbf{X}_{(N-1)}(:, k) \right] w_k \end{aligned} \quad (9.46)$$

Utilizando (9.44) em (9.46), encontramos:

$$\left[ \mathbf{X}_0 \mathbf{w} \mid \cdots \mid \mathbf{X}_{(N-1)} \mathbf{w} \right] = \sum_{k=0}^L \mathbf{F}_k w_k \quad (9.47)$$

Relembrando que, na formulação do CTLS, denominamos  $\mathbf{W}^H = \sum_{k=0}^L \mathbf{F}_k w_k$  em (8.63), podemos então escrever:

$$\left[ \mathbf{X}_0 \mathbf{w} \mid \mathbf{X}_1 \mathbf{w} \mid \cdots \mid \mathbf{X}_{(N-1)} \mathbf{w} \right] = \mathbf{W}^H \quad (9.48)$$

Substituindo esta expressão em (9.45), segue-se que:

$$\mathbf{D}_w = \mathbf{W}^H \mathbf{W} \quad (9.49)$$

Como vimos no Capítulo 8 que  $\mathbf{W}$  tem posto completo, verificamos através de (9.49) que  $\mathbf{D}_w$  também o tem, corroborando a hipótese inicial desta seção.

Substituindo (9.49) em (9.38), podemos finalmente escrever a função objetivo do STLS como:

$$\min_{\tilde{\mathbf{x}}, \mathbf{w}} \sum_{i=0}^{N-1} (x_i - \tilde{x}_i)^* (x_i - \tilde{x}_i) = \min_{\mathbf{w}} \left\{ \mathbf{w}^H \mathbf{X}^H (\mathbf{W}^H \mathbf{W})^{-1} \mathbf{X} \mathbf{w} \right\} \quad (9.50)$$

Esta expressão é idêntica à função objetivo do método CTLS e, portanto, podemos dizer que o método STLS minimiza o mesmo erro e fornece a mesma solução dada pelo CTLS. Poderíamos questionar a equivalência entre STLS e CTLS, observando que, embora minimizem uma mesma função objetivo, a eles são aplicadas restrições diferentes, ou seja, no STLS exigimos que  $\mathbf{v}^H \mathbf{D}_u \mathbf{v} = \mathbf{u}^H \mathbf{D}_u \mathbf{u} = 1$ , enquanto que no CTLS basta apenas que  $\mathbf{w}^H \mathbf{w} = 1$ . Que conseqüências têm estas restrições sobre a minimização? Nossa procura pela resposta a esta questão foi inicialmente desencorajada pelos comentários de De Moor em [20], que afirmavam que estas diferentes restrições impossibilitavam a equivalência. Porém, contrariando esta afirmação, recentemente, Lemmerling, De Moor e Van Huffel [52] demonstraram numericamente que a minimização de (9.32), seja utilizando as restrições  $\mathbf{v}^H \mathbf{D}_u \mathbf{v} = \mathbf{u}^H \mathbf{D}_u \mathbf{u} = 1$ , ou unicamente  $\mathbf{v}^H \mathbf{v} = 1$ , converge para um mesmo vetor ótimo  $\mathbf{v}$ . Assim, concluíram que  $\mathbf{D}_u$  pode ser qualquer matriz definida positiva, corroborando a equivalência entre STLS e CTLS.

As diferenças entre as formulações STLS e CTLS surgem apenas quando  $\tilde{\mathbf{X}}$  ou  $\Delta \mathbf{X}$  são estruturadas de forma diferente de  $\mathbf{X}$ . Por exemplo, consideremos os problemas STLS e CTLS onde, respectivamente, impomos sobre as matrizes  $\tilde{\mathbf{X}}$  e  $\Delta \mathbf{X}$  uma mesma estrutura, porém diferente daquela de  $\mathbf{X}$ . Neste caso, a matriz resultante da diferença  $(\mathbf{X} - \tilde{\mathbf{X}})$ , no STLS, não possui a estrutura que foi imposta e, conseqüentemente, não possui a mesma estrutura da matriz  $\Delta \mathbf{X}$  do CTLS. Isto significa que, neste caso, não podemos encontrar uma relação entre as matrizes  $\{\mathbf{F}_k\}$  do CTLS e as matrizes  $\{\mathbf{X}_i\}$  do STLS de tal maneira que estes métodos formulem exatamente o mesmo problema, isto é, sejam equivalentes.

## 9.4 Conclusões

Assim, como o STLS é equivalente ao CTLS, levando em consideração os resultados do Capítulo 8, podemos afirmar que o método STLS é também equivalente aos métodos WTLS e IQML, quando  $\tilde{\mathbf{X}}$  ou  $\Delta \mathbf{X}$  são estruturadas da mesma forma que  $\mathbf{X}$ . Este resultado é de grande abrangência, envolvendo todos os métodos mais recentes para a estimação de parâmetros. Na literatura não encontramos afirmação tão abrangente e detalhada, o que nos leva a crer que este resultado constitui-se em uma contribuição significativa à compreensão do poder de solução dos métodos estudados. No capítulo seguinte propomo-nos a desenvolver métodos LS que levem em conta a informação de estrutura, na busca de algoritmos com maior robustez, como a apresentada pelo FBLP Modificado em relação ao TLS-FBLP.

# Capítulo 10

## Estruturando o LS

No Capítulo 7 verificamos que, para um número pequeno de amostras disponíveis, como  $N = 25$ , o método FBLP Modificado apresentou melhor desempenho do que o TLS-FBLP para SNR's abaixo de 10 dB. Acreditamos que isto se deve à maior robustez do FBLP Modificado à perda de estrutura provocada pelo truncamento da SVD de  $\mathbf{A}$ . Nos Capítulos 8 e 9, analisamos os métodos IQML, WTLS, CTLS e STLS, que permitem manter a estrutura da matriz de dados estendida  $\mathbf{X}$ , mesmo reduzindo seu posto segundo o critério TLS. Embora os métodos estudados nos Capítulos 8 e 9 apresentem desempenho equivalente ao da Máxima Verossimilhança, a maior robustez do LS ao truncamento da SVD nos incentiva a buscar a inclusão da informação de estrutura da matriz em sua otimização. Neste sentido, introduzimos neste capítulo a abordagem estruturada dos métodos CTLS e STLS ao caso LS.

Na otimização LS consideram-se perturbações apenas em um dos lados de  $\mathbf{A}\mathbf{w}' \approx \mathbf{b}$ , quer seja em  $\mathbf{A}$  ou em  $\mathbf{b}$ , no sentido de se escrever esta relação como uma igualdade. Enquanto isso, no TLS consideram-se perturbações em ambos os lados. Desta forma, podemos interpretar o LS como um caso particular do TLS onde, por exemplo, as perturbações sobre a primeira coluna de  $\mathbf{X} = \begin{bmatrix} \mathbf{b} & | & \mathbf{A} \end{bmatrix}$  (ou seja,  $\mathbf{b}$ ) são desconsideradas no processo de otimização, de forma que a matriz de perturbações fica expressa apenas por  $\Delta\mathbf{X} = \begin{bmatrix} \mathbf{0} & | & \Delta\mathbf{A} \end{bmatrix}$ . Então, o LS pode ser formulado como um TLS com elementos fixos.

Neste capítulo propomos quatro métodos inéditos, até onde temos notícia. Primeiramente, introduzimos o método de Mínimos Quadrados Restrito (CLS), que é uma versão LS do método CTLS. Embora na literatura encontremos este nome associado a outro método, resolvemos aqui adotá-lo no sentido de manter a nomenclatura e facilitar a associação entre os dois métodos. A seguir, procurando estabelecer a formulação que nos permitirá incorporar na otimização LS a informação de estrutura adotada no STLS, desenvolvemos uma abordagem genérica, introduzindo o método STLS com Elementos Fixos. Este método revelou-se também uma generalização do STLS. Em seguida, como um caso particular dessa formulação, propomos o método Mínimos Quadrados Estruturado (SLS). Adicionalmente, formulamos uma versão LS também para o WTLS, dando origem ao método Mínimos Quadrados Branqueado (WLS).

## 10.1 Mínimos Quadrados Restrito - CLS

Repetindo os procedimentos anteriormente adotados no CTLS, buscamos agora desenvolver e determinar um estimativa de mínimos quadrados que leve em conta a estrutura existente na matriz de dados. Colocado de forma semelhante ao problema TLS, o problema LS considera perturbações em apenas um dos membros da equação de predição, ou seja, apenas no vetor  $\mathbf{b}$  ou na matriz de dados  $\mathbf{A}$  simplesmente. Desse modo, podemos descrever a predição linear no contexto LS por:

$$(\mathbf{A} + \Delta\mathbf{A}) \mathbf{w}' = \mathbf{b} \quad (10.1)$$

Assim, resolver o problema LS consiste simplesmente em otimizar  $\mathbf{w}'$  tal que a norma do produto  $\Delta\mathbf{A}\mathbf{w}'$  seja minimizada. Se adicionarmos a consideração de que, na ausência de ruído, o posto de  $\mathbf{A} = \mathbf{A}_E$  é originariamente incompleto, aproximamos a matriz  $(\mathbf{A} + \Delta\mathbf{A})$ , de posto completo, por uma matriz de posto incompleto, definindo a solução de Tufts e Kumaresan. Entretanto, considerando que  $\mathbf{A}$  possui alguma estrutura especial, como feito pelo CTLS, podemos definir uma nova solução. Neste sentido, inicialmente reorganizamos a equação anterior em uma formulação parecida com a do TLS:

$$\left( \left[ \mathbf{b} \mid \mathbf{A} \right] + \left[ \mathbf{0} \mid \Delta\mathbf{A} \right] \right) \begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix} = \mathbf{0} \quad (10.2)$$

Agora, seguimos a formulação do CTLS, representando a informação de estrutura de  $\Delta\mathbf{A}$  como:

$$\left( \mathbf{X} + \left[ \mathbf{0} \mid \mathbf{F}_1\Delta\mathbf{x} \mid \cdots \mid \mathbf{F}_L\Delta\mathbf{x} \right] \right) \begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix} = \mathbf{0} \quad (10.3)$$

onde  $\mathbf{X} = \left[ \mathbf{b} \mid \mathbf{A} \right]$  e as matrizes  $\mathbf{F}_i$ ,  $i = 1, \dots, L$ , e o vetor  $\Delta\mathbf{x}$  são os mesmos utilizados em (8.56) e (9.41). Cada matriz  $\mathbf{F}_i$  tem a mesma dimensão de  $\mathbf{W}^H$ .

Expandindo e explicitando os produtos matriciais, obtemos:

$$\mathbf{X} \begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix} + \left( \sum_{i=1}^L \mathbf{F}_i w_{(i)} \right) \Delta\mathbf{x} = \mathbf{0} \quad (10.4)$$

Definindo  $\mathbf{W}_{LS}^H = \sum_{i=1}^L \mathbf{F}_i w_{(i)}$ , de mesmas dimensões que  $\mathbf{W}^H$ , ou seja,  $(N - L) \times N$ :

$$\mathbf{W}_{LS}^H = \begin{bmatrix} w_{(L)} & w_{(L-1)} & \cdots & w_{(1)} & 0 & \cdots & 0 & 0 \\ 0 & w_{(L)} & w_{(L-1)} & \cdots & w_{(1)} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & w_{(L)} & w_{(L-1)} & \cdots & w_{(1)} & 0 \end{bmatrix}$$

Assim, podemos escrever:

$$\mathbf{X} \begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix} + \mathbf{W}_{LS}^H \Delta\mathbf{x} = \mathbf{0} \quad (10.5)$$

Então, o valor de  $\Delta \mathbf{x}$  que satisfaz (10.5) é dado, em função da solução  $\mathbf{w}'$ , por:

$$\Delta \mathbf{x} = - \left( \mathbf{W}_{LS}^H \right)^\# \mathbf{X} \begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix} \quad (10.6)$$

Desse modo, minimizar a norma de  $\Delta \mathbf{A}$  em (10.2) corresponde a minimizar a norma de  $\Delta \mathbf{x}$  em (10.5):

$$\min \|\Delta \mathbf{x}\|_2^2 = \min_{\mathbf{w}'} \begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix}^H \mathbf{X}^H \mathbf{W}_{LS}^\# \left( \mathbf{W}_{LS}^H \right)^\# \mathbf{X} \begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix} \quad (10.7)$$

A matriz  $\mathbf{W}_{LS}^H$  não é singular, embora sua última coluna contenha apenas zeros, pois seu posto é  $(N - L)$ . Assim, o produto  $\mathbf{W}_{LS}^\# \left( \mathbf{W}_{LS}^H \right)^\#$  equivale a  $\left( \mathbf{W}_{LS}^H \mathbf{W}_{LS} \right)^{-1}$ , permitindo-nos escrever:

$$\begin{aligned} \min \|\Delta \mathbf{x}\|_2^2 &= \min_{\mathbf{w}'} \begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix}^H \mathbf{X}^H \left( \mathbf{W}_{LS}^H \mathbf{W}_{LS} \right)^{-1} \mathbf{X} \begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix} \\ &= \min_{\mathbf{w}} \mathbf{w}^H \mathbf{X}^H \left( \mathbf{W}_{LS}^H \mathbf{W}_{LS} \right)^{-1} \mathbf{X} \mathbf{w} \end{aligned} \quad (10.8)$$

Como  $\left( \mathbf{W}_{LS}^H \mathbf{W}_{LS} \right)^{-1}$  difere de  $\left( \mathbf{W}^H \mathbf{W} \right)^{-1}$ , concluímos que (10.8) não equivale à minimização da função objetivo de Máxima Verossimilhança, dada por (8.11). Da mesma forma que resolvemos (8.11), podemos realizar esta minimização iterativamente através de:

$$\min_{\mathbf{w}^{(k+1)}} \left\| \left( \mathbf{W}_{LS}^{(k)H} \right)^\# \mathbf{X} \mathbf{w}^{(k+1)} \right\| \quad (10.9)$$

resolvendo-a no sentido TLS. Para o caso de exponenciais complexas sem amortecimento, adotamos uma formulação FBLP como descrita em (8.77) para o WTLS.

Na próxima seção definiremos uma generalização da formulação de De Moor [18] que nos possibilitará, em seguida, obter uma versão LS para o STLS.

## 10.2 Mínimos Quadrados Totais Estruturado com Elementos Fixos

De Moor [18] propôs uma formulação para o TLS convencional (não estruturado) quando apenas alguns elementos particulares da matriz  $\mathbf{X}$  podem ser alterados para reduzir seu posto, denominando-o TLS com Elementos Fixos. Este método é interessante em aplicações como, por exemplo, em reconstrução de sinais resultantes de medidas, onde apenas algumas amostras do sinal contêm erros de medição devido a imprecisões do instrumento de medida utilizado. Nesta seção, associamos a abordagem estruturada do STLS à formulação do problema TLS (não-estruturado) com elementos fixos. Ambos métodos foram propostos por De Moor, porém jamais associados, até onde temos notícia. Assim, acreditamos que a associação proposta aqui produz um novo método que é uma generalização dos métodos STLS e CTLS. Denominamos

este novo método de Mínimos Quadrados Totais Estruturado com Elementos Fixos ou, simplesmente, STLS com Elementos Fixos. Este método permite-nos tratar problemas estruturados onde temos certeza do valor exato de alguns elementos de  $\mathbf{X}$ , os quais não devem ser alterados na otimização.

Quando apenas alguns elementos de  $\mathbf{X}$  podem ser modificados no processo de otimização, é conveniente que separemos  $\mathbf{X}$  em duas matrizes,  $\mathbf{X} = \mathbf{X}_{\mathcal{F}} + \mathbf{X}_{\mathcal{M}}$ , onde a matriz  $\mathbf{X}_{\mathcal{F}}$  contém os elementos de  $\mathbf{X}$  que devem permanecer fixos e  $\mathbf{X}_{\mathcal{M}}$  aqueles suscetíveis de mudança. Isto é efetuado no método TLS com Elementos Fixos, onde a matriz  $\tilde{\mathbf{X}}$  não é estruturada. Porém, no método STLS decompomos a matriz  $\mathbf{X}$  como uma combinação afim das amostras disponíveis, dada por  $\mathbf{X} = \sum_{i=0}^{N-1} x_i \mathbf{X}_i + \mathbf{X}_N$ , na qual o conjunto de matrizes  $\{\mathbf{X}_i\}$  é o responsável pela estruturação de  $\mathbf{X}$ . Assim, para formularmos um método STLS com Elementos Fixos, devemos considerar que um mesmo elemento  $x_i$  que ocupa diferentes posições na matriz  $\mathbf{X}$ , determinadas pela matriz  $\mathbf{X}_i$ , pode ocupar simultaneamente uma posição na qual deve se manter fixo e outra na qual pode ser modificado pela otimização. A maneira que encontramos para associar estas duas abordagens aparentemente conflitantes foi separar cada matriz  $\mathbf{X}_i$  em duas outras:

$$\mathbf{X}_i = \mathbf{X}_{i\mathcal{F}} + \mathbf{X}_{i\mathcal{M}} \quad (10.10)$$

onde  $\mathbf{X}_{i\mathcal{F}}$  contém apenas os elementos de  $\mathbf{X}_i$  correspondentes aos elementos fixos de  $\mathbf{X}$ , e  $\mathbf{X}_{i\mathcal{M}}$  apenas os elementos de  $\mathbf{X}_i$  correspondentes àqueles de  $\mathbf{X}$  suscetíveis de sofrerem modificações na otimização. Dessa maneira, decompomos:

$$\begin{aligned} \mathbf{X} &= \mathbf{X}_{\mathcal{F}} + \mathbf{X}_{\mathcal{M}} \\ &= \sum_{i=0}^{N-1} x_i \mathbf{X}_i + \mathbf{X}_N \\ &= \sum_{i=0}^{N-1} x_i \mathbf{X}_{i\mathcal{F}} + \sum_{i=0}^{N-1} x_i \mathbf{X}_{i\mathcal{M}} + \mathbf{X}_N \end{aligned} \quad (10.11)$$

Entretanto, como alguns elementos de  $\mathbf{X}$  devem permanecer fixos, a matriz  $\tilde{\mathbf{X}}$ , aproximação de  $\mathbf{X}$ , deve obrigatoriamente possuir estes mesmos elementos fixos de  $\mathbf{X}$  em suas mesmas posições, ou seja,  $\tilde{\mathbf{X}} = \mathbf{X}_{\mathcal{F}} + \tilde{\mathbf{X}}_{\mathcal{M}}$ . Dessa forma, devemos decompor  $\tilde{\mathbf{X}}$  como:

$$\begin{aligned} \tilde{\mathbf{X}} &= \mathbf{X}_{\mathcal{F}} + \tilde{\mathbf{X}}_{\mathcal{M}} \\ &= \sum_{i=0}^{N-1} x_i \mathbf{X}_{i\mathcal{F}} + \sum_{i=0}^{N-1} \tilde{x}_i \mathbf{X}_{i\mathcal{M}} + \mathbf{X}_N \end{aligned} \quad (10.12)$$

tal que a otimização no contexto do método STLS com Elementos Fixos seja descrita por:

$$\min_{\mathbf{x}, \mathbf{w}} \sum_{i=0}^{N-1} (x_i - \tilde{x}_i)^* (x_i - \tilde{x}_i) \quad \text{tal que } \tilde{\mathbf{X}}\mathbf{w} = \mathbf{0} \text{ e } \mathbf{w}^H \mathbf{w} = 1 \quad (10.13)$$

sujeito a (10.11) e (10.12). Assim, tomando seu Lagrangiano, temos:

$$\mathcal{L}(\bar{\mathbf{x}}, \mathbf{w}, \mathbf{l}, \lambda) = \sum_{i=0}^{N-1} (x_i - \tilde{x}_i)^* (x_i - \tilde{x}_i) + \mathbf{l}^H \tilde{\mathbf{X}} \mathbf{w} + \lambda (1 - \mathbf{w}^H \mathbf{w}) \quad (10.14)$$

que, substituindo (10.12) em (10.14), resulta em:

$$\mathcal{L}(\bar{\mathbf{x}}, \mathbf{w}, \mathbf{l}, \lambda) = \sum_{i=0}^{N-1} (x_i - \tilde{x}_i)^* (x_i - \tilde{x}_i) + \mathbf{l}^H \left( \sum_{i=0}^{N-1} x_i \mathbf{X}_{i\mathcal{F}} + \sum_{i=0}^{N-1} \tilde{x}_i \mathbf{X}_{i\mathcal{M}} + \mathbf{X}_N \right) \mathbf{w} + \lambda (1 - \mathbf{w}^H \mathbf{w}) \quad (10.15)$$

Igualando a zero suas primeiras derivadas encontramos as seguintes restrições:

$$\begin{aligned} \tilde{\mathbf{X}} \mathbf{w} &= \mathbf{0} & \mathbf{w}^H \mathbf{w} &= 1 \\ \tilde{\mathbf{X}}^H \mathbf{l} &= \mathbf{w} \lambda^* & (x_i - \tilde{x}_i) &= \left( \mathbf{l}^H \mathbf{X}_{i\mathcal{M}} \mathbf{w} \right)^*, \quad i = 0, \dots, N-1 \end{aligned} \quad (10.16)$$

Utilizando (10.11) e (10.12), podemos escrever:

$$\begin{aligned} (\mathbf{X} - \tilde{\mathbf{X}}) \mathbf{w} &= \sum_{i=0}^{N-1} (x_i - \tilde{x}_i) \mathbf{X}_{i\mathcal{M}} \mathbf{w} \\ &= \sum_{i=0}^{N-1} \left( \mathbf{l}^H \mathbf{X}_{i\mathcal{M}} \mathbf{w} \right)^* \mathbf{X}_{i\mathcal{M}} \mathbf{w} \\ &= \sum_{i=0}^{N-1} (\mathbf{X}_{i\mathcal{M}} \mathbf{w}) (\mathbf{X}_{i\mathcal{M}} \mathbf{w})^H \mathbf{l} \end{aligned} \quad (10.17)$$

Considerando que em (10.16) temos  $\tilde{\mathbf{X}} \mathbf{w} = \mathbf{0}$ , a expressão anterior pode ser escrita como:

$$\mathbf{X} \mathbf{w} = \sum_{i=0}^{N-1} (\mathbf{X}_{i\mathcal{M}} \mathbf{w}) (\mathbf{X}_{i\mathcal{M}} \mathbf{w})^H \mathbf{l} \quad (10.18)$$

De maneira análoga, é possível demonstrar que:

$$\mathbf{X}^H \mathbf{l} = \sum_{i=0}^{N-1} \left( \mathbf{X}_{i\mathcal{M}}^H \mathbf{l} \right) \left( \mathbf{X}_{i\mathcal{M}}^H \mathbf{l} \right)^H \mathbf{w} \quad (10.19)$$

Então, fazendo a normalização  $\mathbf{z} = \mathbf{l} / \|\mathbf{l}\|$  e  $\|\mathbf{l}\| = \tau$  e definindo  $\mathbf{D}_{w\mathcal{M}} = \sum_{i=0}^{N-1} (\mathbf{X}_{i\mathcal{M}} \mathbf{w}) (\mathbf{X}_{i\mathcal{M}} \mathbf{w})^H$  e  $\mathbf{D}_{z\mathcal{M}} = \sum_{i=0}^{N-1} \left( \mathbf{X}_{i\mathcal{M}}^H \mathbf{z} \right) \left( \mathbf{X}_{i\mathcal{M}}^H \mathbf{z} \right)^H$ , podemos descrever o STLS com Elementos Fixos por:

$$\begin{aligned} \mathbf{X} \mathbf{w} &= \mathbf{D}_{w\mathcal{M}} \mathbf{z} \tau & \text{tal que} & & \mathbf{z}^H \mathbf{z} &= 1 \\ \mathbf{X}^H \mathbf{z} &= \mathbf{D}_{z\mathcal{M}} \mathbf{w} \tau & \text{tal que} & & \mathbf{w}^H \mathbf{w} &= 1 \end{aligned} \quad (10.20)$$

Esta expressão é idêntica a (9.27) que descreve o STLS no Capítulo 9, exceto pelas definições de  $\mathbf{D}_{w\mathcal{M}}$  e  $\mathbf{D}_{z\mathcal{M}}$ . Em princípio, a solução deste sistema de equações segue os mesmos passos da

solução do problema STLS. Porém, nem sempre é possível encontrar uma solução para o STLS com Elementos Fixos, como mostra o exemplo a seguir:

**Exemplo 10.1:** Reduza o posto de uma matriz  $\mathbf{X}$  particionada como:

$$\begin{aligned} \mathbf{X} &= \begin{bmatrix} 2 & 1 \\ 3 & 2 \\ 4 & 3 \end{bmatrix} \\ &= \mathbf{X}_{\mathcal{F}} + \mathbf{X}_{\mathcal{M}} \\ &= \begin{bmatrix} 2 & 0 \\ 3 & 2 \\ 4 & 3 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \end{aligned} \quad (10.21)$$

Obviamente, modificando apenas o elemento "1" não podemos reduzir o posto de  $\mathbf{X}$ . ■

No sentido de facilitar a compreensão destas questões e fornecer uma notação mais conveniente para a implementação do método acima descrito, podemos escrever a matriz  $\mathbf{X}_{\mathcal{M}}$  como:

$$\mathbf{X}_{\mathcal{M}} = \mathbf{X} .* \mathbf{M} \quad (10.22)$$

onde a operação ".\*" é o produto matricial elemento a elemento, aqui em notação Matlab<sup>®</sup>. A matriz  $\mathbf{M}$  tem as mesmas dimensões de  $\mathbf{X}$ ,  $(N - L) \times (L + 1)$ , e atua como uma máscara sobre esta matriz, de forma que, para  $k = 1, \dots, (N - L)$  e  $m = 1, \dots, (L + 1)$ ,  $\mathbf{M}_{k,m} = 0$  se  $\mathbf{X}_{k,m}$  for fixo, e  $\mathbf{M}_{k,m} = 1$  se  $\mathbf{X}_{k,m}$  for passível de mudança.

Dessa forma,  $\mathbf{X}_{i\mathcal{M}}$  pode também ser escrito como:

$$\mathbf{X}_{i\mathcal{M}} = \mathbf{X}_i .* \mathbf{M} \quad (10.23)$$

onde as matrizes  $\mathbf{X}_i$  são as mesmas utilizadas no método STLS. Isto nos leva à seguinte formulação explícita:

$$\begin{aligned} \mathbf{X}\mathbf{w} &= \left\{ \sum_{i=0}^{N-1} [(\mathbf{X}_i .* \mathbf{M}) \mathbf{w}] [(\mathbf{X}_i .* \mathbf{M}) \mathbf{w}]^H \right\} \mathbf{z}\tau & \text{tal que} & \quad \mathbf{z}^H \mathbf{z} = 1 \\ \mathbf{X}^H \mathbf{z} &= \left\{ \sum_{i=0}^{N-1} [(\mathbf{X}_i .* \mathbf{M})^H \mathbf{z}] [(\mathbf{X}_i .* \mathbf{M})^H \mathbf{z}]^H \right\} \mathbf{w}\tau & \text{tal que} & \quad \mathbf{w}^H \mathbf{w} = 1 \end{aligned} \quad (10.24)$$

ou seja, fazendo  $\mathbf{w} = \mathbf{v} / \|\mathbf{v}\|$ ,  $\mathbf{z} = \mathbf{u} / \|\mathbf{u}\|$  e  $\tau = \sigma \|\mathbf{u}\| \|\mathbf{v}\|$ , temos que:

$$\begin{aligned} \mathbf{X}\mathbf{v} &= \left\{ \sum_{i=0}^{N-1} [(\mathbf{X}_i .* \mathbf{M}) \mathbf{v}] [(\mathbf{X}_i .* \mathbf{M}) \mathbf{v}]^H \right\} \mathbf{u}\sigma & \text{tal que} & \quad \mathbf{u}^H \mathbf{D}_{v\mathcal{M}} \mathbf{u} = 1 \\ \mathbf{X}^H \mathbf{u} &= \left\{ \sum_{i=0}^{N-1} [(\mathbf{X}_i .* \mathbf{M})^H \mathbf{u}] [(\mathbf{X}_i .* \mathbf{M})^H \mathbf{u}]^H \right\} \mathbf{v}\sigma & \text{tal que} & \quad \mathbf{v}^H \mathbf{D}_{u\mathcal{M}} \mathbf{v} = 1 \end{aligned} \quad (10.25)$$

Seguindo ainda esta notação, de (10.16) podemos ainda escrever:

$$\tilde{x}_i = x_i - \mathbf{u}^H (\mathbf{X}_i .* \mathbf{M}) \mathbf{v} \sigma \quad (10.26)$$

e, finalmente:

$$\begin{aligned} \tilde{\mathbf{X}} &= \mathbf{X}_{\mathcal{F}} + \tilde{\mathbf{X}}_{\mathcal{M}} \\ &= \sum_{i=0}^{N-1} x_i (\mathbf{X}_i .* \bar{\mathbf{M}}) + \sum_{i=0}^{N-1} \tilde{x}_i (\mathbf{X}_i .* \mathbf{M}) + \mathbf{X}_N \end{aligned} \quad (10.27)$$

onde  $\bar{\mathbf{M}}$  é a matriz negação lógica da matriz  $\mathbf{M}$ , ou seja, é a matriz  $\mathbf{M}$  com seus elementos “1” trocados por “0” e seus elementos “0” trocados por “1”. No sentido de verificarmos o funcionamento deste método, a seguir, propomos um exemplo numérico simples.

**Exemplo 10.2:** Tomemos uma seqüência de dados  $\mathbf{x} \in \mathbb{R}^6$  onde  $\mathbf{x}^T = [6 \ 5 \ 4 \ 3 \ 2 \ 1]$ . Considerando  $L = 1$ , formamos a seguinte matriz Toeplitz  $(N - L) \times (L + 1)$ , ou seja,  $5 \times 2$ :

$$\mathbf{X} = \begin{bmatrix} 5 & 6 \\ 4 & 5 \\ 3 & 4 \\ 2 & 3 \\ 1 & 2 \end{bmatrix} \quad (10.28)$$

É fácil verificar que o posto desta matriz é completo e igual a  $(L + 1) = 2$ . Desejamos, agora, reduzir o posto dessa matriz, mantendo sua estrutura Toeplitz, porém fixando o segundo elemento da primeira coluna (“4”) e o primeiro (“6”), o terceiro (“4”) e o quinto (“2”) elemento da segunda coluna. Então, devemos utilizar uma matriz  $\mathbf{M}$  de máscara dada por:

$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \quad (10.29)$$

Aplicamos o método STLS com Elementos Fixos à matriz  $\mathbf{X}$ , através das equações (10.25), (10.26) e (10.27), e obtivemos uma matriz de aproximação  $\tilde{\mathbf{X}}$  dada por:

$$\tilde{\mathbf{X}} = \begin{bmatrix} 4,8990 & 6 \\ 4 & 4,8990 \\ 3,2660 & 4 \\ 2,6667 & 3,2660 \\ 1,6330 & 2 \end{bmatrix} \quad (10.30)$$

Observando esta matriz  $\tilde{\mathbf{X}}$ , verificamos que seu posto é incompleto e igual a  $L = 1$ . Podemos observar também que os elementos fixados mantiveram-se intactos durante a otimização. Além disso, na medida do possível, a estrutura Toeplitz foi preservada, sendo violada somente na

diagonal que envolve o quarto elemento da primeira coluna e o quinto da segunda, pois este último foi mantido fixo, como desejado. ■

Restringindo a formulação do método STLS com Elementos Fixos, podemos chegar aos diversos métodos estudados: considerando que todos os elementos da matriz de dados estendida possam ser alterados durante a otimização, chegamos ao STLS; se as matrizes de dados estendidas com ou sem ruído forem estruturadas da mesma forma, temos o CTLS; se não impusermos estrutura, caímos no caso TLS convencional; se considerarmos, além disso, que a coluna da matriz de dados estendida relativa ao vetor desejado deve permanecer fixa, nos recolhemos ao caso LS. Assim, o método STLS com Elementos Fixos é uma generalização do STLS, qualificando-se como o mais geral de todos os métodos de predição linear baseados nos critérios TLS e LS.

Dessa maneira, na próxima seção utilizaremos a notação anterior no sentido de formularmos uma versão LS para o método STLS.

### 10.3 Mínimos Quadrados Estruturado - SLS

Nesta seção, particularizamos o STLS com Elementos Fixos no sentido de estender a abordagem estruturada à otimização LS. Dessa forma, até onde temos notícia, produzimos um novo método que denominamos Mínimos Quadrados Estruturado (SLS). Isto ocorre porque podemos interpretar o LS como um caso particular do TLS onde, por exemplo, as perturbações sobre a primeira coluna de  $\mathbf{X} = [\mathbf{b} \mid \mathbf{A}]$  (ou seja,  $\mathbf{b}$ ) são desconsideradas no processo de otimização, de forma que a matriz de perturbações fica expressa apenas por  $\Delta\mathbf{X} = [\mathbf{0} \mid \Delta\mathbf{A}]$  e temos  $\tilde{\mathbf{X}} = [\mathbf{b} \mid \tilde{\mathbf{A}}]$ . Então, o LS pode ser formulado como um TLS com elementos fixos.

Na solução LS com redução de posto, como no FBLP Modificado, mantemos fixo o vetor  $\mathbf{b}$ , enquanto utilizamos a pseudo-inversa de uma aproximação de posto reduzido da matriz de dados,  $\mathbf{w}_{LSMod} = \mathbf{A}_{Mod}^{\#} \mathbf{b}$ . Dessa forma, podemos dizer que apenas os elementos de  $\mathbf{A}$  foram modificados, enquanto os elementos de  $\mathbf{b}$  permaneceram fixos na otimização. Isto nos inspira a formular o SLS como um caso particular do STLS com Elementos Fixos, apenas fixando a matriz de máscara como:

$$\mathbf{M} = \begin{bmatrix} 0 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \cdots & 1 \end{bmatrix} \quad (10.31)$$

Assim, escrevendo as matrizes  $\mathbf{X}_i$  como  $\mathbf{X}_i = [\mathbf{b}_i \mid \mathbf{A}_i]$ , onde  $\mathbf{A}_i$  e  $\mathbf{b}_i$  compõem-se de "0" e "1", cada matriz  $\mathbf{X}_{iM}$  ficaria expressa por:

$$\begin{aligned} \mathbf{X}_{iM} &= \mathbf{X}_i .* \mathbf{M} \\ &= [\mathbf{b}_i \mid \mathbf{A}_i] .* \mathbf{M} \\ &= [\mathbf{0} \mid \mathbf{A}_i] \end{aligned} \quad (10.32)$$

Substituindo este valor de  $\mathbf{X}_{iM}$  em (10.24), encontramos o seguinte conjunto de equações para o SLS:

$$\begin{aligned} \mathbf{X}\mathbf{w} &= \left\{ \sum_{i=0}^{N-1} \left( \begin{bmatrix} \mathbf{0} & \mathbf{A}_i \end{bmatrix} \mathbf{w} \right) \left( \begin{bmatrix} \mathbf{0} & \mathbf{A}_i \end{bmatrix} \mathbf{w} \right)^H \right\} \mathbf{z}\tau & \text{tal que} & \quad \mathbf{z}^H \mathbf{z} = 1 \\ \mathbf{X}^H \mathbf{z} &= \left\{ \sum_{i=0}^{N-1} \left( \begin{bmatrix} \mathbf{0} & \mathbf{A}_i \end{bmatrix}^H \mathbf{z} \right) \left( \begin{bmatrix} \mathbf{0} & \mathbf{A}_i \end{bmatrix}^H \mathbf{z} \right)^H \right\} \mathbf{w}\tau & \text{tal que} & \quad \mathbf{w}^H \mathbf{w} = 1 \end{aligned} \quad (10.33)$$

No sentido de obter uma expressão em função de  $\mathbf{A}$  e de  $\mathbf{b}$ , particionamos:

$$\mathbf{w} = \begin{bmatrix} w_0 \\ \mathbf{w}' \end{bmatrix} \quad (10.34)$$

Substituindo em (10.33), encontramos:

$$\mathbf{X}\mathbf{w} = \left\{ \sum_{i=0}^{N-1} (\mathbf{A}_i \mathbf{w}') (\mathbf{A}_i \mathbf{w}')^H \right\} \mathbf{z}\tau \quad (10.35)$$

e também calculamos:

$$\begin{aligned} \begin{bmatrix} \mathbf{b}^H \mathbf{z} \\ \mathbf{A}^H \mathbf{z} \end{bmatrix} &= \left\{ \sum_{i=0}^{N-1} \begin{bmatrix} 0 \\ \mathbf{A}_i^H \mathbf{z} \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{A}_i^H \mathbf{z} \end{bmatrix}^H \right\} \begin{bmatrix} w_0 \\ \mathbf{w}' \end{bmatrix} \tau \\ &= \sum_{i=0}^{N-1} \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & (\mathbf{A}_i^H \mathbf{z}) (\mathbf{A}_i^H \mathbf{z})^H \end{bmatrix} \begin{bmatrix} w_0 \\ \mathbf{w}' \end{bmatrix} \tau \\ &= \sum_{i=0}^{N-1} \begin{bmatrix} 0 \\ (\mathbf{A}_i^H \mathbf{z}) (\mathbf{A}_i^H \mathbf{z})^H \mathbf{w}' \end{bmatrix} \tau \end{aligned} \quad (10.36)$$

encontrando:

$$\mathbf{A}^H \mathbf{z} = \left\{ \sum_{i=0}^{N-1} (\mathbf{A}_i^H \mathbf{z}) (\mathbf{A}_i^H \mathbf{z})^H \right\} \mathbf{w}'\tau \quad (10.37)$$

e

$$\mathbf{b}^H \mathbf{z} = 0 \quad (10.38)$$

Então, definindo  $\mathbf{D}_{Aw'} = \left\{ \sum_{i=0}^{N-1} (\mathbf{A}_i \mathbf{w}') (\mathbf{A}_i \mathbf{w}')^H \right\}$  e  $\mathbf{D}_{Az} = \left\{ \sum_{i=0}^{N-1} (\mathbf{A}_i^H \mathbf{z}) (\mathbf{A}_i^H \mathbf{z})^H \right\}$  o método SLS pode finalmente ser descrito pelas expressões:

$$\begin{cases} \mathbf{X}\mathbf{w} = \mathbf{D}_{Aw'} \mathbf{z}\tau & \text{tal que} \quad \mathbf{z}^H \mathbf{z} = 1 \\ \mathbf{A}^H \mathbf{z} = \mathbf{D}_{Az} \mathbf{w}'\tau \\ \mathbf{b}^H \mathbf{z} = 0 \end{cases} \quad \text{tal que} \quad \mathbf{w}^H \mathbf{w} = 1 \text{ e } |w_0| \neq \{0 \text{ ou } 1\} \quad (10.39)$$

Normalmente, no critério LS utilizamos  $w_0 = -1$ , de forma a escrevermos a equação de predição como  $\tilde{\mathbf{A}}\mathbf{w}' = \mathbf{b}$ . Porém, no método SLS partimos da consideração  $\mathbf{w}^H\mathbf{w} = 1$ , originária da formulação do método STLS na qual nos baseamos. Isto nos impede de impor  $w_0 = -1$ , pois teria como conseqüência que  $\|\mathbf{w}'\| = 0$ . Entretanto, vimos no Capítulo 8 que podemos realizar a otimização iterativamente a partir de  $\mathbf{w}^H\mathbf{w} = 1$ , obtendo um vetor  $\mathbf{w}'$  ótimo que satisfaz  $\tilde{\mathbf{A}}\mathbf{w}' = -w_0\mathbf{b}$ . Para ajustarmos esta igualdade, basta ponderar o vetor  $\mathbf{w}'$  dividindo-o por  $-w_0$ , de forma a garantir que  $\tilde{\mathbf{A}}\mathbf{w}' = \mathbf{b}$ .

A solução de (10.39) pode ser obtida iterativamente a partir de  $\mathbf{w}^{(k+1)} = \mathbf{X}^\# \mathbf{D}_{Aw'}^{(k)} \mathbf{z}^{(k)} \tau^{(k)}$  e de  $\mathbf{z}^{(k+1)} = (\mathbf{X}^H)^\# \begin{bmatrix} 0 \\ \mathbf{D}_{Az}^{(k)} \mathbf{w}'^{(k+1)} \tau^{(k+1)} \end{bmatrix}$ .

Utilizando (10.16) e (10.32), a norma do erro a ser minimizado pelo método SLS é dada por:

$$\begin{aligned}
 \sum_{i=0}^{N-1} (x_i - \tilde{x}_i) (x_i - \tilde{x}_i)^* &= \sum_{i=0}^{N-1} (\mathbf{1}^H \mathbf{X}_{iM} \mathbf{w}) (\mathbf{1}^H \mathbf{X}_{iM} \mathbf{w})^* \\
 &= \sum_{i=0}^{N-1} (\mathbf{z}^H \mathbf{X}_{iM} \mathbf{w}) (\mathbf{z}^H \mathbf{X}_{iM} \mathbf{w})^* \tau^2 \\
 &= \mathbf{z}^H \sum_{i=0}^{N-1} \left( \begin{bmatrix} 0 & \mathbf{A}_i \end{bmatrix} \mathbf{w} \right) \left( \mathbf{z}^H \begin{bmatrix} 0 & \mathbf{A}_i \end{bmatrix} \mathbf{w} \right)^* \tau^2 \\
 &= \mathbf{z}^H \sum_{i=0}^{N-1} (\mathbf{A}_i \mathbf{w}') (\mathbf{A}_i \mathbf{w}')^H \mathbf{z} \tau^2 \\
 &= \mathbf{z}^H \mathbf{D}_{Aw'} \mathbf{z} \tau^2
 \end{aligned} \tag{10.40}$$

Como o posto de  $\mathbf{D}_{Aw'}$  é completo, a partir de (10.39), podemos afirmar que:

$$\mathbf{z} \tau = \mathbf{D}_{Aw'}^{-1} \mathbf{X} \mathbf{w} \tag{10.41}$$

Substituindo em (10.40), temos:

$$\begin{aligned}
 \sum_{i=0}^{N-1} (x_i - \tilde{x}_i) (x_i - \tilde{x}_i)^* &= \mathbf{w}^H \mathbf{X}^H \left( \mathbf{D}_{Aw'}^{-1} \right)^H \mathbf{D}_{Aw'} \mathbf{D}_{Aw'}^{-1} \mathbf{X} \mathbf{w} \\
 &= \mathbf{w}^H \mathbf{X}^H \left( \mathbf{D}_{Aw'}^{-1} \right)^H \mathbf{X} \mathbf{w} \\
 &= \mathbf{w}^H \mathbf{X}^H \mathbf{D}_{Aw'}^{-1} \mathbf{X} \mathbf{w}
 \end{aligned} \tag{10.42}$$

pois  $\mathbf{D}_{Aw'}$  é Hermitiana por construção.

Concentrando-nos agora em  $\mathbf{D}_{Aw'}$ , podemos escrevê-la como:

$$\begin{aligned}
 \mathbf{D}_{Aw'} &= \sum_{i=0}^{N-1} (\mathbf{A}_i \mathbf{w}') (\mathbf{A}_i \mathbf{w}')^H \\
 &= \left[ \mathbf{A}_0 \mathbf{w}' \mid \cdots \mid \mathbf{A}_{(N-1)} \mathbf{w}' \right] \left[ \mathbf{A}_0 \mathbf{w}' \mid \cdots \mid \mathbf{A}_{(N-1)} \mathbf{w}' \right]^H
 \end{aligned} \tag{10.43}$$

Lembrando que:

$$\mathbf{F}_k = \left[ \mathbf{X}_0(:, k) \mid \mathbf{X}_1(:, k) \mid \cdots \mid \mathbf{X}_{(N-1)}(:, k) \right] \quad (10.44)$$

onde  $\mathbf{X}_i(:, k)$  é a  $k$ -ésima coluna de  $\mathbf{X}_i$ , para  $k = 0, \dots, L$ , a partir de (10.32) podemos escrever:

$$\mathbf{F}_0 = \left[ \mathbf{b}_0 \mid \mathbf{b}_1 \mid \cdots \mid \mathbf{b}_{(N-1)} \right] \quad (10.45)$$

e também, para  $k = 1, \dots, L$ :

$$\mathbf{F}_k = \left[ \mathbf{A}_0(:, k) \mid \mathbf{A}_1(:, k) \mid \cdots \mid \mathbf{A}_{(N-1)}(:, k) \right] \quad (10.46)$$

onde agora numeramos as colunas de  $\mathbf{A}_i$  de 1 a  $L$ , simplesmente. Recordando que na formulação do método CLS definimos:

$$\mathbf{W}_{LS}^H = \sum_{k=1}^L \mathbf{F}_k w_{(k)} \quad (10.47)$$

podemos substituir (10.46) em (10.47), obtendo:

$$\begin{aligned} \mathbf{W}_{LS}^H &= \sum_{k=1}^L \left[ \mathbf{A}_0(:, k) \mid \mathbf{A}_1(:, k) \mid \cdots \mid \mathbf{A}_{(N-1)}(:, k) \right] w_{(k)} \\ &= \left[ \sum_{k=1}^L \mathbf{A}_0(:, k) w_{(k)} \mid \cdots \mid \sum_{k=1}^L \mathbf{A}_{(N-1)}(:, k) w_{(k)} \right] \\ &= \left[ \mathbf{A}_0 \mathbf{w}' \mid \cdots \mid \mathbf{A}_{(N-1)} \mathbf{w}' \right] \end{aligned} \quad (10.48)$$

Substituindo (10.48) em (10.43), obtemos:

$$\mathbf{D}_{Aw'} = \mathbf{W}_{LS}^H \mathbf{W}_{LS} \quad (10.49)$$

Portanto, a norma do erro a ser minimizado pelo SLS é dada por:

$$\begin{aligned} \sum_{i=0}^{N-1} (x_i - \tilde{x}_i) (x_i - \tilde{x}_i)^* &= \mathbf{w}^H \mathbf{X}^H \mathbf{D}_{Aw'}^{-1} \mathbf{X} \mathbf{w} \\ &= \mathbf{w}^H \mathbf{X}^H \left( \mathbf{W}_{LS}^H \mathbf{W}_{LS} \right)^{-1} \mathbf{X} \mathbf{w} \end{aligned} \quad (10.50)$$

Esta expressão é idêntica a (10.8) e, portanto, permite-nos concluir, enfim, que os métodos SLS e CLS são equivalentes.

## 10.4 Mínimos Quadrados Branqueado - WLS

Nesta seção particularizaremos a proposta de Hua e Sarkar [37] para o caso em que existem erros apenas na matriz de dados  $\mathbf{A}$ , definindo o método Mínimos Quadrados Branqueado (WLS). Recordando (8.68), a presença de ruído acaba por gerar um erro de predição dado por:

$$\mathbf{X}\mathbf{w} = \mathbf{e} \quad (10.51)$$

Conforme a formulação do CTLS, podemos escrever (10.51) como:

$$\begin{aligned} (\mathbf{X} + \Delta\mathbf{X})\mathbf{w} &= \mathbf{0} \\ \mathbf{X}\mathbf{w} &= -\Delta\mathbf{X}\mathbf{w} \end{aligned} \quad (10.52)$$

logo:

$$\begin{aligned} \mathbf{e} &= -\Delta\mathbf{X}\mathbf{w} \\ &= -\left[ \Delta\mathbf{b} \mid \Delta\mathbf{A} \right] \begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix} \end{aligned} \quad (10.53)$$

Entretanto, considerando perturbações apenas na matriz  $\mathbf{A}$ , escrevemos simplesmente:

$$\begin{aligned} \mathbf{e} &= -\left[ \mathbf{0} \mid \Delta\mathbf{A} \right] \begin{bmatrix} -1 \\ \mathbf{w}' \end{bmatrix} \\ &= -\Delta\mathbf{A}\mathbf{w}' \end{aligned} \quad (10.54)$$

ou seja, o erro está associado apenas à matriz  $\mathbf{A}$  e à partição  $\mathbf{w}'$  de  $\mathbf{w}$ . Porém, utilizando (10.5), podemos afirmar que  $\Delta\mathbf{A}\mathbf{w}' = \mathbf{W}_{LS}^H \Delta\mathbf{x}$ . Portanto, o erro de predição pode ser escrito como:

$$\mathbf{e} = -\mathbf{W}_{LS}^H \Delta\mathbf{x} \quad (10.55)$$

Calculamos a matriz de covariância deste vetor de erro de predição fazendo:

$$\begin{aligned} E\{\mathbf{e}\mathbf{e}^H\} &= E\{\mathbf{W}_{LS}^H \Delta\mathbf{x} \Delta\mathbf{x}^H \mathbf{W}_{LS}\} \\ &= \mathbf{W}_{LS}^H E\{\Delta\mathbf{x} \Delta\mathbf{x}^H\} \mathbf{W}_{LS} \end{aligned} \quad (10.56)$$

Porém, conforme Abatzoglou e Mendel [2],  $E\{\Delta\mathbf{x} \Delta\mathbf{x}^H\} = E\{\mathbf{y}\mathbf{y}^H\} = \sigma_y^2 \mathbf{I}$ . Dessa forma, concluímos que:

$$E\{\mathbf{e}\mathbf{e}^H\} = \sigma_y^2 \mathbf{W}_{LS}^H \mathbf{W}_{LS} \quad (10.57)$$

Assim, devido à estrutura da matriz de dados (expressa pela matriz  $\mathbf{W}_{LS}^H$ ), verificamos que o vetor  $\mathbf{e}$  não é branco e, para produzirmos o erro de estimação  $\mathbf{e}_w$  ótimo, utilizamos  $(\mathbf{W}_{LS}^H \mathbf{W}_{LS})^{-1/2}$ , raiz quadrada de  $(\mathbf{W}_{LS}^H \mathbf{W}_{LS})^{-1}$ , para branquear o vetor  $\mathbf{e}$ :

$$(\mathbf{W}_{LS}^H \mathbf{W}_{LS})^{-1/2} \mathbf{X} \mathbf{w} = \mathbf{e}_w \quad (10.58)$$

Então, minimizarmos a norma do erro  $\mathbf{e}_w$  corresponde a fazermos:

$$\begin{aligned} \min (\mathbf{e}_w^H \mathbf{e}_w) &= \min_{\mathbf{w}} \left\{ \mathbf{w}^H \mathbf{X}^H \left[ (\mathbf{W}_{LS}^H \mathbf{W}_{LS})^{-1/2} \right]^H (\mathbf{W}_{LS}^H \mathbf{W}_{LS})^{-1/2} \mathbf{X} \mathbf{w} \right\} \\ &= \min_{\mathbf{w}} \left\{ \mathbf{w}^H \mathbf{X}^H \left[ (\mathbf{W}_{LS}^H \mathbf{W}_{LS})^H \right]^{-1/2} (\mathbf{W}_{LS}^H \mathbf{W}_{LS})^{-1/2} \mathbf{X} \mathbf{w} \right\} \\ &= \min_{\mathbf{w}} \left\{ \mathbf{w}^H \mathbf{X}^H (\mathbf{W}_{LS}^H \mathbf{W}_{LS})^{-1/2} (\mathbf{W}_{LS}^H \mathbf{W}_{LS})^{-1/2} \mathbf{X} \mathbf{w} \right\} \\ &= \min_{\mathbf{w}} \left\{ \mathbf{w}^H \mathbf{X}^H (\mathbf{W}_{LS}^H \mathbf{W}_{LS})^{-1} \mathbf{X} \mathbf{w} \right\} \end{aligned} \quad (10.59)$$

Assim, a solução WLS é dada pelo vetor  $\mathbf{w}$  que minimiza (10.59). Novamente, (10.59) é equivalente a (10.50) e a (10.8) e, portanto, os métodos CLS, SLS e WTLS minimizam a norma do mesmo erro e fornecem a mesma solução. A solução WLS é obtida iterativamente, otimizando o vetor  $\mathbf{w}^{(k+1)}$  do espaço "quase" nulo da matriz  $\mathbf{R}_{wLS}^{(k)} = \mathbf{X}^H (\mathbf{W}_{LS(k)}^H \mathbf{W}_{LS(k)})^{-1} \mathbf{X}$ , ou seja:

$$\mathbf{R}_{wLS}^{(k)} \mathbf{w}^{(k+1)} = \mathbf{0} \quad (10.60)$$

cuja solução é obtida no sentido TLS a partir da SVD de  $\mathbf{R}_{wLS}^{(k)}$  até a convergência, como descrito no Capítulo 8. Para o caso de exponenciais complexas não-amortecidas, uma versão FBLP para o método WLS pode ser obtida fazendo-se simplesmente:

$$\mathbf{R}_{wLS}^{(k)} = \mathbf{X}^H (\mathbf{W}_{LS(k)}^H \mathbf{W}_{LS(k)})^{-1} \mathbf{X} + \tilde{\mathbf{I}} \left[ \mathbf{X}^H (\mathbf{W}_{LS(k)}^H \mathbf{W}_{LS(k)})^{-1} \mathbf{X} \right]^* \tilde{\mathbf{I}} \quad (10.61)$$

Esta expressão é obtida a partir das relações de simetria entre  $\mathbf{X}_f$  e  $\mathbf{X}_b$ .

### 10.4.1 Desempenho

No sentido de verificar o desempenho do método WLS para a estimação de frequências, na figura 10.1 apresentamos sua curva de desempenho para  $L = 24$ , bem como a curva do método FBLP para a mesma ordem. Utilizamos no WLS sua versão FBLP dada pela equação (10.61), aplicada sobre 20 experimentos, conforme efetuado no WTLS. Podemos observar que o método WLS mostrou-se ligeiramente menos preciso do que o FBLP, pois sua curva de desempenho posicionou-se mais afastada do Limite de Cramér-Rao. Porém, o método WLS apresentou limiar somente em 6 dB de SNR, constituindo uma redução de 9 dB em relação à SNR de limiar do

método FBLP para  $L = 24$  e de 1 dB em relação á SNR de limiar do FBLP Modificado para  $L = 17$ . Ainda enfrentamos problemas numéricos no cálculo da curva de desempenho do WLS para  $L = 17$ , porém temos a esperança de que seu limiar ocorra bem próximo dos 3 dB de SNR da ML. Assim, embora não possamos afirmar que os métodos CLS, SLS e WLS atendem ao critério ML, verificamos que seu desempenho demonstra significativa melhora em relação àquele do FBLP não estruturado. Sua introdução vem oferecer caminhos alternativos para a pesquisa de procedimentos matemáticos que permitam melhorar sua precisão, tornando-os competitivos com os métodos iterativos de ML já conhecidos.

## 10.5 Conclusões

Neste capítulo propusemos quatro métodos que, até onde temos notícia, são inéditos. Destacamos o método STLS com Elementos Fixos, o qual é uma generalização do STLS, qualificando-se como o mais geral de todos os métodos de predição linear baseados nos critérios TLS e LS. Restringindo sua formulação, ao considerarmos que a coluna da matriz de dados estendida relativa ao vetor desejado deve permanecer fixa, pudemos incorporar a estrutura da matriz de dados na otimização LS, originando o método inédito Mínimos Quadrados Restrito Estruturado (SLS). Também formulamos versões LS para os métodos CTLS e WTLS, produzindo os novos métodos CLS e WLS, respectivamente. Demonstramos a equivalência destes três métodos e verificamos que seu desempenho apresentou limiar somente em 6 dB de SNR, constituindo uma redução de 9 dB em relação à SNR de limiar do método FBLP para a ordem de predição 24 e também de 1 dB para o melhor desempenho do FBLP Modificado.

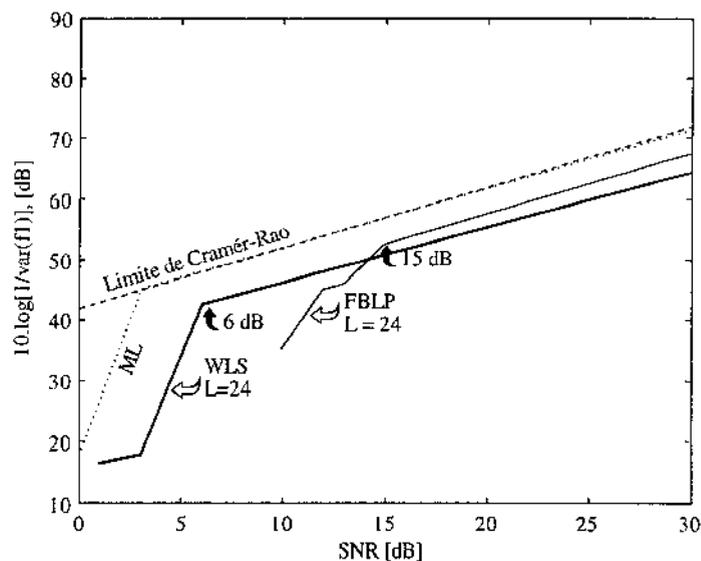


Figura 10.1: Desempenho do método WLS, na versão FBLP, para  $L = 24$ .

# Capítulo 11

## Conclusões e Sugestões

### 11.1 Conclusões

Neste trabalho estudamos a estimação de frequências nos casos críticos que conjugam a disponibilidade de apenas um número reduzido de amostras do sinal, a diferença entre as frequências a serem estimadas menor do que o recíproco do intervalo de observação e a relação sinal-ruído (SNR) assumindo valores tão baixos quanto 3 dB.

No Capítulo 2, reproduzimos o desenvolvimento da teoria da estimação de frequências pelo critério de Máxima Verossimilhança (ML), de forma a esclarecer as dificuldades normalmente encontradas nesse tipo de estimação. Pudemos verificar que a minimização direta da função objetivo de Máxima Verossimilhança exige uma busca multidimensional, que é computacionalmente proibitiva para a estimação de várias frequências.

No sentido de definir o limite máximo de desempenho para os métodos de estimação de frequências estudados, no Capítulo 3 reproduzimos o cálculo do Limite de Cramér-Rao (CRB), como proposto por Rife e Boorstin. Pudemos observar que, mesmo para o caso simples de apenas duas exponenciais complexas, a obtenção de uma expressão analítica explícita para os CRB's demanda muito esforço, o que explica porque esses limites são calculados somente numericamente. As expressões (3.49) e (3.50) permitiram-nos observar que, quando as amplitudes das exponenciais são iguais, seus CRB's também são iguais. Além disso, esses limites não dependem dos valores individuais das frequências e fases, mas sim de suas respectivas diferenças. Também verificamos que, para frequências muito afastadas, os CRB's confundem-se com aquele referente ao caso de uma única exponencial e que, conforme [85], esses limites referentes a duas exponenciais complexas são sempre maiores ou iguais àquele referente uma única exponencial complexa.

No Capítulo 4, através de métodos baseados em predição linear no contexto *forward-backward*, mostramos que podemos estimar as frequências a partir dos zeros de um polinômio definido pelos coeficientes de um filtro de erro de predição, projetado segundo o método FBLP. Porém, para valores de SNR abaixo de 12 dB, o método FBLP mostrou-se ineficiente. Procurando analisar intrinsecamente os efeitos do ruído sobre a estimação, no Capítulo 5 detalhamos a decomposição em valores singulares (SVD) como ferramenta de descrição dos subespaços de

sinal e de ruído. No Capítulo 6, utilizamos a SVD para restringir a solução FBLP ao subespaço de sinal, o que define o método FBLP Modificado e resulta em melhora significativa do desempenho, fazendo a SNR de limiar cair para cerca de 7 dB.

No Capítulo 7, utilizamos o critério de Mínimos Quadrados Totais (TLS) na otimização do método FBLP. O critério TLS admite erros não somente no vetor desejado  $\mathbf{b}$ , como efetuado por Mínimos Quadrados convencional (LS), mas também na matriz de dados  $\mathbf{A}$  e otimiza o vetor  $\mathbf{w}$  de coeficientes de predição através de um sistema homogêneo de equações, como no método de Pisarenko. Isto nos permitiu propor um método para a estimação de frequências, o qual denominamos TLS-FBLP, que supera significativamente o desempenho do FBLP convencional, reduzindo a SNR de limiar para a estimação. Melhoramos nossa proposta inicial, formulando uma versão mais econômica do TLS-FBLP que utiliza o subespaço de sinal. Demonstramos, de maneira inédita, que a solução TLS pode ser interpretada como uma solução LS modificada onde a matriz de dados e o vetor desejado sofrem restrição conjunta ao subespaço de sinal. Comparando os desempenhos do TLS-FBLP com o do método FBLP Modificado de Tufts e Kumaresan, pudemos observar que este último, no melhor caso, apresenta uma SNR de limiar cerca de 1 dB inferior àquela do método proposto. Entretanto, além desta diferença ser pequena, o método TLS-FBLP é menos sensível à variação da ordem de predição. Verificamos também que os melhores desempenhos dos métodos ocorrem quando a matriz de dados ou a matriz de dados estendida são aproximadamente quadradas. No caso de as matrizes serem exatamente quadradas e de posto cheio, as otimizações LS e TLS fornecem soluções únicas para  $\mathbf{w}$ . Isto define uma indicação analítica para as ordens correspondentes aos melhores desempenhos do FBLP Modificado e TLS-FBLP, superando a definição empírica de Tufts e Kumaresan.

Como dito acima, constatamos, com surpresa, que o melhor desempenho do TLS-FBLP é ligeiramente inferior ao do FBLP Modificado. Além disso, os desempenhos de ambos mostraram-se inferiores ao da Máxima Verossimilhança. Acreditamos que esta sub-otimalidade se deve ao uso do truncamento da SVD, o qual não preserva a estrutura que é inserida pelo filtro de predição linear na composição da matriz de dados. Quando descartamos a contribuição do subespaço de ruído, produzindo uma matriz corrigida, descartamos também a estrutura original da matriz de dados. Embora esta correção proporcione significativa melhora no desempenho da estimação de frequências, ela desestrutura o esquema de predição linear. A partir dessas observações concluímos que o TLS-FBLP é mais sensível à perda de estrutura do que o FBLP Modificado, pois no primeiro a correção atinge tanto a matriz de dados quanto o vetor desejado, enquanto que no segundo o vetor de dados é preservado.

No Capítulo 8 detalhamos a formulação dos métodos recentes Máxima Verossimilhança Quadrática Iterativa (IQML), Mínimos Quadrados Totais Restrito (CTLS) e Mínimos Quadrados Totais Branqueado (WTLS), os quais buscam atingir o mesmo desempenho da ML. Demonstramos analiticamente que estes métodos minimizam a norma do mesmo erro sujeito às mesmas restrições e fornecem, portanto, soluções equivalentes. Conseqüentemente, o problema de minimização quadrático no IQML pode também ser interpretado como um problema TLS. Os desempenhos destes métodos superaram os do FBLP Modificado e do TLS-FBLP, porém a um custo computacional maior, devido a utilizarem pelo menos 5 iterações para cada estimativa do vetor de coeficientes. Encontramos certa dificuldade na implementação desses métodos, devido à instabilidade dos algoritmos disponíveis. Embora não seja nosso principal objetivo neste trabalho, buscamos na literatura e também propusemos alguns procedimentos matemáticos que

permitissem corrigir esta deficiência. Porém, salientamos que a melhoria da estabilidade desses métodos ainda é matéria de muitas pesquisas na comunidade científica internacional [17].

No Capítulo 9 destacamos a formulação do método Mínimos Quadrados Totais Estruturado (STLS), por ser ele o mais abrangente e mais recente dentre os métodos que impõem a manutenção da estrutura da matriz de dados na solução pelo critério TLS. Procuramos descrevê-lo em maiores detalhes e de maneira didática, devido à originalidade de sua formulação. Demonstramos que o método STLS é equivalente ao CTLS quando a matriz de dados estendida deve ser aproximada por uma matriz de mesma estrutura. Porém, quando estas matrizes são estruturadas de forma diferente a equivalência não mais se verifica. Isto acontece, por exemplo, quando a matriz de dados isenta de ruído possui estrutura diferente da apresentada pela matriz de dados ruidosa. Levando em consideração os resultados do Capítulo 8, pudemos afirmar também que o método STLS é também equivalente aos métodos WTLS e IQML. Esta equivalência envolve todos os métodos mais recentes para a estimação de parâmetros. Na literatura internacional não encontramos descrição tão abrangente, atualizada e detalhada, o que nos leva a crer que este resultado constitui-se em uma contribuição significativa à compreensão dos métodos de estimação em predição linear.

No Capítulo 10 propusemos quatro métodos. Entre eles destacamos o método STLS com Elementos Fixos que é uma generalização do STLS, qualificando-se como o mais geral de todos os métodos de predição linear baseados nos critérios TLS e LS. Restringindo sua formulação, podemos chegar aos diversos métodos estudados. Considerando que todos os elementos da matriz de dados estendida possam ser alterados durante a otimização, porém de forma a manter a estrutura da matriz de dados, chegamos ao STLS. Daí, se as matrizes de dados estendidas com ou sem ruído forem estruturadas da mesma forma, temos o CTLS. Se não impusermos estrutura, caímos no caso TLS convencional. Ainda mais, se considerarmos que a coluna da matriz de dados estendida relativa ao vetor desejado deve permanecer fixa, nos recolhemos ao caso LS. Esta última consideração permitiu-nos incorporar a estrutura da matriz de dados na otimização LS, originando os métodos inéditos Mínimos Quadrados Restrito (CLS), Mínimos Quadrados Branqueado (WLS) e Mínimos Quadrados Estruturado (SLS). Demonstramos a equivalência destes três métodos e verificamos que, para a ordem de predição 24, seu desempenho acompanhou o Limite de Cramér-Rao até 6 dB de SNR, ficando 9 dB abaixo da SNR de limiar do método FBLP (não estruturado) para a mesma ordem e também 1 dB abaixo da menor SNR de limiar do método FBLP Modificado.

## 11.2 Questões para Prosseguimento

Por mais que tenhamos discutido e comparado os diversos métodos estudados, muito ainda há que se investigar na área de estimação de frequências via predição linear. Verificamos que métodos baseados em predição linear, como CTLS, WTLS, STLS e IQML, já permitem aproximar o desempenho da Máxima Verossimilhança sob um custo computacional menor do que o exigido pela maximização direta da função objetivo de verossimilhança. Porém, o esforço computacional desses métodos ainda é muito superior ao exigido pelos métodos não-iterativos como o TLS-FBLP e o FBLP Modificado. Assim, resta agora a busca de algoritmos mais simples, robustos e eficientes para a solução dos métodos iterativos aqui estudados.

O método IQML utiliza simplesmente a predição linear *forward* para estimar as frequências. No sentido de manter os zeros sobre a circunferência de raio unitário (CRU), lança mão da restrição de simetria sobre os coeficientes do filtro de predição. Isto é eficiente apenas se utilizarmos a ordem de predição igual ao número de frequências a estimar,  $L = M$ . Para ordens superiores, a imposição da restrição de simetria faria com que os zeros espúrios também se posicionassem sobre a CRU, o que inviabilizaria a seleção dos zeros de sinal. A predição linear *forward-backward*, por sua vez, força os zeros de sinal a se posicionarem sobre a CRU, enquanto os zeros espúrios permanecem no interior da mesma, o que nos permite utilizar um critério de seleção. Pensando nisso, a formulação de uma versão FBLP para o método IQML permitiria a utilização de ordens de predição mais altas, como no WTLS.

Uma análise crítica do método STLS, permite observar que a solução por ele obtida, a partir da SVD Riemanniana, é dada somente em termos do vetor singular à direita relativo ao menor valor singular. Quando  $L = M$ , esta é realmente a forma de solução adequada, pois nesse caso a dimensão do subespaço de ruído é 1, ou seja, o menor valor singular é também o único valor singular do subespaço de ruído. Ao utilizarmos ordens de predição maiores, a dimensão do subespaço de ruído seria igual à multiplicidade do menor valor singular. Neste caso, a melhor solução seria calculada combinando-se os vetores singulares do subespaço de ruído ou, alternativamente, aqueles relativos ao subespaço de sinal. Isto exigiria uma revisão da formulação da SVD Riemanniana, no sentido de utilizar uma combinação dos diversos vetores singulares  $u_i$  e  $v_i$  relativos ao subespaço de ruído, ou mesmo aqueles relativos ao subespaço de sinal, ao invés simplesmente os vetores singulares relativos ao último valor singular.

Outra linha de pesquisa interessante é a utilização da modelagem por espaço de estados ao invés de por função de transferência. Algumas simulações preliminares do método proposto por Kung, Arun e Bhaskar Rao [50] mostraram que a modelagem por espaço de estados aumentou a robustez do FBLP Modificado e do TLS-FBLP, reduzindo suas SNR's de limiar para 4 dB. Entretanto, notamos que sua curva de desempenho manteve-se sempre muito afastada do Limite de Cramér-Rao, indicando pobreza de precisão. A melhor compreensão desta abordagem e sua associação aos métodos aqui estudados, pode permitir corrigir esta deficiência de precisão e levar o limiar até 3 dB de SNR, bem como tornar mais robustos e estáveis os métodos iterativos.

Com o avanço ininterrupto da capacidade e da velocidade de processamento dos computadores, o emprego dessas técnicas fica a cada dia mais próximo de atender às crescentes exigências da solução de problemas cada vez mais sofisticados, seja nos modernos sistemas de comunicações pessoais [89], seja no controle automático de qualidade nas indústrias [33], seja na monitoração ambiental [71] ou na busca de diagnósticos mais confiáveis e intervenções cirúrgicas mais precisas [26]. Seja qual for a aplicação, a utilização da teoria aqui estudada promete melhoria significativa de desempenho em relação ao que hoje encontramos.

# Apêndice A

## Diferenciação do Lagrangiano

No Capítulo 9 utilizamos Multiplicadores de Lagrange para tratar as restrições de otimização do STLS. Porém, no sentido de nos mantermos concentrados na otimização do método STLS, deixamos a demonstração da diferenciação do Lagrangiano para este apêndice. O Lagrangiano,  $\mathcal{L}(\tilde{\mathbf{X}}, \mathbf{w}, \mathbf{l}, \lambda)$ , é uma função escalar de variáveis que são vetores complexos e, portanto, sua diferenciação exige atenção especial a dois conceitos matemáticos distintos: a natureza vetorial das variáveis; e o fato de os elementos dos vetores serem complexos. Dessa maneira, antes de derivarmos o Lagrangiano, precisamos introduzir algumas definições básicas.

### A.1 Definições Básicas

Apresentamos as definições em dois blocos, conforme são os conceitos matemáticos envolvidos na diferenciação do Lagrangiano. Inicialmente definimos a diferenciação de uma função real de variáveis complexas e, em seguida, estendemos a abordagem para o caso de diferenciação com respeito a um vetor.

#### A.1.1 Derivada Complexa de uma Função Escalar

O fato de  $\mathcal{L}(\tilde{\mathbf{X}}, \mathbf{w}, \mathbf{l}, \lambda)$  possuir parcelas que são funções escalares reais de variáveis complexas, faz do Lagrangiano uma função não-analítica e que, normalmente, não poderia ser diferenciada. Entretanto, em Haykin [32] e Kay [42], esta complicação é resolvida pela introdução de uma definição para a derivada complexa de funções não-analíticas, conforme demonstrado primeiramente por Schwartz em 1967 [90].

A derivada de uma função  $f(\theta_k)$  escalar com respeito a um parâmetro complexo  $\theta_k$  é definida como:

$$\frac{\partial f(\theta_k)}{\partial \theta_k} = \frac{1}{2} \left( \frac{\partial f(\theta_k)}{\partial \theta_{k \text{ Re}}} - j \frac{\partial f(\theta_k)}{\partial \theta_{k \text{ Im}}} \right) \quad (\text{A.1})$$

Esta definição satisfaz duas exigências básicas:

$$\begin{aligned}
\frac{\partial \theta_k}{\partial \theta_k} &= \frac{1}{2} \left( \frac{\partial \theta_k}{\partial \theta_{k \text{ Re}}} - j \frac{\partial \theta_k}{\partial \theta_{k \text{ Im}}} \right) \\
&= \frac{1}{2} \left( \frac{\partial (\theta_{k \text{ Re}} + j \theta_{k \text{ Im}})}{\partial \theta_{k \text{ Re}}} - j \frac{\partial (\theta_{k \text{ Re}} + j \theta_{k \text{ Im}})}{\partial \theta_{k \text{ Im}}} \right) \\
&= \frac{1}{2} (1 - j^2)
\end{aligned} \tag{A.2}$$

logo:

$$\frac{\partial \theta_k}{\partial \theta_k} = 1 \tag{A.3}$$

e

$$\begin{aligned}
\frac{\partial \theta_k^*}{\partial \theta_k} = \frac{\partial \theta_k}{\partial \theta_k^*} &= \frac{1}{2} \left( \frac{\partial \theta_k^*}{\partial \theta_{k \text{ Re}}} - j \frac{\partial \theta_k^*}{\partial \theta_{k \text{ Im}}} \right) \\
&= \frac{1}{2} \left( \frac{\partial (\theta_{k \text{ Re}} - j \theta_{k \text{ Im}})}{\partial \theta_{k \text{ Re}}} - j \frac{\partial (\theta_{k \text{ Re}} - j \theta_{k \text{ Im}})}{\partial \theta_{k \text{ Im}}} \right) \\
&= \frac{1}{2} (1 + j^2)
\end{aligned} \tag{A.4}$$

logo:

$$\frac{\partial \theta_k^*}{\partial \theta_k} = 0 \tag{A.5}$$

Com esta definição podemos, por exemplo, derivar a função real  $f(\theta_k) = |\theta_k|^2 = \theta_k^* \theta_k$  fazendo:

$$\begin{aligned}
\frac{\partial (\theta_k^* \theta_k)}{\partial \theta_k} &= \frac{1}{2} \left( \frac{\partial (\theta_k^* \theta_k)}{\partial \theta_{k \text{ Re}}} - j \frac{\partial (\theta_k^* \theta_k)}{\partial \theta_{k \text{ Im}}} \right) \\
&= \frac{1}{2} \left( \frac{\partial (\theta_{k \text{ Re}}^2 + \theta_{k \text{ Im}}^2)}{\partial \theta_{k \text{ Re}}} - j \frac{\partial (\theta_{k \text{ Re}}^2 + \theta_{k \text{ Im}}^2)}{\partial \theta_{k \text{ Im}}} \right) \\
&= \frac{1}{2} (2\theta_{k \text{ Re}} - j 2\theta_{k \text{ Im}}) \\
&= \theta_{k \text{ Re}} - j \theta_{k \text{ Im}}
\end{aligned} \tag{A.6}$$

logo:

$$\frac{\partial (\theta_k^* \theta_k)}{\partial \theta_k} = \theta_k^* \tag{A.7}$$

A consistência de (A.7) pode ser verificada se reescrevermos  $|\theta_k|^2$  em função de duas variáveis, como  $f(\theta_k, \theta_k^*) = \theta_k^* \theta_k$ , onde  $\theta_k$  e  $\theta_k^*$  são consideradas independentes. Desse modo, pela regra da derivada de um produto, fazemos:

$$\frac{\partial(\theta_k^* \theta_k)}{\partial \theta_k} = \frac{\partial \theta_k^*}{\partial \theta_k} \theta_k + \theta_k^* \frac{\partial \theta_k}{\partial \theta_k} \quad (\text{A.8})$$

Utilizando as definições (A.3) e (A.5) na equação anterior, encontramos:

$$\begin{aligned} \frac{\partial(\theta_k^* \theta_k)}{\partial \theta_k} &= 0 \cdot \theta_k + \theta_k^* \cdot 1 \\ &= \theta_k^* \end{aligned} \quad (\text{A.9})$$

portanto, consistente com (A.7).

### A.1.2 Extensão para o Caso Vetorial

Estendemos as definições anteriores para o caso vetorial escrevendo a derivada de uma função escalar com respeito a um vetor complexo  $\boldsymbol{\theta} = [\theta_1 \ \dots \ \theta_L]^T$  como:

$$\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial f(\boldsymbol{\theta})}{\partial \theta_1} \\ \vdots \\ \frac{\partial f(\boldsymbol{\theta})}{\partial \theta_L} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \left( \frac{\partial f(\boldsymbol{\theta})}{\partial \theta_{1 \text{ Re}}} - j \frac{\partial f(\boldsymbol{\theta})}{\partial \theta_{1 \text{ Im}}} \right) \\ \vdots \\ \frac{1}{2} \left( \frac{\partial f(\boldsymbol{\theta})}{\partial \theta_{L \text{ Re}}} - j \frac{\partial f(\boldsymbol{\theta})}{\partial \theta_{L \text{ Im}}} \right) \end{bmatrix} \quad (\text{A.10})$$

Para diferenciar o Lagrangiano precisaremos diferenciar formas lineares como  $\boldsymbol{\theta}^H \mathbf{b}$  e Hermitianas como  $\boldsymbol{\theta}^H \mathbf{A} \boldsymbol{\theta}$ , onde  $\mathbf{A}$  é Hermitiana. Assim, primeiramente, calcularemos as derivadas complexas dessas funções. Fazendo  $f(\boldsymbol{\theta}) = \boldsymbol{\theta}^H \mathbf{b}$ , temos:

$$\frac{\partial(\boldsymbol{\theta}^H \mathbf{b})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial \boldsymbol{\theta}^H \mathbf{b}}{\partial \theta_1} \\ \vdots \\ \frac{\partial \boldsymbol{\theta}^H \mathbf{b}}{\partial \theta_L} \end{bmatrix} \quad (\text{A.11})$$

Mas podemos escrever  $f(\boldsymbol{\theta})$  como:

$$f(\boldsymbol{\theta}) = \boldsymbol{\theta}^H \mathbf{b} = \sum_{i=1}^L \theta_i^* b_i \quad (\text{A.12})$$

daí, pelas propriedades de linearidade das derivadas, temos:

$$\begin{aligned} \frac{\partial(\boldsymbol{\theta}^H \mathbf{b})}{\partial \theta_k} &= \sum_{i=1}^L \frac{\partial \theta_i^*}{\partial \theta_k} b_i \\ &= \frac{\partial \theta_k^*}{\partial \theta_k} b_k \end{aligned} \quad (\text{A.13})$$

que, utilizando (A.5), fica:

$$\begin{aligned}\frac{\partial (\boldsymbol{\theta}^H \mathbf{b})}{\partial \theta_k} &= 0 \cdot b_k \\ &= 0\end{aligned}\tag{A.14}$$

Assim, temos finalmente:

$$\frac{\partial (\boldsymbol{\theta}^H \mathbf{b})}{\partial \boldsymbol{\theta}} = \mathbf{0}\tag{A.15}$$

Para  $f(\boldsymbol{\theta}) = \mathbf{b}^H \boldsymbol{\theta}$  temos:

$$\begin{aligned}\frac{\partial (\mathbf{b}^H \boldsymbol{\theta})}{\partial \theta_k} &= \sum_{i=1}^L b_i^* \frac{\partial \theta_i}{\partial \theta_k} \\ &= b_k^* \frac{\partial \theta_k}{\partial \theta_k}\end{aligned}\tag{A.16}$$

que, utilizando (A.3), fica expressa por:

$$\begin{aligned}\frac{\partial (\mathbf{b}^H \boldsymbol{\theta})}{\partial \theta_k} &= b_k^* \cdot 1 \\ &= b_k^*\end{aligned}\tag{A.17}$$

Então, em termos vetoriais:

$$\frac{\partial (\mathbf{b}^H \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{b}^*\tag{A.18}$$

Tomando agora  $f(\boldsymbol{\theta}) = \boldsymbol{\theta}^H \mathbf{A} \boldsymbol{\theta}$ , onde  $\mathbf{A}$  é Hermitiana, temos:

$$\begin{aligned}\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \frac{\partial (\boldsymbol{\theta}^H \mathbf{A} \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ &= \frac{\partial (\boldsymbol{\theta}^H \mathbf{A}^H \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ &= \frac{\partial [(\mathbf{A} \boldsymbol{\theta})^H \boldsymbol{\theta}]}{\partial \boldsymbol{\theta}}\end{aligned}\tag{A.19}$$

que, utilizando (A.18), fica expressa por:

$$\frac{\partial (\boldsymbol{\theta}^H \mathbf{A} \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = (\mathbf{A} \boldsymbol{\theta})^*\tag{A.20}$$

## A.2 Diferenciando o Lagrangiano do TLS

Da mesma forma que no Capítulo 9, primeiramente derivaremos o Lagrangiano do TLS, considerando que a matriz  $\tilde{\mathbf{X}}$  não possui estrutura. Na seção seguinte, trataremos do Lagrangiano do método STLS. Então, recordamos a expressão de  $\mathcal{L}(\tilde{\mathbf{X}}, \mathbf{w}, \mathbf{1}, \lambda)$  do TLS não-estruturado:

$$\mathcal{L}(\tilde{\mathbf{X}}, \mathbf{w}, \mathbf{1}, \lambda) = \sum_{i=1}^{(N-L)} \sum_{k=1}^{(L+1)} (x_{ik} - \tilde{x}_{ik})^* (x_{ik} - \tilde{x}_{ik}) + \mathbf{1}^H \tilde{\mathbf{X}} \mathbf{w} + \lambda(1 - \mathbf{w}^H \mathbf{w}) \quad (\text{A.21})$$

Dadas as definições anteriores, podemos agora minimizá-lo:

- Derivando em relação a  $\lambda$ , temos:

$$\frac{\partial \mathcal{L}(\tilde{\mathbf{X}}, \mathbf{w}, \mathbf{1}, \lambda)}{\partial \lambda} = 0 \quad \Rightarrow \quad 1 - \mathbf{w}^H \mathbf{w} = 0 \quad (\text{A.22})$$

logo:

$$\mathbf{w}^H \mathbf{w} = 1 \quad (\text{A.23})$$

- Derivando em relação a  $\mathbf{w}$ , encontramos:

$$\frac{\partial \mathcal{L}(\tilde{\mathbf{X}}, \mathbf{w}, \mathbf{1}, \lambda)}{\partial \mathbf{w}} = 0 \quad \Rightarrow \quad \frac{\partial \mathbf{1}^H \tilde{\mathbf{X}} \mathbf{w}}{\partial \mathbf{w}} - \lambda \frac{\partial \mathbf{w}^H \mathbf{w}}{\partial \mathbf{w}} = 0$$

Reorganizando, obtemos:

$$\frac{\partial (\tilde{\mathbf{X}}^H \mathbf{1})^H \mathbf{w}}{\partial \mathbf{w}} = \lambda \frac{\partial \mathbf{w}^H \mathbf{w}}{\partial \mathbf{w}} \quad (\text{A.24})$$

que, utilizando (A.18), resulta em:

$$(\tilde{\mathbf{X}}^H \mathbf{1})^* = \lambda \mathbf{w}^* \quad (\text{A.25})$$

ou seja:

$$\tilde{\mathbf{X}}^H \mathbf{1} = \lambda^* \mathbf{w} \quad (\text{A.26})$$

- Na parcela  $\mathbf{1}^H \tilde{\mathbf{X}} \mathbf{w}$ , podemos observar que, na verdade,  $\mathcal{L}(\cdot)$  é função de  $\mathbf{1}^*$  e não de  $\mathbf{1}$ . Logo devemos derivar em relação a  $\mathbf{1}^*$ :

$$\frac{\partial \mathcal{L}(\tilde{\mathbf{X}}, \mathbf{w}, \mathbf{1}, \lambda)}{\partial \mathbf{1}^*} = \mathbf{0} \quad \Rightarrow \quad \frac{\partial \mathbf{1}^H \tilde{\mathbf{X}} \mathbf{w}}{\partial \mathbf{1}^*} = \mathbf{0} \quad (\text{A.27})$$

Porém, primeiramente devemos observar que:

$$\frac{\partial f(\theta_k)}{\partial \theta_k^*} = \frac{1}{2} \left( \frac{\partial f(\theta_k)}{\partial \theta_{k \text{ Re}}} + j \frac{\partial f(\theta_k)}{\partial \theta_{k \text{ Im}}} \right) \quad (\text{A.28})$$

e, portanto:

$$\frac{\partial \theta^H \mathbf{b}}{\partial \theta^*} = \frac{\partial \sum_{i=1}^L \theta_i^* b_i}{\partial \theta^*} = \mathbf{b} \quad (\text{A.29})$$

Então, temos:

$$\frac{\partial \mathbf{1}^H \tilde{\mathbf{X}} \mathbf{w}}{\partial \mathbf{1}^*} = \tilde{\mathbf{X}} \mathbf{w} \quad (\text{A.30})$$

Finalmente, isto nos permite concluir que:

$$\tilde{\mathbf{X}} \mathbf{w} = \mathbf{0} \quad (\text{A.31})$$

- Derivando em relação a  $\tilde{\mathbf{X}}$ :

Em verdade, não derivamos propriamente em relação à matriz  $\tilde{\mathbf{X}}$ , mas sim em relação a seus elementos  $x_{ik}$ ,  $i = 1, \dots, (N - L)$  e  $k = 1, \dots, (L + 1)$ . Porém, como anteriormente definimos a derivada de uma função escalar em relação a um vetor, podemos particionar  $\tilde{\mathbf{X}}$  de forma a derivar o Lagrangiano em relação a suas colunas, que são vetores. Dessa forma, particionando  $\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_0 \mid \dots \mid \tilde{\mathbf{x}}_L]$  e  $\mathbf{X} = [\mathbf{x}_0 \mid \dots \mid \mathbf{x}_L]$ , a derivada do Lagrangiano em relação a  $\tilde{\mathbf{X}}$  pode ser expressa por:

$$\frac{\partial \mathcal{L}(\tilde{\mathbf{X}}, \mathbf{w}, \mathbf{1}, \lambda)}{\partial \tilde{\mathbf{X}}} = \left[ \frac{\partial \mathcal{L}(\tilde{\mathbf{X}}, \mathbf{w}, \mathbf{1}, \lambda)}{\partial \tilde{\mathbf{x}}_0} \mid \dots \mid \frac{\partial \mathcal{L}(\tilde{\mathbf{X}}, \mathbf{w}, \mathbf{1}, \lambda)}{\partial \tilde{\mathbf{x}}_L} \right] = \mathbf{0} \quad (\text{A.32})$$

Assim, para  $k = 0, \dots, L$ , a derivada em relação à  $k$ -ésima coluna é calculada através de:

$$\frac{\partial \mathcal{L}(\tilde{\mathbf{X}}, \mathbf{w}, \mathbf{1}, \lambda)}{\partial \tilde{\mathbf{x}}_k} = \frac{\partial \left( \|\mathbf{X} - \tilde{\mathbf{X}}\|_F^2 + \mathbf{1}^H \tilde{\mathbf{X}} \mathbf{w} \right)}{\partial \tilde{\mathbf{x}}_k} \quad (\text{A.33})$$

Porém, temos as seguintes relações:

$$\begin{aligned} \|\mathbf{X} - \tilde{\mathbf{X}}\|_F^2 &= \text{Tr} \left[ (\mathbf{X} - \tilde{\mathbf{X}})^H (\mathbf{X} - \tilde{\mathbf{X}}) \right] \\ &= \text{Tr} \left\{ \left[ (\mathbf{x}_0 - \tilde{\mathbf{x}}_0) \mid \dots \mid (\mathbf{x}_L - \tilde{\mathbf{x}}_L) \right]^H \left[ (\mathbf{x}_0 - \tilde{\mathbf{x}}_0) \mid \dots \mid (\mathbf{x}_L - \tilde{\mathbf{x}}_L) \right] \right\} \end{aligned}$$

$$\begin{aligned}
&= \text{Tr} \left\{ \begin{bmatrix} (\mathbf{x}_0 - \bar{\mathbf{x}}_0)^H (\mathbf{x}_0 - \bar{\mathbf{x}}_0) & \cdots & (\mathbf{x}_0 - \bar{\mathbf{x}}_0)^H (\mathbf{x}_L - \bar{\mathbf{x}}_L) \\ \vdots & \ddots & \vdots \\ (\mathbf{x}_L - \bar{\mathbf{x}}_L)^H (\mathbf{x}_0 - \bar{\mathbf{x}}_0) & \cdots & (\mathbf{x}_L - \bar{\mathbf{x}}_L)^H (\mathbf{x}_L - \bar{\mathbf{x}}_L) \end{bmatrix} \right\} \\
&= \sum_{i=0}^L (\mathbf{x}_i - \bar{\mathbf{x}}_i)^H (\mathbf{x}_i - \bar{\mathbf{x}}_i)
\end{aligned} \tag{A.34}$$

e também:

$$\begin{aligned}
\mathbf{I}^H \tilde{\mathbf{X}} \mathbf{w} &= \mathbf{I}^H \left[ \bar{\mathbf{x}}_0 \mid \cdots \mid \bar{\mathbf{x}}_L \right] \mathbf{w} \\
&= \sum_{i=0}^L \mathbf{I}^H \tilde{\mathbf{x}}_i w_i
\end{aligned} \tag{A.35}$$

Estas relações permitem escrever (A.33) como:

$$\frac{\partial \mathcal{L}(\tilde{\mathbf{X}}, \mathbf{w}, \mathbf{l}, \lambda)}{\partial \bar{\mathbf{x}}_k} = \frac{\partial \left\{ \sum_{i=0}^L \left[ (\mathbf{x}_i - \bar{\mathbf{x}}_i)^H (\mathbf{x}_i - \bar{\mathbf{x}}_i) + \mathbf{I}^H \tilde{\mathbf{x}}_i w_i \right] \right\}}{\partial \bar{\mathbf{x}}_k} \tag{A.36}$$

ou seja:

$$\frac{\partial \mathcal{L}(\tilde{\mathbf{X}}, \mathbf{w}, \mathbf{l}, \lambda)}{\partial \bar{\mathbf{x}}_k} = \frac{\partial \left[ (\mathbf{x}_k - \bar{\mathbf{x}}_k)^H (\mathbf{x}_k - \bar{\mathbf{x}}_k) + \mathbf{I}^H \tilde{\mathbf{x}}_k w_k \right]}{\partial \bar{\mathbf{x}}_k} \tag{A.37}$$

que, utilizando a regra da cadeia, pode ser escrita como:

$$\frac{\partial \mathcal{L}(\tilde{\mathbf{X}}, \mathbf{w}, \mathbf{l}, \lambda)}{\partial \bar{\mathbf{x}}_k} = -(\mathbf{x}_k - \bar{\mathbf{x}}_k)^* + \mathbf{I}^* w_k \tag{A.38}$$

Substituindo (A.38) em (A.32), encontramos:

$$\frac{\partial \mathcal{L}(\tilde{\mathbf{X}}, \mathbf{w}, \mathbf{l}, \lambda)}{\partial \tilde{\mathbf{X}}} = -(\mathbf{X} - \tilde{\mathbf{X}})^* + \mathbf{I}^* \mathbf{w}^T = \mathbf{0} \tag{A.39}$$

e, por fim, temos:

$$(\mathbf{X} - \tilde{\mathbf{X}}) = \mathbf{l} \mathbf{w}^H \tag{A.40}$$

## A.3 Diferenciando o Lagrangiano do STLS

Recordamos agora o Lagrangiano do método STLS:

$$\mathcal{L}(\tilde{\mathbf{x}}, \mathbf{w}, \mathbf{l}, \lambda) = \sum_{i=0}^{N-1} (x_i - \tilde{x}_i)^* (x_i - \tilde{x}_i) + \mathbf{I}^H \tilde{\mathbf{X}} \mathbf{w} + \lambda (1 - \mathbf{w}^H \mathbf{w}) \tag{A.41}$$

Devemos observar que o cálculo das derivadas de  $\mathcal{L}(\bar{\mathbf{x}}, \mathbf{w}, \mathbf{l}, \lambda)$  com respeito a  $\mathbf{w}$ ,  $\mathbf{l}$  e  $\lambda$ , é o mesmo apresentado para o TLS não-estruturado. Apenas a diferenciação com respeito a  $\bar{\mathbf{x}}$  é que difere da abordagem anterior e, por isso mesmo, será desenvolvida nesta seção. Desta maneira, a derivada do Lagrangiano em relação ao vetor  $\bar{\mathbf{x}}$ ,  $N \times 1$ , é dada por:

$$\frac{\partial \mathcal{L}(\bar{\mathbf{x}}, \mathbf{w}, \mathbf{l}, \lambda)}{\partial \bar{\mathbf{x}}} = \begin{bmatrix} \frac{\partial \mathcal{L}(\bar{\mathbf{x}}, \mathbf{w}, \mathbf{l}, \lambda)}{\partial \tilde{x}_0} \\ \vdots \\ \frac{\partial \mathcal{L}(\bar{\mathbf{x}}, \mathbf{w}, \mathbf{l}, \lambda)}{\partial \tilde{x}_{(N-1)}} \end{bmatrix} = \mathbf{0} \quad (\text{A.42})$$

Assim, para  $k = 0, \dots, (N-1)$ , a derivada em relação ao  $k$ -ésimo elemento é calculada através de:

$$\frac{\partial \mathcal{L}(\bar{\mathbf{x}}, \mathbf{w}, \mathbf{l}, \lambda)}{\partial \tilde{x}_k} = \frac{\partial \left[ \sum_{i=0}^{N-1} (x_i - \tilde{x}_i)^* (x_i - \tilde{x}_i) + \mathbf{l}^H \tilde{\mathbf{X}} \mathbf{w} \right]}{\partial \tilde{x}_k} = 0 \quad (\text{A.43})$$

Porém, lembrando que  $\tilde{\mathbf{X}} = \sum_{i=0}^{N-1} \tilde{x}_i \mathbf{X}_i + \mathbf{X}_N$ , podemos escrever  $\mathbf{l}^H \tilde{\mathbf{X}} \mathbf{w}$  como:

$$\mathbf{l}^H \tilde{\mathbf{X}} \mathbf{w} = \mathbf{l}^H \left( \sum_{i=0}^{N-1} \tilde{x}_i \mathbf{X}_i + \mathbf{X}_N \right) \mathbf{w} \quad (\text{A.44})$$

$$= \sum_{i=0}^{N-1} \tilde{x}_i \mathbf{l}^H \mathbf{X}_i \mathbf{w} + \mathbf{l}^H \mathbf{X}_N \mathbf{w} \quad (\text{A.45})$$

que, substituída em (A.43), permite-nos escrever:

$$\frac{\partial \mathcal{L}(\bar{\mathbf{x}}, \mathbf{w}, \mathbf{l}, \lambda)}{\partial \tilde{x}_k} = \frac{\partial \left\{ \sum_{i=0}^{N-1} \left[ (x_i - \tilde{x}_i)^* (x_i - \tilde{x}_i) + \tilde{x}_i \mathbf{l}^H \mathbf{X}_i \mathbf{w} \right] + \mathbf{l}^H \mathbf{X}_N \mathbf{w} \right\}}{\partial \tilde{x}_k} = 0 \quad (\text{A.46})$$

ou seja:

$$\frac{\partial \left[ (x_k - \tilde{x}_k)^* (x_k - \tilde{x}_k) + \tilde{x}_k \mathbf{l}^H \mathbf{X}_k \mathbf{w} \right]}{\partial \tilde{x}_k} = 0 \quad (\text{A.47})$$

e, portanto:

$$-(x_k - \tilde{x}_k)^* + \mathbf{l}^H \mathbf{X}_k \mathbf{w} = 0 \quad (\text{A.48})$$

Assim, para satisfazer (A.43), devemos ter:

$$(x_k - \tilde{x}_k) = \left( \mathbf{l}^H \mathbf{X}_k \mathbf{w} \right)^*, \quad k = 0, \dots, (N-1) \quad (\text{A.49})$$

■

## Apêndice B

### Quadro Resumo dos Métodos Estudados

Método	Complexidade	Resolução	Estabilidade
FBLP Modificado	Baixa	Baixa	Alta
TLS-FBLP	Baixa	Baixa	Alta
ML	Alta	Alta	Alta
IQML	Média	Alta	Média
CTLS	Média	Alta	Média
WTLS	Média	Alta	Média
STLS	Média	Alta	Média
STLS com Elementos Fixos	Média	Alta	Média
SLS	Média	Média	Baixa
CLS	Média	Média	Baixa
WLS	Média	Média	Baixa

# Apêndice C

## Glossário de Símbolos

$N$	— número de amostras
$M$	— número de exponenciais complexas
$L$	— ordem de predição
$x[n]$	— $n$ -ésima amostra do sinal
$a_i$	— amplitude da $i$ -ésima exponencial
$\omega_i$	— frequência angular da $i$ -ésima exponencial
$\phi_i$	— fase da $i$ -ésima exponencial
$y[n]$	— $n$ -ésima amostra do ruído
$\mathbf{x}$	— vetor de amostras do sinal
$\mathbf{y}$	— vetor de amostras do ruído
$\hat{\mathbf{R}}_{yy}$	— matriz de autocorrelação do ruído
$\sigma_y^2$	— variância do ruído
$\mathbf{E}$	— matriz de Vandermonde sobre $\exp(j\omega_i n)$
$\mathbf{a}_c$	— vetor de amplitudes complexas
$()^T$	— transposto de
$()^*$	— conjugado complexo de
$()^H$	— transposto conjugado (hermitiano) de
$()^{-1}$	— inversa de
$()^\#$	— pseudo-inversa de
$\text{Re}()$	— parte real de
$\text{Im}()$	— parte imaginária de
$\log()$	— logaritmo decimal de
$\ln()$	— logaritmo natural de
$\mathbf{J}$	— matriz de informação de Fischer
$E\{\}$	— esperança matemática ou média de

$var \{ \}$	— esperança matemática ou média de
SNR	— relação sinal-ruído
$f_L[n]$	— erro de predição <i>forward</i> de ordem $L$
$b_L[n]$	— erro de predição <i>backward</i> de ordem $L$
$( )_f$	— índice relativo a predição <i>forward</i>
$( )_b$	— índice relativo a predição <i>backward</i>
$\mathbf{A}$	— matriz de dados da predição
$\mathbf{b}$	— vetor desejado da predição
$\mathbf{w}$	— vetor de coeficientes do filtro de predição
$\mathbf{w}'$	— vetor de coeficientes do filtro de erro de predição
$\varepsilon \{ \}$	— energia do erro de predição
$\mathbf{R}$	— matriz de autocorrelação
$\mathbf{r}$	— vetor de correlação
$W(z)$	— função de transferência do filtro de erro de predição
$\mathbb{C}$	— conjunto dos números complexos
$\mathbb{R}$	— conjunto dos números reais
$\mathcal{D}()$	— conjunto domínio de, espaço das linhas de
$\mathcal{I}()$	— conjunto imagem de, espaço das colunas de
$\mathcal{N}()$	— espaço nulo de
SVD	— decomposição em valores e vetores singulares
$\mathbf{U}_Z$	— matriz de vetores singulares à esquerda de $\mathbf{Z}$
$\mathbf{V}_Z$	— matriz de vetores singulares à direita de $\mathbf{Z}$
$\mathbf{\Sigma}_Z$	— matriz diagonal contendo valores singulares de $\mathbf{Z}$
$\mathbf{u}_{zi}$	— $i$ -ésimo vetor singular à esquerda de $\mathbf{Z}$
$\mathbf{v}_{zi}$	— $i$ -ésimo vetor singular à direita de $\mathbf{Z}$
$\sigma_{zi}$	— $i$ -ésimo valor singular de $\mathbf{Z}$
$\mathbf{A}_E$	— matriz de dados do caso sem ruído
$\mathbf{Y}$	— matriz de ruído
$\mathbf{C}$	— matriz diagonal contendo cossenos dos ângulos canônicos entre os subespaços de sinal e de ruído
$\mathbf{A}_{LSr}$	— aproximação de posto reduzido para a matriz de dados
$\mathbf{w}'_{LS}$	— vetor $\mathbf{w}'$ otimizado por mínimos quadrados convencional
$\mathbf{w}'_{LSr}$	— vetor $\mathbf{w}'$ otimizado por mínimos quadrados convencional com redução de posto.
$\mathbf{w}'_{Mod.}$	— vetor $\mathbf{w}'$ otimizado por mínimos quadrados convencional com restrição ao subespaço de sinal.
$\mathbf{w}'_{Mod.}$	— vetor $\mathbf{w}'$ otimizado por mínimos quadrados totais

$\Delta A$	— perturbações na matriz de dados
$\Delta b$	— perturbações no vetor desejado
$\tilde{A}$	— aproximação da matriz de dados: $\tilde{A} = A + \Delta A$
$X$	— matriz de dados estendida
$\Delta X$	— perturbações na matriz de dados estendida
$\tilde{X}$	— aproximação da matriz de dados estendida: $\tilde{X} = X + \Delta X$
$\  \cdot \ , \  \cdot \ _2$	— norma euclidiana
$\  \cdot \ _F$	— norma de Frobenius
$H$	— matriz de transformação de Householder
$W^H$	— matriz de coeficientes do filtro de predição
$\bar{I}$	— matriz de reflexão lateral
$T^H$	— matriz complementar de $W^H$ para forma circulante
$\Theta^H$	— matriz circulante de coeficientes do filtro de predição
$\Lambda$	— matriz diagonal de contendo autovalores de $\Theta^H$
$Q$	— matriz circulante $Q = \Theta^H \Theta$
$\hat{Q}$	— versão definida positiva e bem condicionada de $Q$
$\tilde{Q}$	— versão bem condicionada de $W^H W$
$\Delta x$	— perturbações no vetor de amostras $x$
$F_i$	— matriz de estruturação de cada coluna de $\Delta X$
$X_i$	— matriz de estruturação de $\tilde{X}$ que mapeia $x_i$ em $X$
$e$	— vetor erro de predição
$e_w$	— vetor erro de predição branqueado
$\tilde{x}$	— aproximação do vetor $x$ : $\tilde{x} = x + \Delta x$
$\mathcal{L}()$	— Lagrangiano em função de
$l$	— vetor de multiplicadores de Lagrange
$\lambda$	— escalar multiplicador de Lagrange
$\text{Tr}()$	— traço de
$D_w$	— matriz de ponderação em $w$
$X_{\mathcal{F}}$	— parcela fixa de $X$
$X_{\mathcal{M}}$	— parcela modificável de $X$
$\tilde{X}_{\mathcal{M}}$	— parcela modificável de $\tilde{X}$
$X_{i\mathcal{F}}$	— parcela fixa de $X_i$
$X_{i\mathcal{M}}$	— parcela modificável de $X_i$
$D_{w\mathcal{M}}$	— matriz de ponderação em $w$ para $X_{i\mathcal{M}}$
$M$	— matriz de máscara sobre $X_i$
$b_i, A_i$	— partições de $X_i$ para $b$ e $A$

# Bibliografia

- [1] Abatzoglou, T.J. e Mendel, J.M., 'Constrained total least squares', *Proceedings of the IEEE ICASSP'87*, pp. 1485-1488, Dallas, 1987.
- [2] Abatzoglou, T.J., Mendel, J.M. e Harada, G., 'The constrained total least squares technique and its applications to harmonic superresolution', *IEEE Trans. on Signal Processing*, v. TSP-39, no. 5, pp. 1070-1086, May 1991.
- [3] Antunes, E.J.B., *Detecção de Frequências através de Predição Linear*, Tese de Mestrado apresentada à Faculdade de Engenharia Elétrica da UNICAMP, Universidade Estadual de Campinas, Campinas-SP, maio de 1992.
- [4] Antunes, E.J.B. e Lopes, A., 'Frequencies estimation of real sinusoids', *Proceedings of the International Symposium on Signals, Systems and Electronics (ISSSE'92)*, pp. 520-523, Paris, 1992.
- [5] Behrens, R.T. e Scharf, L.L., 'Signal processing applications of oblique projection operators', *IEEE Trans. on Signal Processing*, v. TSP-42, no. 5, pp. 1413-24, June 1994.
- [6] Bresler, Y. e Macovski, A., 'Exact maximum likelihood parameter estimation of superimposed exponential signals in noise', *IEEE Trans. on ASSP*, v. ASSP-34, no. 5, pp. 1081-1089, October 1986.
- [7] Cadzow, J.A., Baseghi, B. e Hsu, T., 'Singular-value decomposition approach to time series modeling', *IEE Proceedings*, v. 130, pt. F, pp. 202-210, April 1983.
- [8] Cadzow, J.A., 'Signal enhancement - A composite property mapping algorithm', *IEEE Trans. on ASSP*, v. ASSP-36, no. 1, pp. 49-62, January 1988.
- [9] Chen, H., Van Huffel, S. e Vandewalle, J., 'Structured total least norm and its application to parameter estimation of exponentially damped sinusoids', *Proceedings of the PRO-RISC/ IEEE workshop*, Mierlo, 1995.
- [10] Chen, H., Van Huffel, S. e Van Ormondt, D., 'Application of the structured total least norm technique in spectral estimation', *Proceedings of the 8th. European Signal Processing Conference (EUSIPCO'96)*, pp. 706-709, Trieste, 1996.
- [11] Chen, H., Van Huffel, S., Dowling, E.M. e DeGroat, R.D., 'TLS based methods for exponential parameter estimation', *Recent Advances in Total Least Squares and Errors-in-Variables Modeling*, (Van Huffel S., ed.), pp. 295-305, Philadelphia, PA: SIAM, 1997.

- [12] Clark, M.P. e Scharf, L.L., 'On the complexity of IQML algorithms', *IEEE Trans. on Signal Processing*, v. TSP-40, no.7, pp. 1811-1813, July 1992.
- [13] Davila, C.E. e Azmoodeh, M., 'Efficient estimation of the signal subspace without eigendecomposition', *IEEE Trans. on Signal Processing*, vol. TSP-42, no. 1, pp. 236-239, January 1994.
- [14] De Moor, B. e Vandewalle, J., 'A unifying theorem for linear and total linear least squares', *IEEE Trans. on Automatic Control*, vol. TAC-35, no. 5, pp. 563-566, May 1990.
- [15] De Moor, B. e Golub, G.H., 'The restricted singular value decomposition: properties and applications', *SIAM J. Matrix Analysis Appl.*, v. 12, no. 3, pp. 401-425, July 1991.
- [16] De Moor, B., 'Generalizations of the singular value and qr decomposition', *SIAM J. on Matrix Analysis and Appl.*, v. 13, no. 4, pp. 993-1014, October 1992.
- [17] De Moor, B., 'On the structure of generalized singular value and qr decompositions', *SIAM J. on Matrix Analysis and Appl.*, 1993.
- [18] De Moor, B., 'Structured total least squares and  $L_2$  approximation problems', Special issue of *Linear Algebra and Its Applications*, on Numerical Linear Algebra Methods in Control, Signals and Systems, v. 188-189, pp. 163-207, July 1993.
- [19] De Moor, B., 'The singular value decomposition and long and short spaces of noisy matrices', *IEEE Trans. on Signal Processing*, v. TSP-41, no. 9, pp. 2826-2838, September 1993.
- [20] De Moor, B., 'Total least squares for affinely structured matrices and the noisy realization problem', *IEEE Trans. on Signal Processing*, v. TSP-42, no. 11, pp. 3104-3113, November 1994.
- [21] DeGroat, R.D., Dowling, E.M. e Linebarger, D.A., 'The constrained MUSIC problem', *IEEE Trans. on Signal Processing*, v. TSP-41, no. 3, pp. 1445-1449, March 1993.
- [22] Dologlou, I., Van Huffel, S. e Van Ormondt, D., 'Improved signal enhancement procedures applied to exponential data modeling', *IEEE Trans. on Signal Processing*, v. TSP-45, no. 3, pp. 799-803, March 1997.
- [23] Dowling, E.M. e DeGroat, R.D., 'The equivalence of the total least squares and minimum norm methods', *IEEE Trans. on Signal Processing*, v. TSP-39, no. 8, pp. 1891-1892, August 1991.
- [24] Dowling, E.M., DeGroat, R.D., Linebarger, D.A., Scharf, L.L. e Vis, M., 'Reduced polynomial order linear prediction', *IEEE Signal Processing Letters*, v. 3, no. 3, pp. 92-94, March 1996.
- [25] Evans, A.G. e Fischl, R., 'Optimal least squares time-domain synthesis of recursive digital filters', *IEEE Trans. on Audio and Electroacoustics*, v. AU-21, no. 1, pp. 61-65, February 1973.

- [26] Fenn, A.J. e King, G.A., 'Adaptive radio-frequency hyperthermia-phased array system for improved cancer therapy: phantom target measurements', *International Journal of Hyperthermia*, v. 10, no. 2, pp. 189-208, March-April 1994.
- [27] Fierro, R.D., Golub, G.H., Hansen, P.C. e O'Leary, D.P., 'Regularization by truncated total least squares', 1996.
- [28] Golub, G.H. e Van Loan, C.F., 'An analysis of the total least squares problem', *SIAM J. on Numerical Analysis*, v. 17, no. 6, December 1980.
- [29] Golub, G.H., Hoffman, A. e Stewart, G.W., 'A generalization of the Eckart-Young-Mirsky matrix approximation problem', *Linear Algebra and Its Applications*, no. 88/89, pp. 317-327, 1987.
- [30] Golub, G.H. e Van Loan, C.F., *Matrix Computations*, second ed. Baltimore, MD: John Hopkins University Press, 1989.
- [31] Haykin, S., *Modern Filters*, Macmillan Publishing Co., 1990.
- [32] Haykin, S., *Adaptive Filter Theory*, second ed., Englewood Cliffs, NJ: Prentice-Hall, 1991.
- [33] Heck, L.P., 'Signal processing research in automatic tool wear monitoring', *Proceedings of the IEEE ICASSP'93*, v. I, pp.55-58, Minneapolis, MN, 1993.
- [34] Hildebrand, F.B., *Introduction to Numerical Analysis*, New York, NY: MacGraw-Hill, 1956.
- [35] Hsue, J. e Yagle, A., 'Similarities and differences between one-sided and two-sided linear prediction', *IEEE Trans. on Signal Processing*, v. TSP-43, no. 1, pp. 345-349, January 1995.
- [36] Hu, Y.H., 'Notes on eigenvalue distribution of toeplitz matrices', *Proceedings of the IEEE ICASSP'88*, pp. 2480-2483, 1988.
- [37] Hua, Y. e Sarkar, T.K., 'Perturbation analysis of TK method for harmonic retrieval problems', *IEEE Trans. on ASSP*, v. ASSP-36, no. 2, pp. 228-240, February 1988.
- [38] Hua, Y. e Sarkar, T.K., 'A perturbation property of the TLS-LP method', *IEEE Trans. on ASSP*, v. ASSP-38, no. 11, pp. 2004-2005, November 1990.
- [39] Hua, Y. e Sarkar, T.K., 'On the total least squares linear prediction method for frequency estimation', *IEEE Trans. on ASSP*, v. ASSP-38, no. 12, pp. 2186-2189, December 1990.
- [40] Kay, S.M., 'Accurate frequency estimation at low signal-to-noise ratio', *IEEE Trans. on ASSP*, v. ASSP-32, no. 3, pp. 540-547, June 1984.
- [41] Kay, S.M., *Spectral Estimation: Theory and Application*, Englewood Cliffs, NJ: Prentice-Hall, 1988.
- [42] Kay, S.M., *Fundamentals of Statistical Signal Processing: Estimation Theory*, Englewood Cliffs, NJ: Prentice-Hall, 1993.

- [43] Kaylath, T., 'Norbert Wiener and the development of mathematical engineering', 1996.
- [44] Kot, A.C., Tufts, D.W. e Vaccaro, R.J., 'Analysis of linear prediction by matrix approximation', *IEEE Trans. on Signal Processing*, v. TSP-41, no. 11, pp. 3174-3177, November 1993.
- [45] Krim, H. e Viberg, M., 'Two decades of array signal processing', *IEEE Signal Processing Magazine*, v. 13, no. 4, pp. 67-94, July 1996.
- [46] Kumaresan, R. e Tufts, D.W., 'Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise', *IEEE Trans. on ASSP*, v. ASSP-30, no. 6, pp. 833-840, December 1982.
- [47] Kumaresan, R. e Tufts, D.W., 'Estimating the angles of arrival of plane waves', *IEEE Trans. on Aerospace and Electronic Systems*, v. AES-19, no. 1, pp. 134-139, January 1983.
- [48] Kumaresan, R., Tufts, D.W. e Scharf, L.L., 'A Prony method for noisy data: choosing the signal components and selecting the order in exponential signal models', *Proceedings of the IEEE*, v. 72, no. 2, pp. 230-233, February 1984.
- [49] Kumaresan, R., Scharf, L.L. e Shaw, A.K., 'An algorithm for pole-zero modeling and spectral analysis', *IEEE Trans. on ASSP*, v. ASSP-34, no. 3, pp. 637-640, June 1986.
- [50] Kung, S.Y., Arun, K.S. e Bhaskar Rao, D.V., 'State-space and singular values decomposition-based approximation methods for the harmonic retrieval problem', *J. Optical Society of America*, v. 73, no. 12, pp. 1799-1811, December 1983.
- [51] Lawson, C.L. e Hanson, R.J., *Solving Least Squares Problems*, Englewood Cliffs, NJ: Prentice-Hall, 1974.
- [52] Lemmerling, P., De Moor, B. e Van Huffel, S., 'On the equivalence of constrained total least squares and structured total least squares', *IEEE Trans. on Signal Processing*, v. TSP-44, no. 11, pp. 2908-2911, November 1996.
- [53] Lemos, R.P., *Métodos de Alta Resolução com Restrição de Subespaço para Detecção de Frequências*, Tese de Mestrado apresentada à Faculdade de Engenharia Elétrica da UNICAMP, Universidade Estadual de Campinas, Campinas-SP, março de 1995.
- [54] Lemos, R.P. e Lopes, A., 'Frequency detection by linear prediction using total least squares criterion', *Proceedings of the 38<sup>th</sup> Midwest Symposium on Circuits and Systems*, v. 2, pp. 823-826, R.J., 1995.
- [55] Lemos, R.P. e Lopes, A., 'Detecção de frequências através de predição linear utilizando o critério de mínimos quadrados totais', *Anais do 13<sup>o</sup> Simpósio Brasileiro de Telecomunicações*, v. 2, pp. 506-511, Águas de Lindóia, 1995.
- [56] Lemos, R.P. e Lopes, A., 'Detecção de frequências pela resposta de fase com alta resolução', *Anais do 13<sup>o</sup> Simpósio Brasileiro de Telecomunicações*, v. 2, pp. 512-517, Águas de Lindóia, 1995.

- [57] Lemos, R.P. e Lopes, A., 'A new method for frequency detection by linear prediction using total least squares', *Proceedings of the International Symposium on Signals, Systems and Electronics 1995*, pp. 135-138, San Francisco, E.U.A., 1995.
- [58] Lemos, R.P. e Lopes, A., 'Uma análise das soluções de norma mínima e de mínimos quadrados totais para o método FBLP modificado', *Anais do 14º Simpósio Brasileiro de Telecomunicações/ TELEMO' 96*, v. 2, pp. 832-836, Curitiba, 1996.
- [59] Lemos, R.P. e Lopes, A., 'Total least squares methods with approximate maximum likelihood performance', *Proceedings of the 3rd International Electronic Engineering Conference (Intercon'96)*, pp. 118-120, Trujillo, Peru, 1996.
- [60] Lemos, R.P. e Lopes, A., 'Details on the equivalence of the modified FBLP method and the TLS-FBLP method', *Proceedings of the 1996 Midwest Symposium on Circuits and Systems*, v. 2, pp. 685-688, Ames, E.U.A, 1996.
- [61] Lemos, R.P. e Lopes, A., 'An unifying framework to total least squares and approximate maximum likelihood', *Proceedings of the 1996 Bi-Annual International Telecommunications Symposium (ITS)*, pp. 153-157, Acapulco, México, 1996.
- [62] Lemos, R.P. e Lopes, A., 'Structured total least squares with fixed elements', aceito para publicação no *40th Midwest Symposium on Circuits and Systems*, a se realizar em Sacramento, E.U.A., de 3 a 6 de agosto de 1997.
- [63] Lemos, R.P. e Lopes, A., 'Mínimos quadrados totais estruturado com elementos fixos', submetido ao *15º Simpósio Brasileiro de Telecomunicações*, a se realizar em Recife, de 8 a 11 de setembro de 1997.
- [64] Lemos, R.P. e Lopes, A., 'Estimação de frequências por predição linear com desempenho de máxima verossimilhança', submetido à *Revista Brasileira de Telecomunicações*, 1997.
- [65] Lemos, R.P. e Lopes, A., 'Frequency estimation using a structured least squares method', a ser submetido à revista *IEEE Signal Processing Letters*, 1997.
- [66] Lemos, R.P. e Lopes, A., 'Formulation of least squares methods for structured matrices', a ser submetido à revista *SIAM Journal on Matrix Analysis and its Applications*, 1997.
- [67] Lemos, R.P. e Lopes, A., 'Equivalencia entre métodos basados en mínimos cuadrados totales de máxima similitud', submetido à revista *Información Tecnológica*, 1997.
- [68] Marple, S.L., Jr., *Digital Spectral Analysis*, Englewood Cliffs, NJ: Prentice-Hall, 1987.
- [69] Mataušek, M.R., Stanković, S.S. e Radović, D.R., 'Iterative inverse filtering approach to the estimation of frequencies of noisy sinusoids', *IEEE Trans. on ASSP*, v. ASSP-31, no. 6, pp. 1456-1463, December 1983.
- [70] McWhorter, L.T. e Scharf, L.L., 'Nonlinear maximum likelihood estimation of autoregressive time series', *IEEE Trans. on Signal Processing*, v. TSP-43, no.12, pp. 2909-2919, December 1995.

- [71] Nehorai, A., Porat, B. e Paldi, E., 'Detection and localization of vapor-emitting sources', *IEEE Trans. on Signal Processing*, v. TSP-43, pp. 243-253, January 1995.
- [72] Nuttall, A.H., 'Spectral analysis of an univariate process with bad data points via maximum entropy and linear predictive techniques', *NUSC Scientific and Engineering Studies, Spectral Estimation*, NUSC, New London, CT, March 1976.
- [73] Oppenheim, A.V. e Schaffer, R.W., *Discrete-Time Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [74] Ottersten, B., *Parametric Subspace Fitting Methods for Array Signal Processing*, PhD. dissertation, Inform. Syst. Lab., Dep. Elec. Eng., Stanford Univ., December 1989.
- [75] Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, third ed. New York, NY: MacGraw-Hill, Inc., 1991.
- [76] Park, H., Van Huffel, S. e Eldén, L., 'Fast algorithms for exponential data modeling', *Proceedings of the IEEE ICASSP'94*, Adelaide, 1994.
- [77] Park, H., Rosen, J.B. e Van Huffel, S., 'Structure preserving total least squares and its application to parameter estimation', *SVD and Signal Processing, III: Algorithms, Architectures and Applications*, M. Moonen and B. De Moor, eds., Elsevier Science Publishers B.V. (North-Holland), pp. 399-408, 1995.
- [78] Penrose, R., 'A generalized inverse for matrices', *Proceedings of the Cambridge Philosophical Society*, v. 51, pp. 406-413, 1955.
- [79] Penrose, R., 'On best approximate solutions of linear matrix equations', *Proceedings of the Cambridge Philosophical Society*, v. 52, pp. 17-19, 1956.
- [80] Prony, Baron de (Gaspard Riche), 'Essai expérimental et analytique: sur les lois de la dilatabilité des fluides et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alcool, à différentes températures', *J. de l'École Polytechnique*, Paris, vol.1, no. 2, pp. 24-76, 1795.
- [81] Rahman, M.A. e Yu, K.B., 'Total least squares approach for frequency estimation using linear prediction', *IEEE Trans. on ASSP*, vol. ASSP-35, no. 10, pp. 1440-1454, October 1987.
- [82] Rao, B.D., 'Lowering the threshold SNR of singular value decomposition based methods', *Proceedings of the IEEE ICASSP'88*, pp. 2472-2475, 1988.
- [83] Rao, S.S. e Gnanaprakasam, D.C., 'A criterion for identifying dominant singular values in the SVD based method of harmonic retrieval', *Proceedings of the IEEE ICASSP'88*, pp. 2460-2463, 1988.
- [84] Rife, D.C. e Boorstyn, R.R., 'Single tone parameter estimation from discrete-time observations', *IEEE Trans. on Information Theory*, v. IT-20, pp. 591-598, September 1974.

- [85] Rife, D.C e Boorstyn, R.R., 'Multiple tone parameter estimation from discrete-time observations', *The Bell System Tech. J.*, v. 55, no. 9, pp. 1389-1410, November 1976.
- [86] Rosen, J.B., Park, H. e Glick, J., 'Structured total least norm for nonlinear problems', *Technical Report TR 95-054*, Department of Computer Science, University of Minnesota, July 1995.
- [87] Spiegel, M.R., *Manual de Fórmulas e Tabelas Matemáticas*, Editora MacGraw-Hill do Brasil, 1973.
- [88] Stewart, G.W., *Introduction to Matrix Computations*, New York, NY: Academic Press, Inc., 1973.
- [89] Strom, E., Parkvall, S. e Ottersten, B., 'Near-far resistant propagation delay estimator for asynchronous direct-sequence code division multiple access systems', In Gunter, C.G., editor, *Mobile Communications, Advanced Systems and Components. Proceedings of the 1994 International Zürich Seminar on Digital Communications*, pp. 251-260, Zürich, Suíça, Março de 1994. Springer Verlag.
- [90] Schwartz, L., *Cours d'Analyse*, v. II, pp. 271-278, Paris: Hermann, 1967.
- [91] Tufts, D.W. e Kumaresan, R., 'Estimation of frequencies of multiple sinusoids: making linear prediction perform like maximum likelihood', *Proceedings of the IEEE*, vol. 70, no. 9, pp. 975-989, September 1982.
- [92] Tufts, D.W., Kot, A.C. e Vaccaro, R.J., 'The threshold analysis of SVD-based algorithms', *Proceedings of the IEEE ICASSP'88*, pp. 2416-2419, 1988.
- [93] Ulrych, T.J. e Clayton, R.W., 'Time series modeling and maximum entropy', *Physics Earth Planet Interior*, vol. 12, pp. 188-200, August 1976.
- [94] Van der Vorst, H.A. e Golub, G.H., '150 years old and still alive: eigenproblems', 1996.
- [95] Vanhamme, L., Van den Boogaart, A. e Van Huffel, S., 'Fast and accurate parameter estimation of noisy complex exponentials with use of prior knowledge', *Proceeding of the 8th. European Signal Processing Conference (EUSIPCO'96)*, pp. 407-410, Trieste, 1996.
- [96] Van Huffel, S. e Vandewalle, J., 'The use of total least squares techniques for identification and parameter estimation', *Proceedings of the IFAC'85*, v. 2-c, pp. 1167-1172, York, 1985.
- [97] Van Huffel, S. e Vandewalle, J., *The Total Least Squares Problem: Computational Aspects and Analysis*, Frontiers in Applied Mathematics, v. 9, Philadelphia, PA: SIAM, 1991.
- [98] Van Huffel, S., 'Enhanced resolution based on minimum variance estimation and exponential data modeling', 1992.
- [99] Van Huffel, S., Decanniere, C., Chen, H. e Van Hecke, P., 'Resolution improvement by minimum variance signal enhancement', 1993.

- [100] Van Huffel, S., Chen, H., Decanniere, C. e Van Hecke, P., 'Total least squares based algorithm for time domain NMR data fitting', 1994.
- [101] Van Huffel, S., De Moor, B. e Chen, H., 'Relationships between structured and constrained TLS, with applications to signal enhancement', *Systems and Networks: Mathematical Theory and Applications; Proceedings of the Int. Symposium MTNS-93*, Regensburg, Germany, August 2-6, 1993, v. II, Akademie Verlag, Berlin (Mathematical Research: v. 79), pp. 899-904, 1994.
- [102] Van Huffel, S., Park, H. e Rosen, J.B., 'Formulation and solution of structured total least norm problems for parameter estimation', *IEEE Trans. on Signal Processing*, v. TSP-44, no. 10, pp. 2464-2474, Oct. 1996.
- [103] Vis, M.L. e Scharf, L.L., 'A note on recursive maximum likelihood for autoregressive modeling', *IEEE Trans. on Signal Processing*, v. TSP-42, no. 10, pp. 2881-2883, October 1994.
- [104] Wigren, T. e Nehorai, A., 'Asymptotic Cramér-Rao bounds for estimation of the parameters of damped sine waves in noise', *IEEE Trans. on ASSP*, vol. ASSP-34, no. 4, April 1991.
- [105] Wilkinson, J.H., *The Algebraic Eigenvalue Problem*, Oxford, UK: Clarendon, 1965.
- [106] Xu, G. e Kailath, T., 'Fast subspace decomposition', *IEEE Trans. on Signal Processing*, v. TSP-42, no. 3, pp. 539-551, March 1994.
- [107] Yu, K.B., 'Recursive updating the eigenvalue decomposition of a covariance matrix', *IEEE Trans. on Signal Processing*, v. TSP-39, no. 5, pp. 1136-1144, May 1991.