



UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO
DEPARTAMENTO DE COMUNICAÇÕES

Descrição Unificada de Métodos de Estimação DOA em Arranjo de Sensores

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.

Autor

Ricardo Fialho Colares

Mestre em Engenharia Elétrica pela UNICAMP em 1997

Orientador

Prof. Dr. Amauri Lopes

Doutor em Engenharia Elétrica pela UNICAMP em 1982

Banca Examinadora

Prof. Dr. Amauri Lopes - FEEC/UNICAMP-Presidente

Prof. Dr. Leonardo Silva Resende - UFSC

Prof. Dr. João César Moura Mota - UFCE

Prof. Dr. João Batista T. Yabu-uti - FEEC/UNICAMP

Prof. Dr. João Marcos Travassos Romano - FEEC/UNICAMP

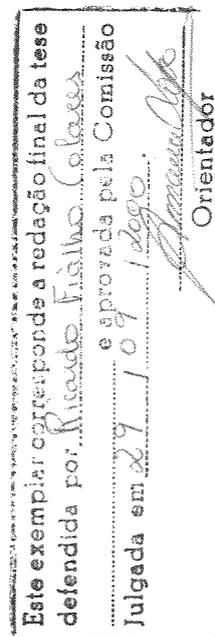
Prof. Dr. Yuzo Yano - FEEC/UNICAMP

Campinas, 29 de setembro de 2000.

UNICAMP

BIBLIOTECA CENTRAL

SEÇÃO CIRCULANTE



200107405

UNIDADE U
N.º CHAMADA: T/UNICAMP
C67d
V. Ex.
TOMBO BC/ 44339
PROC. 16-292101
C D
PREC. R\$ 11,00
DATA 30/05/01
N.º CPD

CM-00155162-9

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

C67d Colares, Ricardo Fialho
Descrição unificada de métodos de estimação DOA em
arranjo de sensores / Ricardo Fialho Colares.--Campinas,
SP: [s.n.], 2000.

Orientador: Amauri Lopes.
Tese (doutorado) - Universidade Estadual de
Campinas, Faculdade de Engenharia Elétrica e de
Computação.

1. Estimativa de parâmetro. 2. Filtros digitais
(Matemática). 3. Otimização matemática. 4. Antenas
ajustáveis. 5. Probabilidades – Processamento de dados.
I. Lopes, Amauri. II. Universidade Estadual de
Campinas. Faculdade de Engenharia Elétrica e de
Computação. III. Título.

Abstract

This work is concerned with the direction-of-arrival estimation of plane waves impinging on a sensors array. This problem arises from applications such as radar, sonar and, more recently, in wireless communications. Among the usual parameter estimation methods, the parametric ones are more accurate than the spectral ones. In this work the parametric methods are presented and classified in three categories. The first one consists of methods based on the well known maximum-likelihood (ML) criterion, such as the IQML and MODE methods. The second class is composed of methods that use the linear prediction concepts, such as the Modified FBLP an the WTLS-LP. The last class is based on the subspace approach such as the MUSIC, MINIMUM-NORM, ESPRIT and WSF methods. Those methods are presented using an unifying formulation and notation, in which case the estimation is reached by minimizing suitable and similar cost functions. The minimization process is carried out by using the least-squares (LS) and total least-squares(TLS) approaches, as well as through variants that are also presented. This unifying framework make it possible to compare the performance and computational effort of the methods presented, showing the advantages and the disadvantages of each one. In addition, we propose new methods based on the existing ones.

Sumário

Este trabalho aborda a aplicação de arranjo de sensores no problema de estimação do ângulo-de-chegada (DOA) de ondas planas. Uma vasta gama de aplicações nas mais diversas áreas se enquadram nesse problema, como em radar, sonar e mais recentemente em comunicações móveis. Dentre os métodos existentes para estimação de parâmetros, destacam-se os paramétricos em relação aos espectrais. Este trabalho classifica os principais métodos paramétricos em três categorias. A primeira consiste de métodos baseados no bastante conhecido critério da máxima-verossimilhança (ML), tais como o IQML e MODE. A segunda categoria compreende os métodos baseados na predição linear, dentre os quais podemos citar o FBLP Modificado e o WTLS-LP. A terceira consiste nos métodos que fazem uso dos subespaços definidos pela matriz de correlação dos dados disponíveis, como os métodos MUSIC, NORMA-MÍNIMA, ESPRIT e WSF. Esses métodos são apresentados neste trabalho de uma forma unificada, na qual as estimativas desejadas são obtidas através da minimização de funções custo adequadas e com mesma estrutura. Nesse processo de otimização fazemos uso do critério dos mínimos-quadrados (LS) e dos mínimos-quadrados totais (TLS), além de suas variações que também são descritas. A visão unificada proposta aqui torna possível uma comparação de desempenho e esforço computacional entre esses métodos, evidenciando suas vantagens e desvantagens. Além disso, torna também possível a sugestão de propostas para novos métodos baseados naqueles existentes.

Conteúdo

1	Introdução	1
1.1	Processamento de Arranjo	1
1.2	O Problema DOA	2
1.3	Objetivos do Trabalho	2
1.4	Organização dos Capítulos	2
2	Abordagem Geral do Problema	4
2.1	Introdução	4
2.2	Técnicas de Processamento	4
2.2.1	Técnicas Espectrais	5
2.2.2	Técnicas Paramétricas	6
2.2.3	Outras Técnicas	6
2.3	Modelo de Sinal Paramétrico para DOA	6
2.3.1	Modelo de Onda	6
2.3.2	Parâmetro DOA	7
2.4	Arranjo Linear Uniforme (U.L.A.)	8
2.4.1	Modelo do Sinal	10
2.5	Aspectos Básicos	12
2.6	Conclusão	15
3	Estratégias para Estimação do Parâmetro DOA	16
3.1	Introdução	16
3.2	Máxima-verossimilhança (ML)	17
3.2.1	Máxima-verossimilhança para Sinais Determinísticos (MLD)	17
3.2.2	Máxima-verossimilhança para Sinais Aleatórios (MLE)	18
3.2.3	ML Baseado em Subespaços	19
3.3	Predição Linear em Arranjo Linear Uniforme	20
3.3.1	Estimação de Parâmetros via Predição Linear	22
3.3.2	Predição Linear Baseada em Subespaços	23
3.4	Reparametrizando a Predição Linear	25
3.5	Reparametrizando o Estimador ML	26
3.6	Estimação Totalmente Baseada em Subespaços	26
3.6.1	O Critério de Ortogonalidade entre Subespaços	27
3.6.2	O Critério de “Distância” entre Subespaços (“Subspace Fitting”)	28
3.7	Quadro Resumo	30

6.6	Resumo Comparativo	85
6.7	Generalização dos Métodos	87
6.8	Conclusão	88
7	Conclusão	90
7.1	Resumo das Conclusões	90
7.2	Sugestões para o Prosseguimento do Trabalho	92

3.8	Conclusão	31
4	Otimização Linear Quadrática sem Restrições	33
4.1	Introdução	33
4.2	Mínimos Quadrados (LS)	34
4.3	Mínimos Quadrados Totais (TLS)	36
4.4	Problema TLS Equivalente	39
4.4.1	Solução TLS Alternativa	40
4.5	Conclusão	41
5	Adicionando Informações ao Problema de Otimização	43
5.1	Introdução	43
5.2	Estruturando os Dados	44
5.2.1	Algoritmo CTLS	45
5.2.2	Algoritmo STLS	47
5.2.3	Equivalência entre os Algoritmos CTLS e STLS	48
5.3	Branqueando o Erro de Predição	49
5.3.1	Algoritmo WTLS	50
5.4	Relação entre WTLS, CTLS e STLS	50
5.5	Incorporando Informações sobre os Subespaços	51
5.6	Generalização dos Procedimentos de Otimização	51
5.7	Conclusão	51
6	Métodos para Estimar DOA	53
6.1	Introdução	53
6.2	Métodos Baseados no Critério ML	54
6.2.1	Método IQML	54
6.2.2	Método MODE	58
6.3	Métodos Baseados na Predição Linear	64
6.3.1	Método de Prony	64
6.3.2	Métodos FLP, BLP e FBLP	65
6.3.3	Método FBLP Modificado	67
6.3.4	Método TLS-LP	67
6.3.5	Método TLS-FBLP	68
6.3.6	Método WTLS-LP	69
6.3.7	Métodos CTLS-LP e STLS-LP para DOA	71
6.4	Métodos Baseados em Subespaços	72
6.4.1	Método de Pisarenko	72
6.4.2	Método MUSIC	75
6.4.3	Método da Norma-mínima	78
6.5	Métodos Baseados na “Distância” entre Subespaços (“Subspace Fitting”)	81
6.5.1	Método ESPRIT	81
6.5.2	Método “Signal Subspace Fitting - SSF”	83
6.5.3	Método “Noise Subspace Fitting - NSF”	84
6.5.4	Método “Subspace Fitting” ponderado (WSF)	84

Capítulo 1

Introdução

1.1 Processamento de Arranjo

O processamento de arranjo de sensores trata do processamento de sinais provenientes da saída de um conjunto de sensores localizados em diferentes pontos do espaço de interesse. Os sinais incidentes no arranjo de sensores são, em geral, oriundos do fenômeno de propagação de ondas [Haykin85]. O objetivo de tal processamento é, portanto, extrair informações úteis do sinal transportado. Alguns parâmetros de interesse são a localização das fontes geradoras dos sinais, a velocidade de propagação das ondas que transmitem o sinal, e as propriedades temporais ou espectrais dos sinais (“assinatura”).

Quanto às aplicações do processamento de sinais obtidos por um arranjo de sensores podemos citar exemplos em diversas áreas de estudo, cada qual com seus objetivos e problemas específicos. Na exploração sísmológica, com o objetivo de levantar as características físicas do solo em uma determinada região. Em sonar passivo, no qual o sinal é captado por um arranjo de hidrofones, com o intuito de se determinar sua estrutura espacial e temporal. Em radar, no qual uma antena de transmissão é utilizada juntamente com um arranjo de antenas receptoras, com o objetivo de estimar o espectro de potência do sinal recebido. Nesse caso o sinal recebido é uma versão do sinal transmitido refletida pelos obstáculos ao longo do caminho de propagação. Na radioastronomia, o interesse é a rádio-emissão de fontes celestes. Nessa aplicação um conjunto de antenas é utilizado para reconstrução de imagens associadas ao espectro das fontes de rádio. Na tomografia, é utilizado para obter a imagem seccionada de objetos. Na maioria dos casos em tomografia, o objeto é “iluminado” em diversas direções e a sua imagem é reconstruída através dos dados obtidos na reflexão dos sinais. Mais recentemente, encontramos aplicação do processamento de arranjo de antenas em comunicação móveis. Em sistemas de comunicações, o objetivo de tal processamento é propiciar o cancelamento de interferências através da formação de um padrão de radiação adequado dado pelo conjunto de antenas, que privilegia a recepção dos sinais provenientes de uma determinada direção, simultaneamente anulando os sinais interferentes em outras direções. O uso de arranjo de antenas em sistemas de comunicações também possibilita combater o efeito “fading” através de técnicas de diversidade. Ainda em sistemas de comunicações, tal processamento é útil na elaboração de técnicas modernas de multiplexagem baseadas na disposição das fontes de sinais no espaço combinados com códigos (S-CDM) [Paulraj97].

Devido aos avanços nas técnicas de processamento digital de sinais, o termo proces

de arranjo de sensores foi ampliado para englobar todo e qualquer processamento aplicado em mais de um sensor, a partir dos quais obtemos versões diferentes do sinal de uma mesma fonte ou diversas fontes de sinais. As versões do sinal recebido por cada sensor são então modeladas de forma adequada ao experimento para permitir a extração dos parâmetros de interesse. Exemplos desse tipo de processamento de arranjo de sensores podem ser encontrados em aplicações biomédicas [Krim96].

1.2 O Problema DOA

O fenômeno de propagação de ondas ocorre sempre quando uma fonte geradora de distúrbio é aplicada em um meio aberto. As ondas assim geradas são denominadas de ondas viajantes. Essas ondas possuem a propriedade de transportar energia, cuja forma depende da natureza física da fonte geradora. Uma outra propriedade é o fato dessas ondas tornarem-se essencialmente planas a longas distâncias da fonte, em um meio ideal (infinito). Na verdade, as ondas planas são provavelmente a mais comum forma de propagação de ondas e são também a base para o processamento de arranjo de sensores de nosso interesse.

O problema DOA, que será o principal objeto de estudo desse trabalho, consiste na obtenção de estimativas precisas dos ângulos de incidência de ondas planas que atingem um arranjo de sensores, a partir da observação do sinal proveniente desse arranjo. Dentre as dificuldades enfrentadas na solução desse problema podemos citar a presença de ruído no sinal observado, a limitação do tempo de observação, a proximidade entre os ângulos de incidências de fontes de sinal distintas e o esforço computacional requerido para obter as estimativas finais.

1.3 Objetivos do Trabalho

Partindo do estudo de uma vasta variedade de métodos existentes para tratar o problema DOA, constatamos a semelhança entre diversas abordagens. Propomos, aqui, apresentar os principais métodos ressaltando suas diferenças essenciais e suas equivalências. Para esse fim apresentaremos inicialmente as ferramentas e procedimentos comuns a esses métodos. Através de uma notação única, será possível descrever os métodos existentes na literatura como uma composição dessas ferramentas e procedimentos [Colares2000b].

Como fruto dessa abordagem unificada, será possível uma comparação qualitativa entre os métodos existentes, além de permitir a proposta de novos métodos que potencializam suas principais vantagens.

1.4 Organização dos Capítulos

Capítulo 2

Neste capítulo, apresentaremos as técnicas mais comuns de estimação de parâmetros. Destacaremos as técnicas paramétricas e o uso da abordagem de subespaços. Apresentaremos o modelo de sinal utilizado nos capítulos seguintes, definindo com precisão o parâmetro DOA. arranjo de sensores do tipo linear e uniforme será escolhido como o tipo de arranjo empre

no problema DOA, e as equações para o sinal obtido através desse arranjo também serão apresentadas.

Alguns resultados estatísticos também serão apresentados nesse capítulo, estabelecendo dessa forma a notação utilizada no decorrer do trabalho.

Capítulo 3

Neste capítulo descreveremos três categorias básicas de procedimentos paramétricos voltados para estimação do parâmetro DOA: o critério da máxima-verossimilhança(ML), a predição linear e aqueles baseados na decomposição de subespaços.

Apresentaremos uma reparametrização do problema DOA original que será adequada para unificar esses grupos de procedimentos através de um problema de minimização com mesma estrutura.

Capítulo 4

Neste capítulo apresentaremos os procedimentos mais comuns utilizados nos problemas de minimização que surgem do problema DOA. O critério clássico dos mínimos-quadrados(LS) e sua extensão denominada de mínimos-quadrados totais(TLS) serão apresentados e confrontados. As soluções possíveis para os dois critérios também serão derivadas.

Destacaremos nesse capítulo as diversas formas de interpretação dos problemas de minimização no contexto LS e TLS.

Capítulo 5

Neste capítulo apresentaremos variações ao procedimento TLS decorrente da inserção de informações adicionais ao problema de otimização original. Essas informações serão tratadas como restrições à solução TLS clássica, dando origem aos algoritmos STLS, CTLS e WTLS.

Os algoritmos citados serão apresentados no contexto da predição linear permitindo a constatação das equivalências existentes entre os mesmos.

Capítulo 6

Neste capítulo apresentaremos os métodos paramétricos existentes para abordar o problema DOA. Os métodos serão classificados dentro das estratégias apresentadas no Capítulo 3, desde a sua concepção inicial até suas variações mais particulares. Alguns métodos de estimação de parâmetros que não foram concebidos para tratar especificamente o problema DOA, também serão descritos por apresentarem estruturas semelhantes e serem passíveis dessa aplicação.

Esse capítulo atinge, portanto, os principais objetivos desse trabalho.

Capítulo 2

Abordagem Geral do Problema

2.1 Introdução

Técnicas modernas de processamento de sinais em arranjo de sensores, referidas aqui como processamento de arranjo, baseiam-se na escolha adequada do modelo estatístico utilizado para descrever os campos de ondas, através dos quais podemos interpretar o processamento de arranjo como um problema estocástico. Tal abordagem nos permite aplicar resultados conhecidos de técnicas de estimação de espectros ou de parâmetros.

Além das técnicas empregadas no processamento, existem algumas classificações para o processamento de arranjo de sensores quanto ao modelo de sinal e tipo de arranjo. O modelo para o sinal será apresentado desde a forma mais geral até uma forma mais específica de nosso interesse, a partir da qual definiremos o parâmetro DOA. Neste modelo, as fontes dos sinais podem ser classificadas como *não-correlacionadas*, no caso de serem *independentes* umas das outras e parcialmente *correlacionadas* ou completamente *coerentes*. Quanto a fonte de radiação, esta pode ser proveniente de um meio difuso e, portanto, distribuída no espaço, ou vir de fontes isoladas com extensão angular finita [Pillai89].

O arranjo de sensores, por sua vez, pode tomar uma variedade de formas geométricas dependendo da aplicação de interesse [Haykin85]. A configuração mais utilizada é a *linear*, na qual sensores de um mesmo tipo são dispostos uniformemente espaçados ao longo de uma linha reta imaginária. Uma outra configuração comum é a *planar*, na qual os sensores formam uma grade retangular ou circunferências concêntricas.

2.2 Técnicas de Processamento

As ferramentas fundamentais para o processamento de sinais através de arranjo de sensores são a correlação e/ou a análise espectral. O uso da correlação se justifica através do conceito de filtro casado, quando dispomos do conhecimento *a priori* do sinal. A análise espectral inclui uma vasta gama de técnicas desde as mais tradicionais baseadas na transformada de Fourier até técnicas que dependem diretamente dos dados, como o caso da máxima-verossimilhança (ML).

As técnicas denominadas de conformação de feixe (“beamforming”), foram precursoras no tratamento de aplicações de arranjo de sensores [Haykin85]. Essa técnica procura ajustar os

ganhos dos sensores, no caso antenas, de forma a obter um padrão de radiação adequado para todo o arranjo. Os ganhos dos sensores são então ajustados de forma que o arranjo apresente um ganho máximo na direção da fonte do sinal desejado e ganho mínimo ou nulo na direção das fontes de sinais interferentes.

De um modo geral, as técnicas de processamento podem ser classificadas como a seguir.

2.2.1 Técnicas Espectrais

As técnicas espectrais fazem uso do espectro, ou de uma função tipo espectral através de um mapeamento adequado, em uma ou mais dimensões, para levantar as características principais dos sinais incidentes em um arranjo. Um espectro estimado pode fornecer ainda uma melhor separação das características entre o sinal e o ruído. No processamento de arranjo de sensores, o interesse se limita aos valores de um conjunto restrito de variáveis como a direção, velocidade e frequências.

Até o final dos anos 60, o método mais comum para obtenção da estimativa de um espectro era a transformada de Fourier, como por exemplo no método do Periodograma. Posteriormente surgiram novos procedimentos de análise espectral baseados em modelos, como o algoritmo de Burg [Kay88].

A análise espectral paramétrica consiste no desenvolvimento de um modelo para o processo estocástico observado, o qual deverá ser perfeitamente especificado através de um conjunto de parâmetros. Diferente dos procedimentos anteriores, que se concentravam na estimação direta da seqüência de autocorrelação do processo, os procedimentos paramétricos simplesmente estimam os parâmetros do modelo especificado. Como exemplo desses modelos estatísticos para esse fim podemos citar os modelos autorregressivos (AR), os de média-ajustável (MA) e a composição desses dois denominada de modelo autorregressivo de média-ajustável (ARMA) [Box76], [Makhoul75]. Embora com algumas limitações, os procedimentos baseados em modelos paramétricos apresentam características de alta-resolução e melhores estimativas que os procedimentos não-paramétricos, principalmente quando dispomos de um pequeno conjunto de dados. No entanto, os procedimentos não-paramétricos baseados na transformada de Fourier ainda são utilizados pela sua simplicidade de cálculo e aplicabilidade. Recentemente, alguns trabalhos apontam para o aprimoramento dos métodos de janelamento baseados na transformada de Fourier discreta [Sacchi98].

Uma das principais figuras de mérito das técnicas de processamento de arranjo é a sua capacidade de resolução. O conceito clássico e intuitivo de resolução espectral refere-se a capacidade de distinguir frequências próximas entre si. Tipicamente a resolução está relacionada às medidas espectrais que apresentam picos na posição das frequências do sinal ou próximas a essas. Quando existe um número de picos espectrais igual ao número de frequências distintas, as frequências são consideradas resolvidas.

Uma das limitações dos métodos espectrais é o fato da sua resolução ser diretamente dependente do tamanho físico do arranjo (“abertura”), independentemente da relação sinal-ruído (SNR).

Existe uma classe importante de estimadores de espectro de potência que vale a pena ser destacada. Tais técnicas de estimação são baseadas no critério em probabilidade da *máxima-verossimilhança* (ML). Esse critério apresenta uma alta capacidade de resolução e nenhuma dificuldade particular na aplicação em arranjos de qualquer geometria ou para o caso multidimensional.

mensional [Ottersten93].

Dentre os procedimentos baseados em modelos paramétricos para a análise espectral destacam-se também aqueles baseados em subespaços que utilizam propriedades geométricas do modelo de sinal, como o método MUSIC espectral. Nesses procedimentos a capacidade de resolução independe do tamanho físico do arranjo de sensores, desde que a quantidade de dados e/ou a SNR sejam suficientemente grandes e o modelo reflita precisamente o cenário real.

2.2.2 Técnicas Paramétricas

Particularmente para cenários envolvendo sinais parcialmente correlacionados ou mesmo coerentes, o desempenho das técnicas espectrais torna-se insatisfatório. Uma alternativa para superar essa limitação é explorar o modelo de sinal utilizado. Essas técnicas, denominadas aqui de paramétricas, são aquelas em que o resultado de interesse surge não mais na forma de uma função, mas através de um conjunto finito de parâmetros, mesmo que tais parâmetros sejam oriundos da parametrização de uma função espectral.

As técnicas paramétricas são, em geral, mais eficientes e robustas que as técnicas espectrais. Porém, sua implementação apresenta algoritmos mais complexos e alto esforço computacional, geralmente exigindo buscas multidimensionais para obter as estimativas finais dos parâmetros. No caso de arranjo linear com sensores uniformemente espaçados, tais buscas podem ser evitadas sem perda considerável de desempenho.

Procedimentos que fazem uso da extração das raízes de um polinômio para obter suas estimativas, bem como os métodos mais modernos baseados no critério da máxima-verossimilhança (ML) e os baseados em subespaços, fazem parte de exemplos de aplicação das técnicas paramétricas.

Nestas técnicas a noção de resolução não é trivial, sendo necessário uma abordagem baseada em estimação de parâmetros.

2.2.3 Outras Técnicas

Técnicas baseadas na estrutura do sinal, diferentemente das técnicas mais tradicionais, utilizam informações adicionais do sinal observado. A estrutura cicloestacionária do sinal, bem como o uso de estatísticas de ordens superiores são geralmente incorporadas nessas técnicas [Shamsunde94].

Quanto ao desempenho, tais técnicas geralmente exigem um grande quantidade de dados para apresentar resultados satisfatórios. Conseqüentemente, apresentam uma convergência lenta comparada com as técnicas descritas anteriormente.

Neste trabalho abordaremos apenas os métodos paramétricos, destacando os baseados no critério ML e aqueles baseados em subespaços.

2.3 Modelo de Sinal Paramétrico para DOA

2.3.1 Modelo de Onda

Considerando a condição de campo distante da fonte geradora, em um meio isotrópico infinito, e fontes pontuais, as ondas viajantes podem ser caracterizadas através dos seguintes

parâmetros: amplitude em um dado instante de tempo (t), frequência e uma posição determinada pelo vetor \mathbf{r} , que indica um ponto no espaço medido a partir da origem do sistema de coordenadas cartesianas, conforme a *Figura 2.1*. Portanto, em uma localização do plano definida pelo vetor \mathbf{r} , a função de onda possui a seguinte dependência com o tempo [Pillai89], [Haykin85]

$$E(t, \mathbf{r}) = s(t)e^{j(\omega_c t - \mathbf{v} \cdot \mathbf{r}^T)}, \quad (2.1)$$

na qual \mathbf{v} é chamado de vetor de onda na direção da propagação da onda, e sua magnitude $|\mathbf{v}| = 2\pi/\lambda_c$ é denominada de número de onda, onde λ_c é o comprimento de onda na direção de propagação.

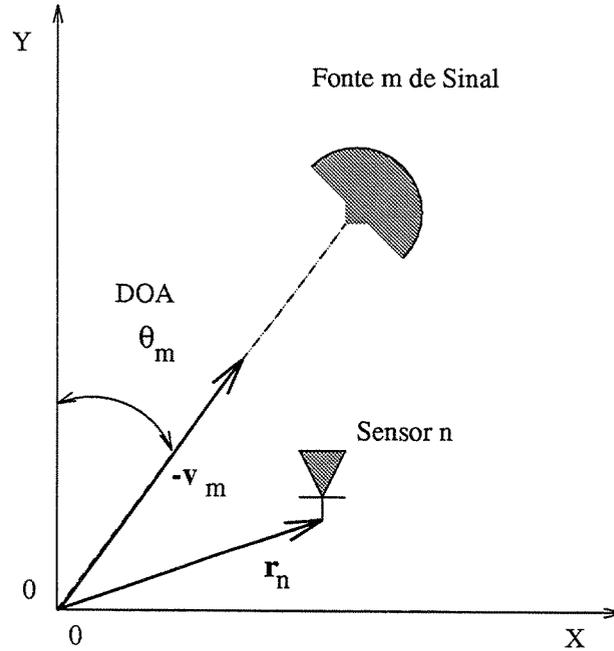


Figura 2.1: Geometria do modelo paramétrico em duas dimensões.

O argumento da função de onda em (2.1) dado por $(\omega_c t - \mathbf{v} \cdot \mathbf{r}^T)$ é denominado de fase. Dessa forma, podemos ainda definir *frente de onda* como o conjunto de pontos com mesma fase, em um tempo fixo t .

Por sua vez, a versão em banda básica da função de onda transmitida pode ser expressa genericamente por

$$E(t, \mathbf{r}) = s(t)e^{-j(\mathbf{v} \cdot \mathbf{r}^T)}, \quad (2.2)$$

na qual $s(t) = A(t)e^{j\alpha(t)}$ é o sinal transmitido, suposto variar mais lentamente do que a portadora $e^{j\omega_c t}$ [Zatman98].

2.3.2 Parâmetro DOA

Considerando um meio linear, no qual o princípio da superposição é válido, podemos considerar os efeitos de várias ondas separadamente. A expressão (2.2) traz informações no tempo e no espaço, sendo, portanto, um modelo adequado para distinguir sinais com parâmetros

espaço-temporal distintos. Dentre esses parâmetros destacamos o parâmetro DOA (direção de chegada) ou AOA (ângulo de chegada), geralmente definido pelo ângulo de elevação em conjunto com o azimute [Haykin85], [Krim96].

Podemos notar que o argumento do vetor \mathbf{v} indica a direção de propagação da onda plana. Portanto, teremos

$$\mathbf{v} = |\mathbf{v}| [\text{sen}\theta, \cos\theta], \quad (2.3)$$

no qual o par $(\text{sen}\theta, \cos\theta)$ indica uma coordenada no plano, e θ indica o ângulo de incidência vertical da onda, definido no sentido horário relativo ao eixo vertical, conforme a *Figura 2.1*. O parâmetro DOA, dado pelo valor do ângulo θ , define, portanto, a direção angular de cada fonte de sinal.

Considerando que um dado sensor n é representado por um receptor pontual em uma coordenada espacial dada pelo vetor $\mathbf{r}_n = [r_{nx}, r_{ny}]^T$, o campo medido nesse sensor devido a uma fonte m com elevação θ_m (DOA) em um dado instante t será dado por

$$E_m(t, \mathbf{r}_n) = s_m(t) e^{-j[|\mathbf{v}_m|(r_{nx}\text{sen}\theta_m + r_{ny}\cos\theta_m)]}. \quad (2.4)$$

Podemos, então, modelar o sinal obtido na saída do n 'ésimo sensor relativo à m 'ésima fonte de sinal como

$$x_{n,m}(t) = g_n(\theta_m) e^{-j|\mathbf{v}_m|(r_{nx}\text{sen}\theta_m + r_{ny}\cos\theta_m)} s_m(t), \quad (2.5)$$

na qual $g_n(\theta_m)$ é o ganho do sensor n na direção θ_m .

O parâmetro DOA, conforme a aplicação, pode trazer informações sobre a posição de alvos em radar, sonar ou a localização de sinais interferentes em sistemas de comunicações.

Em algumas aplicações, o parâmetro DOA pode ser utilizado juntamente com outros parâmetros como velocidade de propagação, polarização e propriedades espectrais dos sinais captados. Desta forma, cada emissor é geralmente associado a um conjunto desses parâmetros, possibilitando uma melhor caracterização e distinção dos mesmos.

Tradicionalmente, técnicas convencionais baseadas em espectros utilizam a medida da potência de saída do arranjo em função do ângulo de direção de chegada para a obtenção do parâmetro DOA. Neste caso, os picos da distribuição de potência são tomados como as verdadeiras direções de chegada dos sinais presentes na porção do meio observada ou “iluminada”.

Nas técnicas paramétricas, o parâmetro DOA é estimado através da escolha do valor que melhor se ajusta ao sinal recebido. O ajuste ótimo para os parâmetros desconhecidos geralmente é obtido através da minimização de uma função custo. Em alguns casos pode ser numericamente favorável realizar uma reparametrização do problema inicial.

2.4 Arranjo Linear Uniforme (U.L.A.)

Os arranjos lineares uniformes são aqueles em que os sensores estão uniformemente dispostos sobre uma linha reta. A grande vantagem do uso de arranjos lineares uniformes é a simplicidade matemática do modelo do sinal na saída do arranjo. Quando uma onda plana atravessa um meio não dispersivo, a linearidade permite representar o sinal na saída de qualquer sensor imerso nesse meio como uma versão retardada ou adiantada no tempo em relação a um sensor

de referência. Desta forma, as diferenças de fase dos sinais recebidos por cada sensor, que estão associadas aos parâmetros DOA's, são transformadas em diferenças temporais e o problema inicial reduz-se a um problema clássico de estimação de atrasos. Por esse motivo, a medida do tempo de atraso é fundamental em problemas de processamento em arranjos lineares.

Uma outra vantagem da geometria linear é a possibilidade de interpretar o arranjo como um filtro temporal com resposta finita (FIR), no qual os ganhos do filtro são dados pelos ganhos dos sensores. Através dessa analogia, todas as técnicas de estimação existentes para séries temporais podem ser aplicadas no problema de determinação do parâmetro DOA no espaço.

Em situações práticas, é possível a existência de várias fontes de sinais em diferentes direções irradiando ao mesmo tempo. Sobre esses sinais geralmente incide ruído gerado pelos amplificadores dos sensores e por outras fontes de ruído não distinguíveis de baixa intensidade.

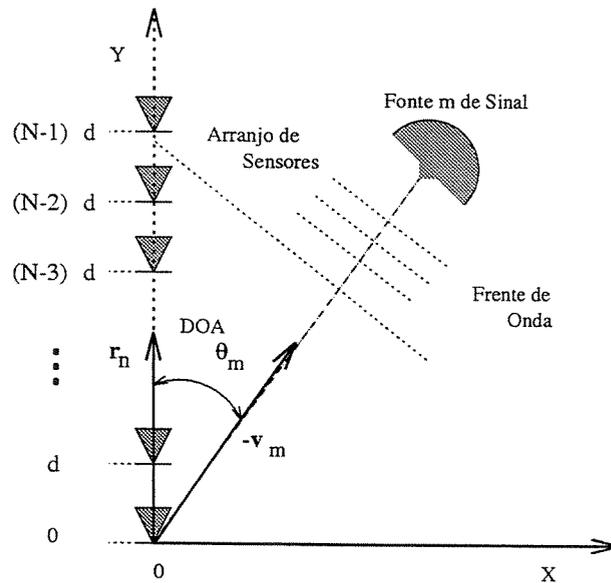


Figura 2.2: Arranjo Linear Uniforme (A.L.U.).

Consideremos, aqui, um arranjo linear vertical composto por N sensores, uniformemente espaçados por uma distância d , com M ondas planas incidentes, de faixa estreita e com a mesma frequência de portadora (ω_c). A Figura 2.2 ilustra a m 'ésima onda incidente na direção indicada pelo vetor \mathbf{v}_m em um arranjo posicionado verticalmente na origem do plano xy . Neste modelo, o n 'ésimo sensor é indicado pelo vetor posição $\mathbf{r}_n = [0, r_{ny}]^T$, com $r_{ny} = nd$.

Considerando o sinal produzido pela m 'ésima onda incidente no n 'ésimo sensor no instante t , conforme (2.5), e tomando como referência o sensor posicionado na origem do arranjo $\mathbf{r}_0 = [0, 0]^T$, teremos na sua saída

$$x_{n,m}(t) = g_n(\theta_m) s_m(t) \exp[-jn(2\pi d/\lambda_c) \cos(\theta_m)], \quad (2.6)$$

com $n = 0, 1, \dots, N-1$, $m = 1, 2, \dots, M$, e $|\mathbf{v}| = 2\pi/\lambda_c$.

Definindo a amplitude complexa do sinal $s_m(t)$ como

$$s_m(t) = A_m(t) \exp[j\alpha_m(t)], \quad (2.7)$$

e o ângulo de fase elétrica, associado a m 'ésima fonte de sinal, como

$$\phi_m = -\frac{2\pi d}{\lambda_c} \cos(\theta_m), \quad (2.8)$$

teremos (2.6) dado por

$$x_{n,m}(t) = s_m(t)a_n(\phi_m), \quad (2.9)$$

na qual $a_n(\phi_m) = g_n(\phi_m) \exp[jn\phi_m]$.

Assumindo que todos os elementos possuem a mesma diretividade $g_n(\phi_m) = g(\phi_m)$, para $n = 0, 1, \dots, N - 1$, e que o ganho independe da direção, como no caso de um arranjo composto por antenas omnidirecionais, teremos $g(\phi_m) = g$, para $m = 1, 2, \dots, M$. Podemos ainda considerar, sem perda de generalidade, $g = 1$, obtendo $a_n(\phi_m) = \exp[jn\phi_m]$.

Podemos concluir, portanto, que para sinais de banda estreita, quando então $s_m(t)$ varia lentamente com o tempo comparado com a portadora, os sinais recebidos pelos sensores diferem entre si apenas por uma defasagem, a qual é um múltiplo de ϕ_m dado em (2.8). Esta, por sua vez, depende somente do espaçamento entre os sensores e do ângulo de chegada da onda plana considerada.

Podemos notar ainda que com o conhecimento de ϕ_m poderemos inferir o ângulo de incidência θ_m através da relação

$$\theta_m = \cos^{-1} \left[-\frac{\lambda\phi_m}{2\pi d} \right]. \quad (2.10)$$

Interpretando o ângulo de fase elétrica ϕ_m como uma frequência e levando em conta a expressão em (2.9), toda técnica existente para estimar frequências de exponenciais complexas poderá ser utilizada para estimar o parâmetro DOA.

A expressão (2.9) define, portanto, o sinal complexo na saída do n 'ésimo sensor devido à m 'ésima onda plana, no instante t . O somatório das contribuições produzidas pelas M ondas planas incidentes resultará no sinal $x_n(t)$ dado pelo n 'ésimo sensor. Isto é,

$$x_n(t) = \sum_{m=1}^M s_m(t)a_n(\phi_m), \quad \text{para } n = 0, 1, \dots, N - 1. \quad (2.11)$$

2.4.1 Modelo do Sinal

A essência do processamento de sinais de um arranjo voltado para a determinação do ângulo de chegada de ondas planas consiste em um problema de estimação de parâmetros. Os parâmetros podem ser: azimuth, elevação, polarização, forma de onda, frequência temporal, etc. O parâmetro DOA, no caso, é composto pelos parâmetros azimuth e elevação do sinal. No caso de arranjos lineares fixos e considerando o azimuth conhecido, apenas o parâmetro elevação é necessário para determinar o parâmetro DOA (*ver Figura 2.2*).

A maior parte das técnicas modernas de abordagem do processamento de arranjo são baseadas em modelos para o sinal. Para a determinação de um modelo útil algumas considerações devem ser estabelecidas com relação ao sinal observado.

Considerando o sinal composto em (2.11) para cada um dos N sensores do arranjo, teremos o sinal na saída desse arranjo, associado a cada onda incidente, dado vetorialmente por

$$\mathbf{x}_m(t) = s_m(t)\mathbf{a}(\phi_m), \text{ para } m = 1, 2, \dots, M, \quad (2.12)$$

com

$$\mathbf{a}(\phi_m) = \begin{bmatrix} 1 & e^{j\phi_m} & \dots & e^{j(N-1)\phi_m} \end{bmatrix}^T, \quad (2.13)$$

onde $\mathbf{a}(\phi_m)$ é denominado de vetor de direção (“steering vector”).

Considerando, agora, o sinal produzido pelas M fontes de sinais em (2.12), teremos

$$[\mathbf{x}(t)]_{N \times 1} = \sum_{m=1}^M s_m(t)\mathbf{a}(\phi_m), \quad (2.14)$$

na qual o subscrito $N \times 1$ indica a dimensão do vetor.

Na prática, o sinal observado na saída de qualquer um dos sensores difere do sinal teórico por uma componente aditiva produzida por ruído. O ruído considerado pode ser proveniente do espaço (ruído ambiental), do amplificador do receptor (ruído térmico) ou mesmo provocado por erros de medida. Este ruído é geralmente modelado como branco, Gaussiano e de média nula. Podemos, assim, modelar o sinal na saída do arranjo de sensores como

$$[\mathbf{y}(t)]_{N \times 1} = [\mathbf{x}(t)]_{N \times 1} + [\mathbf{n}(t)]_{N \times 1}, \quad (2.15)$$

onde $\mathbf{n}(t)$ representa o somatório das fontes de ruído. Consideraremos a hipótese de que as causas naturais responsáveis pela geração do sinal e do ruído não são relacionadas, assumindo que o sinal e o ruído são mutuamente decorrelacionados.

De forma mais compacta e geral podemos reescrever o sinal em (2.15) como

$$[\mathbf{y}(t)]_{N \times 1} = [\mathbf{A}(\phi)]_{N \times M} [\mathbf{s}(t)]_{M \times 1} + [\mathbf{n}(t)]_{N \times 1}, \text{ com } \phi = \{\phi_1, \dots, \phi_M\}, \quad (2.16)$$

onde $\mathbf{s}(t) = \begin{bmatrix} s_1(t) & \dots & s_M(t) \end{bmatrix}^T$ e $\mathbf{A}(\phi) = \begin{bmatrix} \mathbf{a}(\phi_1) & \dots & \mathbf{a}(\phi_M) \end{bmatrix}$ é a matriz cujas colunas são os vetores de direção, a qual possui uma estrutura bastante simples e útil, da forma

$$\mathbf{A}(\phi) = \begin{bmatrix} 1 & 1 & \dots & 1 \\ e^{j\phi_1} & e^{j\phi_2} & \dots & e^{j\phi_M} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j(N-1)\phi_1} & e^{j(N-1)\phi_2} & \dots & e^{j(N-1)\phi_M} \end{bmatrix}_{N \times M}, \text{ com } N \geq M, \quad (2.17)$$

conhecida como matriz de *Vandermonde*, a qual passaremos a nos referir apenas como \mathbf{A} . Podemos notar que para o caso de fontes de sinais totalmente independentes, com ângulos elétricos de valores distintos de tal forma que não exista uma combinação linear entre seus valores, teremos as colunas de \mathbf{A} linearmente independentes e, conseqüentemente, o posto de \mathbf{A} será cheio e igual à M . Caso contrário, o posto de \mathbf{A} será menor que M .

No decorrer deste trabalho consideraremos que a matriz \mathbf{A} tem posto cheio e igual a M independentemente dos valores de ϕ_i , para $i = 1, 2, \dots, M$.

Na prática, dispomos apenas de amostras do sinal em (2.16) tomadas nos instantes $t = k\Delta t$, onde Δt indica o intervalo de amostragem do sinal e k um número inteiro positivo.

Consideraremos, ainda, o fato de dispormos de apenas um quantidade finita de amostras do sinal $\mathbf{y}(t)$, consideradas nos instantes $t = k\Delta t$, para $k = 1, \dots, K$, dando origem à seqüência

$\{\mathbf{y}(k)\}_{k=1}^K$, onde K indica o número de instantes de amostragem ou experimentos disponíveis (“snapshots”).

Dois diferentes tipos de sinal podem surgir do modelo em (2.16) quanto ao sinal $\mathbf{s}(t)$ para $t = k\Delta t$, dado por $\mathbf{s}(k)$. O primeiro é denominado de modelo condicional ou determinístico e assume que a seqüência $\mathbf{s}(k)$ não é aleatória, embora desconhecida. Isto é, a seqüência $\mathbf{s}(k)$ não é alterada em todas as realizações de $\mathbf{y}(k)$. Neste modelo apenas a seqüência $\mathbf{n}(k)$ varia a cada realização. No segundo modelo, denominado de incondicional ou estocástico, a seqüência $\mathbf{s}(k)$ é considerada aleatória, sendo alterada a cada realização de $\mathbf{y}(k)$ [Stoica90c].

A denominação de modelo determinístico/aleatório não é apropriada uma vez que $\mathbf{s}(k)$, para k fixo, pode ser aleatório em termos de suas M componentes, em ambos os modelos. Devido a essa interpretação a terminologia condicional/incondicional será utilizada para os dois modelos descritos.

2.5 Aspectos Básicos

Introduziremos, agora, a notação e formulações matemáticas que servirão de base para os diversos métodos existentes voltados para a obtenção do parâmetro DOA. A apresentação prematura das expressões e resultados a seguir visa evitar a replicação daquelas comuns aos métodos a serem apresentados, bem como facilitar uma comparação futura entre os mesmos.

Conforme foi enfatizado anteriormente, a correlação é uma das ferramentas mais úteis na estimação de parâmetros. Essa informação é obtida através da matriz de correlação espacial para o sinal na saída do arranjo. Partindo de (2.16), com $t = k\Delta t$, teremos

$$\mathbf{R}(k, i) = E \{ \mathbf{y}(k) \mathbf{y}^H(i) \} = \mathbf{A} E \{ \mathbf{s}(k) \mathbf{s}^H(i) \} \mathbf{A}^H + E \{ \mathbf{n}(k) \mathbf{n}^H(i) \}, \quad (2.18)$$

com $k, i = 1, 2, \dots, K$, onde $E \{ \cdot \}$ simboliza esperança estatística, e $E \{ \mathbf{n}(k) \mathbf{n}^H(i) \} = \sigma^2 \mathbf{I} \delta_{k,i}$ é a matriz de correlação do ruído, na qual $\delta_{k,i}$ simboliza a função *delta de Kronecker* ($\delta_{k,i} = 1$, para $k = i$ e $\delta_{k,i} = 0$, caso contrário). O ruído com essas características é denominado de ruído espacialmente branco.

Em um modelo genérico consideramos o sinal $\mathbf{s}(k)$ definido pelos parâmetros amplitude e fase em um dado instante de tempo $t = k\Delta t$. Neste modelo, tanto a amplitude como a fase podem variar com o tempo. Neste caso, a matriz de correlação entre esses sinais será dada por $E \{ \mathbf{s}(k) \mathbf{s}^H(i) \} = \mathbf{P}(k, i)$, onde $\mathbf{P}(k, i)$ é uma matriz Hermitiana.

Para o caso particular em que consideramos $\mathbf{s}(k)$ como um sinal tipo ruído branco, teremos [Stoica90c]

$$\begin{aligned} E \{ \mathbf{s}(k) \mathbf{s}^H(i) \} &= \mathbf{P} \delta_{k,i}, \text{ com } \det(\mathbf{P}) \neq 0 \\ &\text{e} \\ E \{ \mathbf{s}(k) \mathbf{s}^T(i) \} &= 0, \text{ para todo } k \text{ e } i. \end{aligned} \quad (2.19)$$

Nesse modelo também é assumido que o sinal $\mathbf{s}(k)$ e o ruído $\mathbf{n}(i)$ são descorrelacionados para todo k e i .

Levando em conta (2.19), a correlação espacial em (2.18) assume a forma

$$\mathbf{R} = \mathbf{R}(k, k) = E \{ \mathbf{y}(k) \mathbf{y}^H(k) \} = \mathbf{A} E \{ \mathbf{s}(k) \mathbf{s}^H(k) \} \mathbf{A}^H + E \{ \mathbf{n}(k) \mathbf{n}^H(k) \}. \quad (2.20)$$

O sinal $\mathbf{s}(k)$ pode ser qualquer sinal aleatório, porém o modelo utilizado em (2.19) foi escolhido por ser matematicamente tratável.

Considerando a matriz \mathbf{P} em (2.19), podemos estabelecer três situações quanto à correlação cruzada entre os sinais provenientes de M fontes distintas: totalmente decorrelacionados ou independentes, totalmente correlacionados ou coerentes e parcialmente correlacionados.

Genericamente, podemos explicitar a matriz \mathbf{P} como a seguir

$$\mathbf{P} = \begin{bmatrix} \rho_{1,1} & \rho_{1,2} & \cdots & \rho_{1,M-1} & \rho_{1,M} \\ \rho_{2,1} & \rho_{2,2} & \rho_{2,3} & \cdots & \rho_{2,M} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho_{M-1,1} & \cdots & \rho_{M-1,M-2} & \rho_{M-1,M-1} & \rho_{M-1,M} \\ \rho_{M,1} & \rho_{M,2} & \cdots & \rho_{M,M-1} & \rho_{M,M} \end{bmatrix}, \quad (2.21)$$

onde $\rho_{i,j}$ indica a correlação cruzada entre os sinais das fontes i e j .

Para o caso em que os sinais são parcialmente correlacionados, a matriz \mathbf{P} será Hermitiana e $1 < \text{Posto}(\mathbf{P}) \leq M$.

Para o caso em que os sinais são decorrelacionados, a matriz \mathbf{P} será diagonal, com $\rho_{i,j} = 0$ para $i \neq j$, e, conseqüentemente, $\text{Posto}(\mathbf{P}) = M$. No caso particular de sinais decorrelacionados normalizados e com mesma potência, teremos $\mathbf{P} = \mathbf{I}$.

Um outro caso particular com relação ao sinal $\mathbf{s}(k)$, porém de bastante interesse, é quando os sinais são coerentes. Dizemos que dois sinais são coerentes se um é uma versão escalonada e/ou defasada do outro. Isto é, todos os sinais $\{s_m(k)\}$ para $\forall m$ são múltiplos complexos de um sinal comum, $s_0(k)$. Assim $\mathbf{s}(k)$ pode ser expresso como

$$\mathbf{s}(k) = \mathbf{p}s_0(k), \quad (2.22)$$

na qual \mathbf{p} é um vetor de dimensões $M \times 1$ com elementos complexos quaisquer, porém fixos.

Nesse caso podemos notar que a matriz $\mathbf{P} = \rho_0 \mathbf{p}\mathbf{p}^H$, com ρ_0 indicando a potência do sinal $s_0(k)$, possuirá posto unitário.

A coerência entre sinais provenientes de direções distintas é um fenômeno comum, resultado do efeito de propagação através de múltiplos percursos [Sheinvald96], [Krim96]. O resultado dessa coerência se apresenta através de uma deficiência no posto da matriz \mathbf{P} . Por outro lado, para sinais não-correlacionados, a matriz \mathbf{P} será diagonal e possuirá posto cheio.

No decorrer deste trabalho consideraremos a hipótese de que a matriz \mathbf{P} possui posto completo, também aqui denominado de posto cheio.

A forma decomposta da matriz \mathbf{R} em (2.20), considerando a matriz \mathbf{P} , será dada por

$$\mathbf{R} = \mathbf{A}\mathbf{P}\mathbf{A}^H + \sigma^2 \mathbf{I}_{N \times N}, \quad (2.23)$$

a qual será de grande importância para a descrição dos métodos em capítulos seguintes.

Consideraremos, ainda, a seguinte possibilidade de decomposição para a matriz \mathbf{R}

$$\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H, \quad (2.24)$$

na qual \mathbf{U} é uma matriz unitária cujas colunas são os autovetores de \mathbf{R} e $\mathbf{\Lambda}$ uma matriz diagonal cujos elementos são os autovalores de \mathbf{R} ordenados de forma decrescente.

Considerando a existência de apenas $M < N$ fontes de sinal independentes em uma situação sem ruído, teremos teoricamente $\mathbf{\Lambda} = \text{diag} \{ \lambda_1, \lambda_2, \dots, \lambda_M, \lambda_{M+1}, \dots, \lambda_N \}$, com $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M > \lambda_{M+1} = \lambda_{M+2} = \dots = \lambda_N = 0$. Com a presença de ruído branco com variância σ^2 teremos $\mathbf{\Lambda} = \text{diag} \{ \lambda_1 + \sigma^2, \lambda_2 + \sigma^2, \dots, \lambda_M + \sigma^2, \lambda_{M+1}, \dots, \lambda_N \}$, com $\lambda_{M+1} = \lambda_{M+2} = \dots = \lambda_N = \sigma^2$. Portanto, indo além na decomposição apresentada em (2.24), podemos reescrever a matriz \mathbf{R} , de dimensões $N \times N$, como

$$\mathbf{R} = \mathbf{U}_s \mathbf{\Lambda}_s \mathbf{U}_s^H + \mathbf{U}_n \mathbf{\Lambda}_n \mathbf{U}_n^H, \quad (2.25)$$

na qual $\mathbf{\Lambda}_n = \sigma^2 \mathbf{I}_{N-M \times N-M}$ e as colunas de $[\mathbf{U}_n]_{N \times N-M}$ correspondem, respectivamente, aos autovalores e autovetores associados ao subespaço de ruído. Da mesma forma, as colunas de $[\mathbf{U}_s]_{N \times M}$ e $\mathbf{\Lambda}_s = \text{diag} \{ \lambda_1, \lambda_2, \dots, \lambda_M \}$ correspondem, respectivamente, aos autovetores e autovalores associados ao subespaço de sinal.

Comparando a decomposição apresentada em (2.23) com (2.25), podemos observar que se o posto de $\mathbf{A} \mathbf{P} \mathbf{A}^H$ for cheio, os elementos da matriz $\mathbf{\Lambda}_s$ serão dados pelos M maiores autovalores não nulos de \mathbf{R} . Considerando, ainda, que o posto de \mathbf{A} é cheio e igual a M , qualquer vetor ortogonal à matriz \mathbf{A} será um autovetor da matriz \mathbf{R} , cujo autovalor será σ^2 . Para as dimensões envolvidas existem $N - M$ vetores linearmente independentes com essa propriedade, e tais pares (autovetores e autovalores) correspondem ao subespaço de ruído. Os outros M autovalores maiores que σ^2 , por sua vez, estão relacionados com os autovetores associados ao subespaço de sinal. Portanto, todos os autovetores de ruído serão ortogonais às colunas da matriz \mathbf{A} , e as colunas de \mathbf{U}_s deverão gerar o mesmo espaço da matriz \mathbf{A} , enquanto que as colunas de \mathbf{U}_n geram o seu complemento ortogonal, denominado de espaço nulo de \mathbf{A} . [Krim96], [Jansson99].

Os operadores de projeção no subespaço de sinal e subespaço de ruído são então definidos, respectivamente, como [Kay93]

$$\mathbf{\Pi}_A = \mathbf{A} \mathbf{A}^\dagger = \mathbf{U}_s \mathbf{U}_s^H \quad (2.26)$$

e

$$\mathbf{\Pi}_A^\perp = \mathbf{I} - \mathbf{A} \mathbf{A}^\dagger = \mathbf{U}_n \mathbf{U}_n^H, \quad (2.27)$$

onde reconhecemos $\mathbf{A}^\dagger = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H$ como a *pseudo-inversa* de Moore-Penrose. A segunda igualdade em (2.26) e (2.27) pode ser obtida considerando a decomposição em valores singulares da matriz \mathbf{A} . Considerando a existência das matrizes inversas em (2.26) e (2.27), teremos $\mathbf{I} = \mathbf{\Pi}_A + \mathbf{\Pi}_A^\perp$.

No caso de amostras finitas, as esperanças estatísticas presentes em (2.20) serão substituídas por médias algébricas, dando origem à matriz de correlação estimada

$$\hat{\mathbf{R}} = \frac{1}{K} \sum_{k=1}^K \mathbf{y}(k) \mathbf{y}^H(k) = \mathbf{A} \left[\frac{1}{K} \sum_{k=1}^K \mathbf{s}(k) \mathbf{s}^H(k) \right] \mathbf{A}^H + \frac{1}{K} \sum_{k=1}^K \mathbf{n}(k) \mathbf{n}^H(k). \quad (2.28)$$

onde K indica o número de experimentos e $\mathbf{y}(k)$ representa o vetor de amostras N -dimensional presente na saída do arranjo de sensores no instante $t = k\Delta t$, perfazendo um total de KN amostras disponíveis.

Conseqüentemente, a representação de \mathbf{R} decomposta em subespaços em (2.25), quando reescrita para o caso de amostras finitas, será dada por

$$\hat{\mathbf{R}} = \hat{\mathbf{U}}_s \hat{\mathbf{\Lambda}}_s \hat{\mathbf{U}}_s^H + \hat{\mathbf{U}}_n \hat{\mathbf{\Lambda}}_n \hat{\mathbf{U}}_n^H, \quad (2.29)$$

onde $\hat{\mathbf{U}}_s, \hat{\mathbf{\Lambda}}_s, \hat{\mathbf{U}}_n$, e $\hat{\mathbf{\Lambda}}_n$ são estimativas de suas versões estatísticas.

2.6 Conclusão

Nesse capítulo apresentamos o contexto do desenvolvimento desse trabalho. Definimos arranjo de sensores desde a forma mais geral até a mais específica, na qual destacamos os arranjos lineares e uniformes compostos por sensores idênticos. Definimos também o ângulo de incidência de uma onda plana sobre um arranjo de sensores como o parâmetro DOA de nosso interesse. Apresentamos um histórico das técnicas que fazem uso de arranjo de sensores. Na evolução dessa apresentação, destacamos as diferenças entre os procedimentos paramétricos e não-paramétricos de processamento. Deve ficar claro que não houve a pretensão de se apresentar um tutorial sobre técnicas de processamento de arranjos, visto que técnicas precursoras como *beamforming* e novas técnicas de processamento tempo-espaço, dentre outras técnicas, não foram abordadas.

Ficou, então, estabelecido que o contexto deste trabalho compreenderá as técnicas paramétricas de processamento de arranjos de sensores uniformes e lineares aplicado à estimação do parâmetro DOA.

No decorrer da apresentação deste capítulo procurou-se estabelecer uma notação mais próxima daquela encontrada na literatura e que será utilizada nos capítulos seguintes. Algumas modificações foram necessárias para privilegiar a clareza das apresentações que se seguem. Dentre essas, chamamos atenção para o uso de ϕ_m em detrimento de θ_m , onde ϕ_m é o parâmetro determinado pelos procedimentos de estimação e θ_m é o parâmetro DOA de interesse, obtido a partir de ϕ_m através de uma relação não-linear.

Alguns resultados referentes à decomposição espectral da matriz de correlação do sinal foram adiantados neste capítulo. Nos capítulos seguintes faremos uso das expressões aqui apresentadas.

Um outro ponto importante que deve ser salientado é a contribuição em termos de referências bibliográficas voltadas para o processamento de arranjo de sensores, que possibilitará a muitos leitores iniciar pesquisas nesse assunto.

Capítulo 3

Estratégias para Estimação do Parâmetro DOA

3.1 Introdução

Dentre os procedimentos paramétricos para estimação, os algoritmos baseados no critério de probabilidade por *Máxima-verossimilhança* (ML) apresentam uma grande capacidade de resolução e nenhuma dificuldade particular na aplicação em arranjos de qualquer geometria ou para o caso de dados multidimensionais. Porém, tal critério exige um esforço computacional proibitivo [Stoica90b], [Stoica89], [Stoica90d]. Assim, novos procedimentos derivados do critério ML surgiram para superar sua grande carga computacional e ao mesmo tempo manter níveis próximos de desempenho [Bresler86], [Stoica90a].

Dentre as alternativas ao critério ML para estimação de parâmetros destaca-se a predição linear [Makhoul75]. Mostraremos que a predição linear pode ser utilizada para a estimação das frequências do sinal presente na entrada de um filtro preditor otimizado [Tufts82], [Colares97]. Tal procedimento pode ser naturalmente estendido para a estimação do parâmetro DOA de sinais incidentes em um arranjo de sensores através da dualidade existente entre a frequência e o ângulo de incidência em arranjos lineares com sensores uniformemente espaçados [Krim96], [Kumaresa83a].

Consideramos, ainda, uma terceira classe de estimadores, denominada aqui de estimadores totalmente baseados em subespaço. Esta consiste basicamente no uso das propriedades apresentadas pelos subespaços de sinal e ruído definidos pelo sinal observado a partir da decomposição de sua matriz de correlação [Stoica91], [Viberg91a]. A denominação utilizada visa evitar confusões com as versões em subespaços existentes para os procedimentos baseados em predição linear e para o critério ML, as quais também serão apresentadas [Stoica90a]. Para essa terceira classe de estimadores, destacaremos aqueles baseados no critério da “proximidade” entre subespaços (“Subspace Fitting”) [Viberg91b], [Paulraj86], [Jansson99].

Através de uma reparametrização do problema de estimação inicial será possível descrever as três classes de estimadores de forma unificada, que é uma das principais contribuições desse trabalho.

3.2 Máxima-verossimilhança (ML)

O procedimento de estimação através do critério ML será descrito a seguir considerando duas possibilidades para o sinal $\mathbf{s}(t)$; determinístico (condicional) e estocástico (incondicional) [Stoica90c], [Krim96], [Wax91].

3.2.1 Máxima-verossimilhança para Sinais Determinísticos (MLD)

Como consequência do modelo do sinal apresentado no capítulo anterior, quando o sinal $\mathbf{s}(t)$ é considerado determinístico, uma única medida do vetor $\mathbf{y}(t)$ será uma variável aleatória Gaussiana de dimensão N , com matriz de correlação $\sigma^2 \mathbf{I}$, média $\mathbf{A}(\phi)\mathbf{s}(t)$ e função de densidade de probabilidade dada por [Kay93], [Papoulis91]

$$\frac{1}{(\pi\sigma^2)^N} e^{-\|\mathbf{y}(t) - \mathbf{A}(\phi)\mathbf{s}(t)\|^2/\sigma^2}, \quad (3.1)$$

na qual $\|\bullet\|$ denota a norma *Euclidiana*. Tomando-se um conjunto de K medidas independentes desse sinal, nos instantes $t = k\Delta t$, para $k = 1, 2, \dots, K$, a função de verossimilhança é obtida como [Bresler86]

$$L_{MLD}(\phi, \mathbf{s}(k), \sigma^2) = \prod_{k=1}^K (\pi\sigma^2)^{-N} e^{-\|\mathbf{y}(k) - \mathbf{A}(\phi)\mathbf{s}(k)\|^2/\sigma^2}, \quad \text{com } \phi \in \mathcal{R}^M, \quad (3.2)$$

na qual os parâmetros desconhecidos são as direções angulares ϕ_1, \dots, ϕ_M , o vetor do sinal $\mathbf{s}(k)$, e a variância do ruído σ^2 . Os estimadores ML desses parâmetros são aqueles que maximizam o valor da função $L_{MLD}(\phi, \mathbf{s}(k), \sigma^2)$. Por conveniência matemática, as estimativas são definidas como os *argumentos* que minimizam a função $l_{MLD}(\phi, \mathbf{s}(k), \sigma^2) = -\log [L_{MLD}(\phi, \mathbf{s}(k), \sigma^2)]$. Normalizando-se em relação a K e ignorando o termo constante, $N \log \pi$, teremos

$$l_{MLD}(\phi, \mathbf{s}(k), \sigma^2) = N \log \sigma^2 + \frac{1}{K\sigma^2} \sum_{k=1}^K \|\mathbf{y}(k) - \mathbf{A}(\phi)\mathbf{s}(k)\|^2. \quad (3.3)$$

A partir da hipótese de que os parâmetros podem ser estimados de forma separada, demonstra-se que as estimativas dadas por [Stoica90c]

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{N-M} \text{Tr} \left\{ \Pi_{\mathbf{A}}^\perp \hat{\mathbf{R}} \right\} = \frac{1}{N-M} \text{Tr} \left\{ \hat{\mathbf{\Lambda}}_n \right\} \\ &\text{e} \\ \hat{\mathbf{s}}(k) &= \mathbf{A}^\dagger \mathbf{y}(k) = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{y}(k), \end{aligned} \quad (3.4)$$

são aquelas que minimizam a função dada por (3.3), onde $\text{Tr} \{\bullet\}$ denota o *traço* da matriz [Golub89]. A estimativa $\hat{\mathbf{s}}(k)$ em (3.4) foi considerada como aquela que minimiza o erro de estimação quadrático [Sorenson70]. Substituindo (3.4) em (3.3), e observando que $\frac{1}{K} \sum_{k=1}^K \|\Pi_{\mathbf{A}}^\perp \mathbf{y}(k)\|^2 = \text{Tr} \left\{ \Pi_{\mathbf{A}}^\perp \hat{\mathbf{R}} \right\}$, transformamos o problema original no seguinte problema de minimização

$$\min_{\phi} J_{MLD}(\phi) = \min_{\phi} \text{Tr} \left\{ \Pi_{\mathbf{A}}^\perp(\phi) \hat{\mathbf{R}} \right\}, \quad \text{com } \phi \in \mathcal{R}^M, \quad (3.5)$$

no qual o subscrito *MLD* significa ML determinístico, considerando $\mathbf{s}(k)$. O parâmetro ϕ foi explicitado na expressão (3.5) para enfatizar a variável livre no processo de minimização. Neste caso, $\hat{\mathbf{R}}$ é o estimador obtido a partir das amostras disponíveis do sinal observado.

Para a obtenção da estimativa em (3.5) é necessário a solução numérica de um problema de otimização não-linear de quarta ordem M -dimensional, sujeita a problemas de convergência [Bresler86], [Stoica90a].

3.2.2 Máxima-verossimilhança para Sinais Aleatórios (MLE)

Uma variação do procedimento ML apresentado, denominada de ML estocástico (*MLE*), é obtida através do modelamento do sinal $\mathbf{s}(k)$ como um processo aleatório estacionário no tempo do tipo ruído branco, com média nula e matriz de correlação \mathbf{P} , conforme o modelo apresentado no capítulo anterior. Adotaremos a hipótese que a função de densidade de probabilidade deste processo é Gaussiana, embora tal procedimento seja aplicado mesmo nos casos em que os dados não são Gaussianos, sendo tal condição imposta apenas para torná-lo matematicamente tratável [Kay93], [Papoulis91].

Como conseqüência da hipótese Gaussiana para $\mathbf{s}(k)$, e ainda considerando a presença de ruído Gaussiano dado pelo vetor $\mathbf{n}(k)$, uma única medida do vetor $\mathbf{y}(k) = \mathbf{A}(\phi)\mathbf{s}(k) + \mathbf{n}(k)$ também será um processo aleatório estacionário Gaussiano com matriz de correlação \mathbf{R} , média nula e função de densidade de probabilidade dada por [Kay93]

$$\frac{1}{(\pi)^N |\mathbf{R}|^N} e^{-[\mathbf{y}(t)^T \mathbf{R}^{-1} \mathbf{y}(t)]}, \quad (3.6)$$

na qual $\mathbf{R} = \mathbf{A}(\phi)\mathbf{P}\mathbf{A}^H(\phi) + \sigma^2\mathbf{I}$ e $|\bullet|$ denota o valor do *determinante* [Golub89]. Tomando-se um conjunto de K medidas independentes desse sinal, nos instantes $t = k\Delta t$, para $k = 1, 2, \dots, K$, a função de verossimilhança é obtida como em (3.2) [Krim96], [Wax91]

$$L_{MLE} = \prod_{k=1}^K (\pi)^{-N} |\mathbf{R}|^{-N} e^{-[\mathbf{y}(k)^T \mathbf{R}^{-1} \mathbf{y}(k)]}. \quad (3.7)$$

Nesse caso, o conjunto de parâmetros desconhecidos torna-se diferente do caso determinístico. A função de máxima verossimilhança dependerá agora dos parâmetros ϕ , \mathbf{P} e σ^2 em \mathbf{R} .

Tomando-se a função $l_{MLE} = -\log[L_{MLE}]$, e normalizando em relação à K , teremos

$$l_{MLE}(\phi, \mathbf{P}, \sigma^2) = N \log \pi + N \log |\mathbf{R}| + \frac{1}{K} \sum_{k=1}^K [\mathbf{y}(k)^H \mathbf{R}^{-1} \mathbf{y}(k)]. \quad (3.8)$$

Reconhecendo a seguinte igualdade $\frac{1}{K} \sum_{k=1}^K [\mathbf{y}(k)^H \mathbf{R}^{-1} \mathbf{y}(k)] = Tr \left[\mathbf{R}^{-1} \hat{\mathbf{R}} \right]$ e substituindo em (3.8), teremos

$$l_{MLE}(\phi, \mathbf{P}, \sigma^2) = N \log \pi + N \log |\mathbf{R}| + Tr \left[\mathbf{R}^{-1} \hat{\mathbf{R}} \right]. \quad (3.9)$$

Considerando o comportamento assintótico ($K \rightarrow \infty$) da função em (3.9), juntamente com a hipótese de sinal estacionário, teremos $\lim_{K \rightarrow \infty} \hat{\mathbf{R}} = \mathbf{R}$. Conseqüentemente, após a eliminação

dos termos constantes, a função de máxima-verossimilhança será reduzida à seguinte função equivalente

$$l_{MLE}(\phi, \mathbf{P}, \sigma^2) = \log |\mathbf{R}| = \log \left| \mathbf{A}(\phi) \mathbf{P} \mathbf{A}^H(\phi) + \sigma^2 \mathbf{I} \right|. \quad (3.10)$$

Para ϕ fixo pode ser demonstrado que o mínimo da função (3.10) com relação aos parâmetros \mathbf{P} e σ^2 é dado por [Krim96], [Wax91]

$$\begin{aligned} \hat{\sigma}^2(\phi) &= \frac{1}{N-M} \text{Tr} \left\{ \Pi_{\mathbf{A}}^\perp \hat{\mathbf{R}} \right\} \\ \hat{\mathbf{P}}(\phi) &= \mathbf{A}^\dagger \left(\hat{\mathbf{R}} - \hat{\sigma}^2(\phi) \mathbf{I} \right) \mathbf{A}^{H\dagger}, \end{aligned} \quad (3.11)$$

nas quais $\hat{\mathbf{R}}$ é dado por (2.28). A estimativa $\hat{\mathbf{P}}(\phi)$ em (3.11) é aquela que minimiza o erro de estimação quadrático. Devemos ressaltar que não há garantias que a estimativa $\hat{\mathbf{P}}(\phi)$ em (3.11) seja uma estimativa positiva semidefinida de \mathbf{P} . Se \mathbf{P} for positiva definida e se K for suficientemente grande, sendo tal estimador consistente, a estimativa $\hat{\mathbf{P}}(\phi)$ será positiva definida. No entanto, se \mathbf{P} for singular tal condição não será necessariamente verdadeira [Stoica90c], [Krim96], [Wax91].

Substituindo as estimativas (3.11) em (3.10), obtemos

$$\min_{\phi} J_{MLE}(\phi) = \min_{\phi} \log \left(\left| \mathbf{A} \hat{\mathbf{P}}(\phi) \mathbf{A}^H + \hat{\sigma}^2(\phi) \mathbf{I} \right| \right), \text{ com } \phi \in \mathcal{R}^M, \quad (3.12)$$

no qual o subscrito *MLE* significa ML estocástico.

Podemos observar que o problema de minimização em (3.12) é tão complexo quanto o do caso determinístico em (3.5). Demonstra-se que para sinais Gaussianos, a variância do erro das estimativas obtidas por (3.12) atinge assintoticamente o limite inferior de Cramér-Rao (CR) para sinais estocásticos [Stoica90c], [Ottersten92], [Wax91].

Uma terceira variação para o critério ML pode ser obtida para o caso em que os sinais $\{s_m(t)\}$ são coerentes. Nesse caso, poderemos escrever $\mathbf{s}(t) = \mathbf{p}s_0(t)$ conforme (2.22), e o conjunto de parâmetros desconhecidos torna-se ϕ, \mathbf{p} e σ^2 [Sheinvald96].

3.2.3 ML Baseado em Subespaços

Para permitir uma notação unificada dos métodos de DOA, apresentaremos uma alternativa para descrever o estimador ML determinístico através da decomposição em subespaços. Embora tal alternativa seja empregada em alguns métodos na literatura, a descrição a seguir tem como objetivo uma generalização.

Tal alternativa consiste basicamente em escrever a matriz de correlação estimada através de sua versão decomposta em subespaços. Tomando-se a versão estimada da matriz de correlação dada no capítulo anterior por (2.28), teremos

$$\hat{\mathbf{R}} = \hat{\mathbf{U}}_s \hat{\mathbf{\Lambda}}_s \hat{\mathbf{U}}_s^H + \hat{\sigma}^2 \hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H, \quad (3.13)$$

na qual tomamos como estimador consistente para os autovalores de ruído $\hat{\mathbf{\Lambda}}_n = \hat{\sigma}^2 \mathbf{I}$.

Substituindo (3.13) em (3.5) teremos o seguinte problema de minimização

$$\min_{\phi} J_{SML}(\phi) = \min_{\phi} Tr \left\{ \Pi_{\mathbf{A}}^{\perp} \left(\hat{\mathbf{U}}_s \hat{\mathbf{\Lambda}}_s \hat{\mathbf{U}}_s^H + \hat{\sigma}^2 \hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H \right) \right\}, \text{ com } \phi \in \mathcal{R}^M, \quad (3.14)$$

no qual o subscripto *SML* significa ML baseado em subespaço.

Considerando a igualdade $\Pi_{\mathbf{A}}^{\perp} = \mathbf{U}_n \mathbf{U}_n^H$ em (3.14), uma vez que os subespaços foram estimados corretamente devemos esperar $\mathbf{U}_n^H \hat{\mathbf{U}}_s = 0$ quando $\phi = \phi_o$. A partir desse resultado, a função custo descrita em (3.14) permite desprezar o termo associado ao subespaço de ruído, levando ao seguinte problema equivalente

$$\min_{\phi} J_{SML}(\phi) = \min_{\phi} Tr \left\{ \Pi_{\mathbf{A}}^{\perp} \left(\hat{\mathbf{U}}_s \hat{\mathbf{\Lambda}}_s \hat{\mathbf{U}}_s^H \right) \right\}, \text{ com } \phi \in \mathcal{R}^M. \quad (3.15)$$

Algumas variações na determinação da estimativa $\hat{\mathbf{\Lambda}}_s$ podem melhorar o desempenho do estimador em (3.15). Uma possibilidade, que se aplica no caso de sinais totalmente independentes entre si e com potências iguais e normalizadas, é fazer $\hat{\mathbf{\Lambda}}_s = \mathbf{I}$. A vantagem dessa escolha é a redução na complexidade do problema de otimização.

3.3 Predição Linear em Arranjo Linear Uniforme

O procedimento de estimação baseado em predição linear consiste basicamente em se otimizar um filtro de predição para o sinal observado e estimar os parâmetros desejados através das propriedades do filtro resultante [Colares97], [Lemos97], [Antunes92].

Para analisar os conceitos básicos do problema de otimização na predição linear, consideraremos um conjunto de N amostras de um sinal ruidoso $y_n(k)$, $0 \leq n \leq N-1$, dado pelo vetor $\mathbf{y}(k)$, como a entrada do filtro transversal com $(L+1)$ fatores de ganhos, conforme apresentado na *Figura 3.1*. O índice k , conforme dado na Seção 2.4.2, representa a variação de $y_n(k)$ ao longo dos instantes de amostragem $k\Delta t$, $1 \leq k \leq K$.

Pela estrutura do filtro da *Figura 3.1*, podemos observar que o sinal na sua saída será dado por

$$e_n(k) = \sum_{l=0}^L y_{n-l}(k) b_l, \text{ para } n = L, L+1, \dots, N-1. \quad (3.16)$$

Para que esse filtro funcione como um filtro de erro de predição, devemos fazer $b_0 = -1$. Desta forma, a energia total do sinal de erro na saída do filtro, para cada valor de k , será dada por

$$E(k) = \sum_{n=L}^{N-1} |e_n(k)|^2 = \sum_{n=L}^{N-1} \left| \sum_{l=1}^L y_{n-l}(k) b_l - y_n(k) \right|^2. \quad (3.17)$$

A posição do elemento de ganho unitário no filtro corresponde à posição do sensor escolhido, do qual desejamos prever a saída. Uma escolha diferente do elemento unitário pode influir na capacidade de resolução e na polarização das estimativas [Colares97].

Mais especificamente, se definirmos a matriz de dados no instante k como

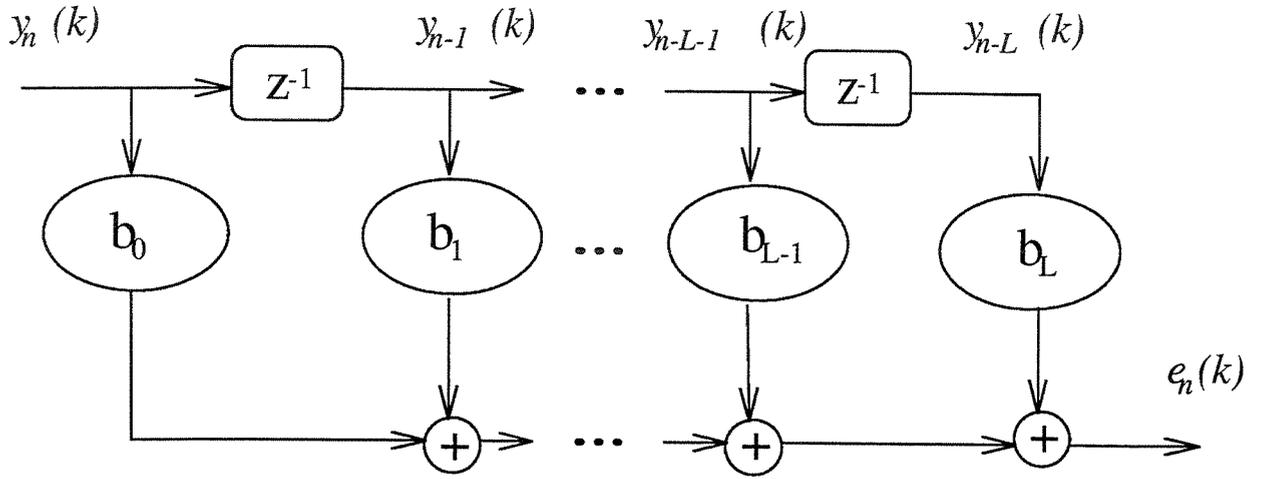


Figura 3.1: Filtro de Erro de Predição.

$$\mathbf{D}(k) \triangleq \begin{bmatrix} y_{L-1}(k) & y_{L-2}(k) & \cdots & y_1(k) & y_0(k) \\ y_L(k) & y_{L-1}(k) & \cdots & y_2(k) & y_1(k) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{N-2}(k) & y_{N-3}(k) & \cdots & y_{N-L}(k) & y_{N-L-1}(k) \end{bmatrix}_{(N-L) \times L}, \quad (3.18)$$

e o vetor dos valores preditos como $\mathbf{d}(k) = [y_L(k) \ y_{L+1}(k) \ \cdots \ y_{N-1}(k)]_{N-L}^T$, poderemos descrever o processo de filtragem através da forma matricial

$$\mathbf{D}(k)\mathbf{b}_k - \mathbf{d}(k) = \mathbf{e}(k), \text{ para cada valor de } k, \quad (3.19)$$

na qual o vetor $\mathbf{b}_k = [b_1, b_2, \dots, b_L]^T$ é composto pelos ganhos do filtro preditor para cada instante k e $\mathbf{e}(k) = [e_L(k), e_{L+1}(k), \dots, e_{N-1}(k)]^T$.

Inicialmente omitindo-se o índice k , para simplificar a notação, poderemos reescrever (3.19) como

$$\mathbf{D}\mathbf{b} - \mathbf{d} = \begin{bmatrix} \mathbf{d} \\ \mathbf{D} \end{bmatrix} \begin{bmatrix} -1 \\ \mathbf{b} \end{bmatrix} = \mathbf{D}'\mathbf{b}' = \mathbf{e}, \quad (3.20)$$

na qual \mathbf{D}' e \mathbf{b}' denotam as versões aumentadas, respectivamente, da matriz de dados \mathbf{D} e do vetor dos ganhos do filtro \mathbf{b} .

Os parâmetros b_l , $l = 1, \dots, L$, são então ajustados minimizando-se a energia do sinal de saída do filtro em (3.17), sujeito à restrição de que o peso do sensor selecionado seja negativo e igual à unidade ($b_0 = -1$). Dessa forma, podemos transformar o problema de estimação através da predição linear em um problema de otimização baseado no critério dos mínimos quadrados, o qual obtém os coeficientes do filtro de erro de predição que minimiza a energia do erro no vetor \mathbf{e} , através da minimização da seguinte função custo

$$J_{PL}(\mathbf{b}) = \{\mathbf{e}^H \mathbf{e}\} = \|\mathbf{e}\|_2^2, \text{ com } b_0 = -1. \quad (3.21)$$

Substituindo o vetor de erro dado por (3.20) em (3.21), teremos a seguinte função custo a ser minimizada

$$J_{PL}(\mathbf{b}) = \{\mathbf{b}'^H \mathbf{D}'^H \mathbf{D}' \mathbf{b}'\}. \quad (3.22)$$

Existem algumas alternativas para a composição da matriz de dados \mathbf{D} e para a estrutura do filtro preditor. O filtro denominado de progressivo (“forward”), dá origem à predição linear progressiva (FLP), que procura prever as amostras $\{y_n(k)\}_{n=L}^{N-1}$ a partir das amostras anteriores $\{y_{n-l}(k)\}_{n=L}^{N-1}$, para $l = 1, 2, \dots, L$, conforme (3.16). Já o filtro denominado de regressivo (“backward”), que dá origem à predição linear regressiva (BLP), estima as amostras $\{y_n(k)\}_{n=0}^{N-L-1}$ a partir das amostras posteriores $\{y_{n+l}(k)\}_{n=0}^{N-L-1}$, para $l = 1, 2, \dots, L$. Embora ambos os filtros utilizem as mesmas amostras finitas do sinal disponível, o resultado que minimiza o erro de predição em cada caso é distinto [Colares97].

Combinando-se as formas de predição progressiva e regressiva, obtém-se uma terceira forma denominada de progressiva-regressiva (“forward-backward”), que dá origem à predição linear progressiva-regressiva (FBLP). Nesse caso, o critério de otimização consiste na minimização da soma das energias dos erros de predição progressivo e regressivo, impondo-se como restrição uma relação adequada entre as soluções de ambos os filtros. Esse procedimento é aquele que melhor utiliza as amostras disponíveis do sinal e, conseqüentemente, apresenta melhores resultados [Colares97].

3.3.1 Estimação de Parâmetros via Predição Linear

Através de uma propriedade particular associada à estrutura do filtro preditor otimizado, constatou-se a possibilidade da utilização de tal procedimento como uma forma indireta de se estimar os parâmetros ϕ_m em (2.8), para $m = 1, 2, \dots, M$, quando os sinais $y_n(k)$, $0 \leq n \leq N-1$, são aplicados na sua entrada.

Para apresentar esta possibilidade consideraremos o polinômio de ordem L dado por

$$b(z) = b_0 z^L + b_1 z^{L-1} + \dots + b_L, \quad (3.23)$$

cujos coeficientes são os ganhos do preditor linear *otimizado* para o sinal $y_n(k)$ sem ruído, ou seja, $x_n(k)$.

Podemos, então, enunciar os teoremas a seguir.

Teorema 1: Se L satisfaz a desigualdade $M \leq L \leq N$, e se \mathbf{b} satisfaz $\mathbf{D}\mathbf{b} = \mathbf{d}$, então $b(z)$ terá M de seus zeros posicionados sobre a circunferência de raio unitário (CRU) em $z_m = \exp[j\phi_m]$, para $m = 1, \dots, M$. Estes zeros são denominados de zeros de sinal [Antunes92], [Kumaresa83b].

Dessa forma, podemos notar que os ângulos indicados pelos zeros de sinal fornecem estimativas dos parâmetros ϕ_m , para $m = 1, \dots, M$.

Teorema 2: Os $(L - M)$ zeros restantes de $b(z)$ estarão localizados no interior da CRU, caso o vetor \mathbf{b} for escolhido dentre aqueles que apresenta a menor norma Euclidiana. Estes zeros são denominados de zeros estranhos. Neste caso, o preditor será denominado de norma-mínima e, conseqüentemente, apresentará fase mínima [Antunes92], [Kumaresa83b].

Para o caso ruidoso os teoremas enunciados não são mais válidos uma vez que todos os zeros

flutuam em torno de suas posições da situação não ruidosa. Conseqüentemente, tais flutuações produzem uma variância nas estimativas obtidas.

Geralmente, a estratégia de identificação dos zeros de sinal consiste em associá-los aos M zeros mais próximos da CRU. O posicionamento dos $L - M$ zeros estranhos dependerá da solução escolhida que satisfaz $\mathbf{D}\mathbf{b} = \mathbf{d}$. Para a solução de norma-mínima, esses zeros estarão uniformemente posicionados no interior da CRU. Essa característica permite a distinção entre os dois conjuntos de zeros, mesmo em baixas relações sinal-ruído (SNR) [Colares97].

Podemos concluir, portanto, que se o vetor dos coeficiente \mathbf{b} for obtido através da otimização de um filtro preditor, as raízes de $b(z)$ fornecerão estimativas para os parâmetros DOA do sinal observado. Podemos notar que as estimativas dos coeficientes de $b(z)$ estão relacionadas através de um modelo linear enquanto que as estimativas dos parâmetros DOA, obtidas a partir da posição angular das raízes de $b(z)$, possuem uma relação não-linear.

Apesar da capacidade de estimação de parâmetro da predição linear ter sido apresentada considerando um único instante de tempo k , a mesma continua válida para um sinal espacial proveniente de um arranjo de sensores, considerando vários instantes de amostragem, $k = 1, 2, \dots, K$, que é o caso denominado de múltiplo experimentos.

A extensão do uso da predição linear para a estimação do parâmetro DOA no caso de dispormos de múltiplos experimentos (“snapshots”), leva a várias possibilidades para formação da matriz de dados, que em geral acarreta em um aumento do esforço computacional, apesar de apresentar a possibilidade de melhor desempenho que no caso de um único experimento.

3.3.2 Predição Linear Baseada em Subespaços

A predição linear baseada em subespaços consiste na utilização da forma decomposta da matriz de dados \mathbf{D} em (3.18). Essa decomposição, por sua vez, consiste em descrever a matriz através de seus valores e vetores singulares e é denominada de decomposição em valores singulares (SVD). Dessa forma torna-se possível descrever a equação normal em (3.19) através dos subespaços gerados pela matriz de dados. Para esse fim, enunciaremos o teorema a seguir.

Teorema 3: Toda matriz \mathbf{D} $m \times n$ pode ser fatorada da forma $\mathbf{D} = \mathbf{U}_D \mathbf{\Sigma}_D \mathbf{V}_D^H$, onde \mathbf{U}_D possui dimensões $m \times m$ e satisfaz $\mathbf{U}_D \mathbf{U}_D^H = \mathbf{U}_D^H \mathbf{U}_D = \mathbf{I}$; \mathbf{V}_D possui dimensões $n \times n$ e satisfaz $\mathbf{V}_D \mathbf{V}_D^H = \mathbf{V}_D^H \mathbf{V}_D = \mathbf{I}$; e $\mathbf{\Sigma}_D$ é uma matriz diagonal de dimensões $m \times n$ composta por elementos reais. Tais elementos reais são denominados de valores singulares e são ordenados tal que $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p > 0$, com $p \leq \min(m, n)$, onde p é o posto de \mathbf{D} . As colunas de \mathbf{U}_D e \mathbf{V}_D , denotadas por \mathbf{u}_i^D e \mathbf{v}_i^D , são denominadas, respectivamente, de vetores singulares à esquerda e à direita de \mathbf{D} [Colares97], [Golub89], [Lemos97].

Desta forma, a SVD decompõe a matriz \mathbf{D} em dois conjuntos de vetores ortogonais: um conjunto \mathbf{U}_D gerador do espaço das colunas de \mathbf{D} , e um outro conjunto \mathbf{V}_D gerador do espaço das linhas de \mathbf{D} . Os valores singulares, por sua vez, constituem os módulos das projeções de \mathbf{D} na direção definida por \mathbf{u}_i^D e \mathbf{v}_i^D [Golub89].

Analisando a SVD de \mathbf{D} , podemos escrever

$$\mathbf{D} = \mathbf{U}_D \mathbf{\Sigma}_D \mathbf{V}_D^H = \mathbf{U}_D \begin{bmatrix} \mathbf{\Sigma}_1^D & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{V}_D^H, \quad (3.24)$$

onde Σ_1^D é uma matriz diagonal de dimensões $p \times p$, onde p é o posto de \mathbf{D} .

Explicitando os termos de (3.24), podemos escrever

$$\mathbf{D} = \sum_{i=1}^p \sigma_i \mathbf{u}_i^D (\mathbf{v}_i^D)^H. \quad (3.25)$$

Assim, podemos definir a *pseudo-inversa* da matriz \mathbf{D} como [Golub89]

$$\mathbf{D}^\dagger = \mathbf{V}_D \begin{bmatrix} (\Sigma_1^D)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{U}_D^H = \sum_{i=1}^p \frac{1}{\sigma_i} \mathbf{v}_i^D (\mathbf{u}_i^D)^H. \quad (3.26)$$

Podemos, então, apresentar o problema de estimação através da predição linear baseada em subespaços como a minimização da função custo em (3.22) com a substituição da matriz \mathbf{D} pela sua versão decomposta, obtendo

$$J_{PL}(\mathbf{b}) = \left\{ \begin{bmatrix} -1 \\ \mathbf{b} \end{bmatrix}^H \left[\mathbf{V}_D \Sigma_D \mathbf{U}_D^H : \mathbf{d} \right]^H \left[\mathbf{V}_D \Sigma_D \mathbf{U}_D^H : \mathbf{d} \right] \begin{bmatrix} -1 \\ \mathbf{b} \end{bmatrix} \right\}. \quad (3.27)$$

Pode ser demonstrado que a solução para o problema de otimização em (3.27), baseada no critério dos mínimos quadrados, pode ser obtida através do uso da *pseudo-inversa* de \mathbf{D} , obtendo como estimativa para \mathbf{b}

$$\hat{\mathbf{b}} = \mathbf{D}^\dagger \mathbf{d}.$$

Indo além na decomposição da matriz \mathbf{D} , podemos considerar, ainda, as seguintes partições na SVD de \mathbf{D}

$$\mathbf{D} = \begin{bmatrix} \mathbf{U}_1^D & \mathbf{U}_2^D \end{bmatrix} \cdot \begin{bmatrix} \Sigma_1^D & \mathbf{0} \\ \mathbf{0} & \Sigma_2^D \end{bmatrix} \cdot \begin{bmatrix} \mathbf{V}_1^D & \mathbf{V}_2^D \end{bmatrix}^H, \quad (3.28)$$

onde \mathbf{U}_1^D é uma partição de \mathbf{U}_D contendo suas p primeiras colunas e \mathbf{U}_2^D suas $m - p$ últimas colunas; \mathbf{V}_1^D é a partição de \mathbf{V}_D contendo as suas p primeiras linhas e \mathbf{V}_2^D as $n - p$ linhas restantes; Σ_2^D é uma matriz retangular, com dimensões $(m-p) \times (n-p)$, contendo os $\min(m-p, n-p)$ últimos elementos ordenados de Σ_D .

Após as definições e considerações anteriores podemos inferir sobre o processo de predição linear considerando os subespaços definidos pela matriz de dados.

No caso de um sinal composto por M exponenciais complexas independentes e sem ruído, o posto de \mathbf{D} será $p = M$ e Σ_2^D será nula. Portanto, existem M valores singulares, não nulos, associados aos vetores singulares de \mathbf{U}_1^D e \mathbf{V}_1^D , denominados, respectivamente, de valores e vetores singulares de sinal.

No caso de um sinal ruidoso, os valores singulares presentes em Σ_2^D , anteriormente nulos, tornam-se positivos e com magnitude crescente com a redução da relação sinal-ruído (SNR). Portanto, para um filtro preditor de ordem L , teremos os $L - M$ menores valores singulares de \mathbf{D} em Σ_2^D e seus correspondentes vetores singulares de \mathbf{U}_2^D e \mathbf{V}_2^D associados exclusivamente à porção ruidosa do sinal, sendo denominados, respectivamente, de valores e vetores singulares de ruído.

Podemos, portanto, dividir o espaço vetorial da matriz de dados \mathbf{D} em dois subespaços: subespaço de sinal e subespaço de ruído. Na verdade, tal denominação é imprecisa visto que o

subespaço de sinal, assim definido, também está contaminado por ruído, sendo por essa razão também denominado de subespaço de sinal mais ruído. Tais subespaços, apesar de estarem relacionados com aqueles associados à matriz de correlação \mathbf{R} , não devem ser confundidos e por esse motivo serão denominados a partir de então de subespaços de sinal e de ruído associados à matriz de dados.

Conhecendo-se os subespaços de sinal e de ruído associados à matriz de dados é possível diminuir a influência do ruído na escolha da solução ótima para o preditor através da sua extração. Portanto, a distinção entre esses subespaços forma a base fundamental dos procedimentos de predição linear baseados em subespaços.

3.4 Reparametrizando a Predição Linear

Para permitir relacionar os procedimentos de estimação baseados no critério ML com os procedimentos que utilizam a predição linear, apresentaremos a seguir uma reparametrização adequada para a predição linear.

Construindo-se a matriz *Toeplitz* \mathbf{B} formada pelos coeficientes $[b_0, b_1, \dots, b_L]$ da forma a seguir

$$\mathbf{B}^H = \begin{bmatrix} b_L & b_{L-1} & \cdots & b_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & b_L & b_{L-1} & \cdots & b_0 \end{bmatrix}_{(N-L) \times N}, \quad (3.29)$$

com $b_0 = -1$ e através da analogia com a convolução entre dois sinais, podemos reconhecer a igualdade

$$\mathbf{B}^H \mathbf{y}(k) = \mathbf{D}' \mathbf{b}', \quad (3.30)$$

na qual $\mathbf{D}' \mathbf{b}'$ é dado em (3.20) e o vetor $\mathbf{y}(k)$ é dado pelas colunas da matriz \mathbf{D}' .

Manipulando a função custo para a predição linear em (3.22) obtemos

$$\begin{aligned} J_{PL}(\mathbf{B}) &= \left\{ \mathbf{y}^H(k) \mathbf{B} \mathbf{B}^H \mathbf{y}(k) \right\} = Tr \left\{ \mathbf{B} \mathbf{B}^H \left[\mathbf{y}(k) \mathbf{y}^H(k) \right] \right\} \\ &= Tr \left\{ \mathbf{B} \mathbf{B}^H \hat{\mathbf{R}}(k) \right\}, \end{aligned} \quad (3.31)$$

na qual utilizamos a igualdade $\hat{\mathbf{R}} = \frac{1}{K} \sum_{k=1}^K \mathbf{y}(k) \mathbf{y}^H(k) = \frac{1}{K} \sum_{k=1}^K \hat{\mathbf{R}}(k)$.

Após esta reparametrização, podemos interpretar a predição linear como a minimização da função custo em (3.31), para um dado $\hat{\mathbf{R}}(k)$. Uma vez obtido $\hat{\mathbf{B}}$ e, conseqüentemente, $\hat{\mathbf{b}}'$, as estimativas $\hat{\phi}_m$, para $m = 1, 2, \dots, M$, são então determinadas através das raízes de $b(z)$.

A minimização da função custo em (3.31) consiste em um problema de otimização de segunda ordem, notadamente mais simples que o enfrentado pelo procedimento de estimação através do critério ML. Essa característica torna a estimação do parâmetro DOA via predição linear uma alternativa para o critério ML. A sua aplicação se restringe ao uso de arranjos lineares e uniformes, em um cenário no qual os sinais possuem potências semelhantes, são levemente correlacionados com moderadas SNR [Krim96], [Tufts82].

3.5 Reparametrizando o Estimador ML

Uma alternativa ao procedimento de minimização para o estimador ML determinístico, em (3.5), surge da idéia de uma reparametrização da matriz de projeção $\Pi_{\mathbf{A}}^{\perp}$ através de bases do *espaço nulo* da matriz \mathbf{A}^H , definido pelas suas linhas [Bresler86].

Tomando-se as matrizes \mathbf{A} e \mathbf{B}^H , definidas, respectivamente, em (2.17) e (3.29), e considerando o **Teorema 1**, podemos notar que o espaço gerado pelas linhas de \mathbf{B}^H , de dimensão $N - L$, é ortogonal ao espaço gerado pelas colunas de \mathbf{A} , de dimensão M , e satisfaz a seguinte relação para um sinal livre de ruído

$$\mathbf{B}^H \mathbf{A} = \mathbf{A}^H \mathbf{B} = \mathbf{0}. \quad (3.32)$$

Para $L = M$, se \mathbf{B}^H tem posto cheio e igual a $N - M$ e \mathbf{A} possui posto cheio e igual a M , as linhas de \mathbf{B}^H formarão uma base para o espaço das colunas de \mathbf{A} , denominado de *espaço nulo* de \mathbf{A}^H . Desta forma, a projeção ortogonal desses subespaços deverão coincidir, implicando em

$$\Pi_{\mathbf{B}} = \mathbf{B}(\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H = \Pi_{\mathbf{A}}^{\perp} \quad (3.33)$$

Como conseqüência da igualdade em (3.33), o critério ML pode ser reparametrizado através dos coeficientes b_l , $l = 1, 2, \dots, M$. As estimativas dos coeficientes podem ser, então, calculadas minimizando-se a função custo

$$J_{ML}(\mathbf{b}') = \text{Tr} \left\{ \left[\mathbf{B}(\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H \right] \hat{\mathbf{R}} \right\}, \text{ com } L = M, \quad (3.34)$$

a qual é equivalente à do método ML em (3.5) se utilizarmos a relação (3.33).

Apesar da minimização de (3.34) ser um problema de otimização não-linear, de quarta ordem, tão complicado quanto aquele encontrado para o estimador ML clássico, existe a possibilidade de uma solução iterativa através de um problema de minimização de segunda ordem, dando origem a novos métodos [Bresler86], [Stoica90a].

3.6 Estimação Totalmente Baseada em Subespaços

Conforme foi apresentado, o procedimento de estimação baseado na predição linear pode fazer uso dos subespaços definidos a partir da matriz de dados \mathbf{D} , enquanto que o critério ML pode fazer uso dos subespaços associados à matriz de correlação \mathbf{R} . Dessa forma, torna-se possível a extração da porção ruidosa na composição de suas funções custo a serem minimizadas, geralmente propiciando melhores resultados.

Existe, porém, procedimentos de estimação baseados em subespaços que não possuem origem nem no conceito de predição ou no critério ML. Tais procedimentos, denominados aqui de estimação totalmente baseada em subespaços, consistem no uso de propriedades existentes entre os subespaço de sinal e subespaço de ruído da matriz \mathbf{R} .

Alguns textos na literatura apresentam os métodos baseados em subespaços como propostas concorrentes com os métodos paramétricos. As técnicas tipo “beamforming”, por exemplo, fazem uso de subespaços em sua concepção e no entanto são técnicas espectrais. Concluímos, portanto, que os procedimentos de estimação baseados em subespaços podem ser utilizados

como variações de técnicas paramétricas ou não-paramétricas (espectrais). Apresentaremos, aqui, tais procedimentos apenas no contexto das técnicas paramétricas de estimação.

O primeiro trabalho publicado que fez uso dos subespaços foi o de Pisarenko em 1973, no qual foi feito uso da matriz de correlação no problema clássico de recuperação harmônica. Porém, foi com o algoritmo MUSIC (“Multiple Signal Classification”) que tal abordagem despertou maior interesse [Krim96], [Schmidt86].

A estimação dos parâmetros do sinal através da propriedade de invariância à rotação apresentada pelos subespaços surgiu com o método ESPRIT, o qual explora uma estrutura específica de arranjo [Roy89]. Métodos baseados em outras propriedades dos subespaços e o uso de reparametrizações também surgiram após o ESPRIT [Krim96]. Podemos afirmar genericamente que tais métodos fazem uso de uma classe de estimadores baseados no critério de “distância” entre os subespaços (“subspace fitting”) [Jansson99], [Viberg91a], [Ottersten92].

3.6.1 O Critério de Ortogonalidade entre Subespaços

Considerando a matriz de correlação estatística de um sinal contaminado por ruído do tipo branco, sabemos que os autovetores associados ao subespaço de ruído, dados pelas colunas de \mathbf{U}_n , são ortogonais ao subespaço associado ao sinal [Krim96], [Stoica91], [Pisarenko73].

Considerando a hipótese de posto cheio para a matriz \mathbf{P} , conforme estabelecido no Capítulo 2, as colunas de $\mathbf{A}(\phi_o)$ são ortogonais aos autovetores do subespaço de ruído de \mathbf{R} . Conseqüentemente, suas colunas geram o subespaço de sinal de \mathbf{R} . Tomando-se o subespaço de sinal como aquele gerado pelas colunas da matriz $\mathbf{A}(\phi_o)$, teremos a validade da seguinte relação

$$\mathbf{U}_n^H \mathbf{a}(\phi_m) = 0, \text{ para } m = 1, 2, \dots, M. \quad (3.35)$$

A relação em (3.35) pode ser, então, utilizada para estimar os parâmetros ϕ_m , para $m = 1, 2, \dots, M$, desde que \mathbf{U}_n^H possa ser corretamente estimado. Neste caso, o procedimento de estimação que considera a relação (3.35) de forma aproximada ($\hat{\mathbf{U}}_n^H \mathbf{a}(\phi_m) \approx 0$), no contexto de uma otimização baseada nos mínimos-quadrados, consistirá no seguinte problema de minimização

$$\min_{\phi_m} \left\| \hat{\mathbf{U}}_n^H \mathbf{a}(\phi_m) \right\|^2, \text{ para } m = 1, 2, \dots, M.$$

Para garantir a unicidade da estimativa, consideraremos que qualquer coleção de M vetores de direção, correspondente às diferentes direções, formam um conjunto linearmente independente. Neste caso, dado $\mathbf{R} = \mathbf{A}\mathbf{P}\mathbf{A}^H + \sigma^2\mathbf{I}$, se $\mathbf{a}(\phi_m)$ satisfaz essa condição e \mathbf{P} possui posto cheio, então $\mathbf{A}\mathbf{P}\mathbf{A}^H$ terá posto igual a N . Dessa forma, a única solução que satisfaz a relação (3.35) será dada pelos parâmetros ótimos ϕ_1, \dots, ϕ_M , os quais podem ser usados para obter exatamente os parâmetros DOA.

A grande desvantagem do uso da relação em (3.35) é o fato desta não ser satisfeita no caso de sinais coerentes, quando teremos uma redução no posto de \mathbf{P} [Krim92]. Para superar essa desvantagem, uma matriz de ponderações geralmente é introduzida na função custo a ser minimizada, podendo assim controlar a participação de cada autovetor de \mathbf{U}_n , no intuito de compensar a correlação existente entre os sinais [Kaveh86], [Kumaresa83a], [Reddi79], [Stoica90d]. Outras técnicas, baseadas no pré-processamento do sinal através de filtragem

espacial, também são freqüentemente usadas para descorrelacionar os sinais nessas situações [Buckley90], [Krim96].

Um procedimento de estimação que considera a relação em (3.35) para todas as colunas de \mathbf{A} , simultaneamente, consiste no seguinte problema de otimização de segunda ordem

$$\min_{\phi} \left\{ \sum_{m=1}^M \left\| \hat{\mathbf{U}}_n^H \mathbf{a}(\phi_m) \right\|_2^2 \right\} = \min_{\phi} \left\{ \left\| \hat{\mathbf{U}}_n^H \mathbf{A}(\phi) \right\|_F^2 \right\}, \text{ com } \phi \in \mathcal{R}^M, \quad (3.36)$$

na qual o subscrito F denota a norma de *Frobenius* [Golub89]. O problema em (3.36) exige uma busca M -dimensional dos parâmetros.

Desenvolvendo o produto em (3.36), teremos o seguinte problema equivalente

$$\min_{\phi} Tr \left\{ \hat{\Pi}_{\mathbf{A}}^{\perp} \left(\mathbf{A}(\phi) \mathbf{A}^H(\phi) \right) \right\}, \text{ com } \phi \in \mathcal{R}^M, \quad (3.37)$$

no qual reconhecemos a igualdade $\hat{\Pi}_{\mathbf{A}}^{\perp} = \hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H$.

É interessante observar que a estimação em (3.37) se dá em relação a \mathbf{A} e não em relação a $\hat{\Pi}_{\mathbf{A}}^{\perp}$, como no caso ML. Por esse motivo, $\hat{\Pi}_{\mathbf{A}}^{\perp}$ é estimado através do subespaço de ruído.

Genericamente, o problema em (3.37) tanto permite a estimação dos parâmetros DOA a partir de uma estimativa do subespaço de ruído, como também através de qualquer outro projetor nesse subespaço.

Considerando a reparametrização em (3.33) e fazendo $\hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H = \hat{\Pi}_{\mathbf{B}}$, a estimativa do projetor ortogonal poderá ser obtida a partir de uma estimativa dos parâmetros \mathbf{b} da predição linear. Além disso, uma outra variação para o problema em (3.37) consiste na utilização de uma estimativa do subespaço de sinal ao invés do subespaço de ruído, na qual consideramos $\hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H = \mathbf{I} - \hat{\mathbf{U}}_s \hat{\mathbf{U}}_s^H$.

Podemos concluir que a qualidade das estimativas obtidas através desse procedimento dependerá diretamente da qualidade das estimativas dos subespaços envolvidos. Tal abordagem exige, portanto, estimativas dos subespaços através de um processo de decomposição da matriz de correlação, que geralmente, requer um grande esforço computacional. Por esse motivo, a escolha de um ou outro subespaço para a implementação desse procedimento dependerá das dimensões envolvidas.

3.6.2 O Critério de “Distância” entre Subespaços (“Subspace Fitting”)

A partir dos procedimentos baseados no critério de ortogonalidade entre os subespaços, surgiu uma nova classe de estimadores baseados no critério da “distância” entre subespaços (“Subspace Fitting”), que apresentam o mesmo desempenho do critério ML, porém um esforço computacional menor [Krim96], [Stoica90b], [Viberg91b], [Ottersten92].

Apresentaremos essa classe de estimadores através de uma reparametrização adequada do critério ML.

Comparando (2.18) com (2.25) e lembrando que $\mathbf{\Lambda}_n = \sigma^2 \mathbf{I}$, teremos

$$\mathbf{A}\mathbf{P}\mathbf{A}^H + \sigma^2\mathbf{I} = \mathbf{U}_s\mathbf{\Lambda}_s\mathbf{U}_s^H + \sigma^2\mathbf{U}_n\mathbf{U}_n^H. \quad (3.38)$$

Pós-multiplicando por \mathbf{U}_s pela direita e reagrupando os termos, teremos a relação

$$\mathbf{U}_s = \mathbf{A}(\phi_o)\mathbf{T}, \quad (3.39)$$

na qual $\mathbf{T} = \mathbf{P}\mathbf{A}^H\mathbf{U}_s(\mathbf{\Lambda}_s - \sigma^2\mathbf{I})^{-1}$ é uma matriz de posto cheio. Visto que os parâmetros ótimos ϕ_m , para $m = 1, \dots, M$, e \mathbf{T} são desconhecidos, é natural procurar seus valores que satisfazem (3.39). Se uma estimativa $\hat{\mathbf{U}}_s$ de \mathbf{U}_s for utilizada, tal solução não existirá. Neste caso, deveremos minimizar alguma “distância” entre $\hat{\mathbf{U}}_s$ e $\mathbf{A}\mathbf{T}$. Para esse propósito a norma de *Frobenius* será utilizada.

As estimativas serão, então, obtidas resolvendo o seguinte problema de otimização não-linear

$$\{\hat{\phi}, \hat{\mathbf{T}}\} = \arg \left\{ \min_{\phi, \mathbf{T}} \left\| \hat{\mathbf{U}}_s - \mathbf{A}\mathbf{T} \right\|_F^2 \right\}, \quad (3.40)$$

no qual $\arg \{\bullet\}$ indica o valor do *argumento* resultante da otimização.

Similar ao procedimento utilizado para o método ML, esse é um problema de mínimos quadrados não-linear separável, cuja solução para o parâmetro linear $\hat{\mathbf{T}}$, com \mathbf{A} fixo, é dada por

$$\hat{\mathbf{T}} = \mathbf{A}^\dagger \hat{\mathbf{U}}_s. \quad (3.41)$$

A estimativa de \mathbf{T} em (3.41) foi escolhida como aquela que minimiza o erro de estimação ao quadrado [Golub89]. Substituindo (3.41) em (3.40) e utilizando a relação $\|\mathbf{M}\|_F^2 = \text{Tr} \{ \mathbf{M}\mathbf{M}^H \}$, obtemos a seguinte função custo a ser minimizada

$$J_{SSF}(\phi) = \text{Tr} \left\{ \mathbf{\Pi}_\mathbf{A}^\perp \hat{\mathbf{U}}_s \hat{\mathbf{U}}_s^H \right\}. \quad (3.42)$$

Uma variação para o problema em (3.42) pode surgir se utilizarmos estimativas do subespaço de ruído em sua formulação, dando origem ao seguinte problema equivalente [Krim96]

$$J_{NSF}(\phi) = \text{Tr} \left\{ \mathbf{\Pi}_\mathbf{A} \hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H \right\}. \quad (3.43)$$

Devemos lembrar que as matrizes \mathbf{A} e \mathbf{U}_s geram o mesmo espaço quando \mathbf{P} tem posto cheio. Em um caso geral, quando o posto de \mathbf{P} é igual a $M' < M$, a matriz \mathbf{U}_s gera um subespaço M' -dimensional da matriz \mathbf{A} . Portanto, nas situações em que \mathbf{P} não possui posto cheio, uma matriz positiva definida de ponderação pode ser adicionada ao problema [Krim96]. A utilização de uma matriz de ponderação permite favorecer a escolha dos M' autovetores em $\hat{\mathbf{U}}_s$ mais representativos do sinal, reduzindo assim a dimensão do problema original.

A adição de uma matriz de ponderação \mathbf{W} no problema em (3.42) leva à seguinte função custo [Krim96], [Viberg91b].

$$J_{WSF}(\phi) = \text{Tr} \left\{ \mathbf{\Pi}_\mathbf{A}^\perp \hat{\mathbf{U}}_s \mathbf{W} \hat{\mathbf{U}}_s^H \right\}, \quad (3.44)$$

na qual surge a questão de como escolher a matriz de ponderação \mathbf{W} que minimiza a variância do erro de estimação.

Para $\mathbf{W} = \mathbf{I}$, podemos observar que (3.44) coincide com a função em (3.42). Esta é uma boa escolha para \mathbf{W} apenas quando \mathbf{P} é uma matriz diagonal com posto cheio. Essa situação compreende um cenário no qual os sinais envolvidos não são correlacionados.

Por outro lado, escolhendo $\mathbf{W} = \hat{\mathbf{\Lambda}}_s$ em (3.44) podemos interpretar o resultado obtido em (3.44) como uma aproximação baseada em subespaço aplicada à matriz de correlação estimada, na qual tomamos a aproximação $\hat{\mathbf{R}} \cong \hat{\mathbf{U}}_s \hat{\mathbf{\Lambda}}_s \hat{\mathbf{U}}_s^H$. Tal procedimento traz melhores resultados, desde que os subespaços sejam corretamente estimados. Nesse caso, podemos ainda notar que o problema em (3.44) é equivalente àquele obtido pelo critério ML determinístico baseado em subespaços, dado em (3.15). Essa constatação faz parte das contribuições desse trabalho.

A função custo em (3.44) generaliza, portanto, os métodos paramétricos baseados em subespaços, assim como (3.5) generaliza os métodos baseados no critério ML.

Devemos enfatizar que os procedimentos baseados em subespaços também podem ser reparametrizados através da substituição do projetor $\Pi_{\mathbf{A}}^\perp$ pelo projetor equivalente dado por $\Pi_{\mathbf{B}}$.

3.7 Quadro Resumo

Máxima-Verossimilhança:

$$\begin{array}{l} \min_{\phi} Tr \left\{ \Pi_{\mathbf{A}}^\perp(\phi) \hat{\mathbf{R}} \right\} \\ \min_{\phi} Tr \left\{ \left[\mathbf{B}(\phi) (\mathbf{B}^H(\phi) \mathbf{B}(\phi))^{-1} \mathbf{B}^H(\phi) \right] \hat{\mathbf{R}} \right\}, L = M. \\ \min_{\phi} Tr \left\{ \Pi_{\mathbf{A}}^\perp(\phi) \left(\hat{\mathbf{U}}_s \hat{\mathbf{\Lambda}}_s \hat{\mathbf{U}}_s^H \right) \right\} \end{array}$$

Predição Linear:

$$\begin{array}{l} \min_{\phi} Tr \left\{ \mathbf{B}(\phi) \mathbf{B}^H(\phi) \left[\mathbf{y}(k) \mathbf{y}^H(k) \right] \right\} \\ \min_{\phi} Tr \left\{ \mathbf{B}(\phi) \mathbf{B}^H(\phi) \hat{\mathbf{R}}(k) \right\}, L \geq M \text{ e } b_0 = -1. \end{array}$$

Subespaços - Ortogonalidade:

$$\min_{\phi} Tr \left\{ \hat{\Pi}_{\mathbf{A}}^\perp \left(\mathbf{A}(\phi) \mathbf{A}^H(\phi) \right) \right\}, L \geq M.$$

Subespaços - “Distância”:

$$\min_{\phi} Tr \left\{ \Pi_{\mathbf{A}}^{\perp}(\phi) \hat{\mathbf{U}}_s \hat{\mathbf{U}}_s^H \right\}, L = M.$$

3.8 Conclusão

Podemos concluir que, nas duas formulações apresentadas para o critério ML, a obtenção das estimativas do parâmetro DOA exige a solução de um problema de minimização quadrática e não-linear em relação ao termo $\Pi_{\mathbf{A}}^{\perp}$. A solução desse problema é obtida através de uma busca multi-dimensional, que, em geral, exige um grande esforço computacional. Ainda assim, tal procedimento não garante a convergência para o mínimo global. Se dispusermos de um bom valor inicial para as estimativas, técnicas como a de Gauss-Newton podem ser utilizadas levando, normalmente, a uma rápida convergência [Cadzow90], [Ottersten93].

A desvantagem por enfrentar problemas de convergência, bem como o alto esforço computacional requerido pelo procedimento ML, torna impraticável sua aplicação em alguns casos, apesar de sua maior precisão.

Tais constatações levam à busca natural por procedimentos alternativos que embora não apresentem a mesma precisão do critério ML, possuem uma maior simplicidade na sua implementação. A predição linear faz parte de uma dessas alternativas.

Uma abordagem baseada em subespaços para a predição linear e para o critério ML também foi apresentada. Nessa abordagem surge a necessidade de se estimar os subespaços definidos pelo sinal e ruído. No caso do critério ML os subespaços são obtidos a partir da decomposição da matriz de correlação \mathbf{R} em autovalores e autovetores, enquanto que na predição linear tais subespaços são obtidos a partir da decomposição em valores e vetores singulares da matriz de dados \mathbf{D} . Tais procedimentos, embora exijam um esforço computacional adicional em relação à abordagem clássica, fornecem melhores estimativas para sinais ruidosos, desde que a distinção entre os dois subespaços seja possível.

Uma outra classe de procedimentos alternativos concentra-se na redução do esforço computacional do procedimento ML, através de manipulações (reparametrizações) do problema original, levando a um problema mais simples, que pode ser resolvido de forma iterativa.

O procedimento de reparametrização apresentado tanto para a predição linear como para o critério ML, permite uma formulação única para ambos, sendo, portanto, uma das principais contribuições deste capítulo. Tal unificação permitirá, em capítulos seguintes, estabelecer relações entre os mesmos quanto ao esforço computacional e desempenho. No momento, comparando as funções custos obtidas para ambos os problemas, podemos concluir que o procedimento de predição linear leva a um problema de minimização de segunda ordem, enquanto que o critério ML leva a uma minimização de quarta ordem. Os procedimentos clássicos para obtenção da solução de tais problemas de minimização, bem como suas variações, serão abordados nos capítulos seguintes, trazendo novas possibilidades de comparações.

Apesar da predição linear ter sido colocada como uma alternativa concorrente com o critério ML, observamos que o conceito de predição linear aplicado através de uma reparametrização do critério ML, possibilita implementação mais eficiente deste último, dando origem a métodos promissores.

Procedimentos totalmente baseados em subespaços também foram apresentados como uma alternativa ao critério ML. Dentre esses procedimentos destacamos aqueles que utilizam a propriedade de ortogonalidade entre os subespaços de sinal e de ruído ou baseiam-se em um critério de “distância” entre um subespaço e sua aproximação. Estes dois também foram interpretados através de uma reparametrização adequada do critério ML.

Concluimos, ainda, que a relação em (3.39) forma a base para os procedimentos baseados no critério da “distância” entre subespaços (“Subspace Fitting”), dando origem ao problema de estimação em (3.42). Variações para essa abordagem consistem no uso desse estimador descrito através do subespaço de ruído (NSF), como em (3.43), e na introdução de matrizes de ponderações (WSF) como em (3.44).

Capítulo 4

Otimização Linear Quadrática sem Restrições

4.1 Introdução

Os procedimentos para estimação do parâmetro DOA, apresentados no capítulo anterior, possuem vários pontos em comum em suas formulações. Um deles consiste na obtenção de uma função custo que deve ser minimizada para a obtenção dos *estimadores/estimativas* finais. Embora as funções custos em cada caso sejam originadas por propriedades distintas, uma análise baseada no conceito de subespaços, juntamente com uma reparametrização adequada do problema original, permitiu uma visão unificada de seus objetivos e formulações.

O objetivo deste capítulo é apresentar algumas técnicas de otimização que são empregadas nos métodos para estimação dos parâmetros DOA, a serem apresentados posteriormente.

Os procedimentos de otimização apresentados aqui são comumente utilizados em problemas de estimação *linear* de parâmetros. Tais procedimentos não estão restritos à minimização de funções custo de segunda ordem, sendo também úteis na solução de problemas de ordens superiores, através de procedimentos iterativos, ou na obtenção de parâmetros com relações não-lineares, por meio de uma reparametrização adequada. No próximo capítulo trataremos das soluções com restrições inspiradas no modelo do sinal.

O problema básico que abordaremos consiste em determinar estimativas dos parâmetros desconhecidos de um modelo linear para o sinal observado da forma

$$b_1 y_1 + \cdots + b_n y_n = c, \quad (4.1)$$

no qual y_1, \dots, y_n e c denotam as variáveis observadas, e $b_1, \dots, b_n \in \mathbb{C}^n$ fazem o papel dos parâmetros que caracterizam o sistema em observação. Para um conjunto de m equações semelhantes a (4.1) com $m > n$, e considerando que as variáveis observadas possuem erros, teremos

$$\mathbf{W}\mathbf{b} \approx \mathbf{w}, \quad (4.2)$$

onde cada linha da matriz de dados \mathbf{W} e do vetor de observações \mathbf{w} contém uma medida das variáveis y_1, \dots, y_n e c , respectivamente. O vetor \mathbf{b} , por sua vez, é definido por $\mathbf{b} = [b_1, b_2, \dots, b_n]^T$.

Os procedimentos de otimização pretendem, portanto, obter as estimativas dos parâmetros em \mathbf{b} que melhor satisfazem a equação (4.2), para o conjunto de observações dadas em \mathbf{W} e \mathbf{w} .

Historicamente, o critério de otimização mais popular é aquele que minimiza o quadrado do erro de estimação, denominado de mínimos-quadrados ou do inglês “Least-squares” (LS) [Sorenson70]. Mais recentemente foi proposto um refinamento do critério dos mínimos-quadrados, denominado de mínimos-quadrados total ou do inglês “Total Least-Squares” (TLS) [Golub80]. O critério TLS difere do LS na hipótese da presença de erros nos dados disponíveis, conforme apresentaremos a seguir.

4.2 Mínimos Quadrados (LS)

O problema de minimização no contexto LS se caracteriza por considerar as observações das variáveis y_1, \dots, y_n livres de erros, sendo todo erro confinado no vetor de observações \mathbf{w} . Apesar dessa hipótese ser pouco realista, na existência de várias possibilidades de imprecisão na medida dos dados, leva a uma simplificação do problema original.

Sendo mais específico, se definirmos, a partir de (4.2), um vetor de erro \mathbf{e} da forma

$$\mathbf{e} = \mathbf{W}\mathbf{b} - \mathbf{w}, \quad (4.3)$$

o critério LS consistirá em obter um vetor solução \mathbf{b}_{LS} que minimiza a norma desse vetor de erro \mathbf{e} , ou seja

$$\mathbf{b}_{LS} = \arg \left\{ \min \left\| \mathbf{W}\mathbf{b} - \mathbf{w} \right\|_2^2 \right\}; \quad \mathbf{W} \in C^{m \times n}, \quad \mathbf{w} \in C^m. \quad (4.4)$$

Dessa forma, podemos dizer que o critério LS procura encontrar uma solução ótima para o problema de minimização de segunda-ordem em (4.4), através de um conjunto de equações não-homôneas em (4.3).

Naturalmente, para computar o vetor solução que satisfaz (4.4) devemos considerar

$$\frac{\partial \left\| \mathbf{e} \right\|_2^2}{\partial \mathbf{b}} = 0. \quad (4.5)$$

Tomando-se $\left\| \mathbf{e} \right\|_2^2 = \mathbf{e}^H \mathbf{e}$, teremos

$$\left\| \mathbf{e} \right\|_2^2 = \mathbf{w}^H \mathbf{w} + \mathbf{b}^H \mathbf{W}^H \mathbf{W} \mathbf{b} - 2\mathbf{b}^H \mathbf{W}^H \mathbf{w}. \quad (4.6)$$

Aplicando a derivada e igualando a zero, constatamos que a solução LS deve satisfazer a equação normal

$$\mathbf{W}^H \mathbf{W} \mathbf{b}_{LS} = \mathbf{W}^H \mathbf{w}. \quad (4.7)$$

A equação em (4.7) será consistente se no caso de dados sem ruído tivermos (4.3) igual a zero, com \mathbf{b} dado por \mathbf{b}_{LS} de (4.7). Para tanto, $\mathbf{W}^H \mathbf{W}$ deverá possuir características específicas [VanHuffel91].

De uma outra forma, reescrevendo a equação normal em (4.7) como

$$\mathbf{W}^H [\mathbf{W} \mathbf{b}_{LS} - \mathbf{w}] = \mathbf{0}, \quad (4.8)$$

com $\mathbf{W}\mathbf{b}_{LS} - \mathbf{w} = \mathbf{e}_{LS}$, podemos constatar que o vetor \mathbf{e}_{LS} é ortogonal às colunas de \mathbf{W} , sendo portanto ortogonal aos dados observados. Desta forma, \mathbf{b}_{LS} resolve o problema (4.4) se e somente se satisfizer (4.8).

Para a solução da equação normal em (4.7) devemos fazer algumas considerações quanto ao posto da matriz $[\mathbf{W}]_{m \times n}$, para $m > n$.

I) Se $\text{Posto}(\mathbf{W}) = n$, $(\mathbf{W}^H \mathbf{W})^{-1}$ existe e então (4.2) possui uma única solução, dada por

$$\mathbf{b}_{LS} = (\mathbf{W}^H \mathbf{W})^{-1} \mathbf{W}^H \mathbf{w}. \quad (4.9)$$

II) Se $\text{Posto}(\mathbf{W}) = p < n$, então (4.2) possui infinitas soluções. Por questão de estabilidade e sensibilidade mínima, uma única solução com norma mínima é escolhida. Neste caso, a ferramenta de decomposição em valores singulares torna-se bastante útil para expressar a solução \mathbf{b}_{LS} de norma mínima.

Considerando, portanto, a decomposição para a matriz \mathbf{W} da mesma forma apresentada em (3.24), teremos

$$\mathbf{W} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H = \sum_{i=1}^p \sigma_i \mathbf{u}_i \mathbf{v}_i^H, \quad (4.10)$$

onde \mathbf{u}_i e \mathbf{v}_i são os vetores singulares de \mathbf{W} que representam, respectivamente, as colunas das matrizes \mathbf{U} e \mathbf{V} , sendo σ_i seus respectivos valores singulares, com $\sigma_1 \geq \dots \geq \sigma_p > \sigma_{p+1} = \dots = \sigma_n = 0$.

Neste caso, a solução \mathbf{b}_{LS} de norma mínima desejada será dada por [VanHuffel91], [Lemos97]

$$\mathbf{b}_{LS} = \mathbf{W}^\dagger \mathbf{w} = \sum_{i=1}^p \frac{1}{\sigma_i} \mathbf{v}_i \mathbf{u}_i^H(\mathbf{w}), \quad (4.11)$$

onde \mathbf{W}^\dagger denota a pseudo-inversa da matriz \mathbf{W} .

Podemos notar que a solução em (4.11) também satisfaz o caso (I) no qual $\text{Posto}(\mathbf{W}) = n$, para a qual reconhecemos a igualdade $(\mathbf{W}^H \mathbf{W})^{-1} \mathbf{W}^H = \mathbf{W}^\dagger$.

Considerando, ainda, a possibilidade de relacionarmos os vetores \mathbf{v}_i e \mathbf{u}_i através da relação $\mathbf{W} \mathbf{v}_i = \mathbf{u}_i \sigma_i$, podemos notar através de (4.11) que para obter o vetor solução \mathbf{b}_{LS} é necessário apenas p vetores singulares e seus respectivos p maiores valores singulares de \mathbf{W} .

Nos procedimentos baseados em subespaços é usual adotar uma solução envolvendo a redução da dimensão do posto considerado em (4.11), inspirada na versão sem ruído para o modelo dos dados observados. Dessa forma, a escolha de um posto reduzido de dimensão $r < p$, segundo algum critério adequado, produzirá uma solução dada por [Haykin89]

$$\mathbf{b}_r = \sum_{i=1}^r \frac{1}{\sigma_i} \mathbf{v}_i (\mathbf{u}_i^H \mathbf{w}). \quad (4.12)$$

Devemos ressaltar que embora o vetor solução \mathbf{b}_r obtido através de (4.12) apresente estimativas úteis quando r é escolhido adequadamente, o mesmo não minimiza o erro quadrático e não possui a menor norma Euclidiana (norma-mínima), como no caso do vetor \mathbf{b}_{LS} obtido por (4.11).

Para o caso particular em que os dados disponíveis são amostras de um sinal proveniente de um arranjo de sensores linear e uniforme, podemos obter as estimativas dos parâmetros DOA através da estimativa de \mathbf{b} , aplicando os conceitos de predição linear, conforme apresentado no Capítulo 3, Seção 3.3.1. Nesse caso, a matriz \mathbf{W} e o vetor \mathbf{w} são dados, respectivamente, pela matriz de dados \mathbf{D} e pelo vetor de observações \mathbf{d} , definidos para o filtro preditor considerado.

4.3 Mínimos Quadrados Totais (TLS)

Abordar um problema de minimização no contexto TLS é mais apropriado para casos em que consideramos erros simultâneos tanto no vetor de observações \mathbf{w} quanto na matriz de dados \mathbf{W} . Quando os erros presentes nessas variáveis são independentes e identicamente distribuídos (i.i.d.), com média zero e matriz de correlação dada pela matriz identidade, a menos de um fator de escala, a solução TLS fornecerá a melhor estimativa estatística [VanHuffel91].

Sendo mais específico, consideraremos a adição de uma perturbação \mathbf{E} em \mathbf{W} e \mathbf{e} em \mathbf{w} na expressão (4.2), tais que

$$(\mathbf{W} + \mathbf{E})\mathbf{b} = \mathbf{w} + \mathbf{e} \quad (4.13)$$

ou, conseqüentemente,

$$\left(\begin{bmatrix} \mathbf{w}:\mathbf{W} \\ \mathbf{e}:\mathbf{E} \end{bmatrix} + \begin{bmatrix} -1 \\ \mathbf{b} \end{bmatrix} \right) = \mathbf{0}. \quad (4.14)$$

A solução TLS da equação homogênea em (4.14) pode ser então formulada como a busca do vetor \mathbf{b}_{TLS} tal que as contribuições das perturbações sejam as menores possíveis, ou seja

$$\left\| \begin{bmatrix} \mathbf{e}:\mathbf{E} \end{bmatrix} \right\|_F^2 \text{ seja mínimo} \quad (4.15)$$

e na qual o vetor composto $(\mathbf{w} + \mathbf{e})$ esteja no espaço definido pelas colunas da matriz composta $(\mathbf{W} + \mathbf{E})$.

Podemos notar que como $\begin{bmatrix} \mathbf{w}:\mathbf{W} \\ -1 \end{bmatrix} \neq \mathbf{0}$, a matriz $\begin{bmatrix} \mathbf{w}:\mathbf{W} \\ -1 \end{bmatrix}$, de dimensões $m \times (n + 1)$, terá posto igual a $n + 1$, para $m > n + 1$. Se o valor de $\left\| \begin{bmatrix} \mathbf{e}:\mathbf{E} \end{bmatrix} \right\|_F^2$ for mínimo e (4.13) for satisfeita, então $\left(\begin{bmatrix} \mathbf{w}:\mathbf{W} \\ \mathbf{e}:\mathbf{E} \end{bmatrix} + \begin{bmatrix} -1 \\ \mathbf{b} \end{bmatrix} \right)$ terá posto menor ou igual a n . Esse resultado é conseqüência do fato de termos (4.14) satisfeita.

Através do teorema de *Eckart-Young-Mirsky* podemos constatar que $\left(\begin{bmatrix} \mathbf{w}:\mathbf{W} \\ \mathbf{e}:\mathbf{E} \end{bmatrix} + \begin{bmatrix} -1 \\ \mathbf{b} \end{bmatrix} \right)$ é a melhor aproximação de posto n para $\begin{bmatrix} \mathbf{w}:\mathbf{W} \\ -1 \end{bmatrix}$, uma vez que as contribuições das perturbações são as menores possíveis e $\begin{bmatrix} \mathbf{e}:\mathbf{E} \end{bmatrix}$ tenha posto unitário [VanHuffel91]. Neste caso, podemos ainda constatar que o valor da norma de *Frobenius* em (4.15) é o mesmo da sua norma *Euclidiana* e igual ao valor do menor valor singular da matriz $\begin{bmatrix} \mathbf{w}:\mathbf{W} \\ -1 \end{bmatrix}$, ou seja

$$\left\| \begin{bmatrix} \mathbf{e}:\mathbf{E} \end{bmatrix} \right\|_F^2 = \left\| \begin{bmatrix} \mathbf{e}:\mathbf{E} \end{bmatrix} \right\|_2^2 = \sigma_{\min}^2.$$

Para a solução da equação homogênea em (4.14), definiremos a matriz composta $\mathbf{C} = \begin{bmatrix} \mathbf{w} : \mathbf{W} \end{bmatrix}$ e consideraremos a seguinte decomposição

$$\mathbf{C} = \mathbf{U}_c \mathbf{\Sigma}_c \mathbf{V}_c^H, \quad (4.16)$$

onde

$$\begin{aligned} \mathbf{U}_c &= \begin{bmatrix} \mathbf{u}_1, & \cdots & \mathbf{u}_m \end{bmatrix}_{m \times m} \\ \mathbf{V}_c &= \begin{bmatrix} \mathbf{v}_1, & \cdots & \mathbf{v}_{n+1} \end{bmatrix}_{(n+1) \times (n+1)} \\ \mathbf{\Sigma}_c &= \begin{bmatrix} \mathbf{\Sigma}_1 & 0 \\ 0 & \sigma_{n+1} \end{bmatrix} = \text{diag} \left[\sigma_1 \ \sigma_2 \ \cdots \ \sigma_{n+1} \right]_{m \times (n+1)}. \end{aligned}$$

Dependendo dos valores singulares obtidos para a matriz \mathbf{C} em (4.16), teremos os casos a seguir para o vetor solução \mathbf{b} no contexto TLS, denotado por \mathbf{b}_{TLS} :

CASO I) A matriz \mathbf{C} possui todos seus valores singulares distintos:

Neste caso a solução TLS, se existir, será única, desde que $v_{n+1,n+1} \neq 0$, onde $v_{n+1,n+1}$ é o elemento presente na última linha da última coluna da matriz \mathbf{V}_c .

Para $\sigma_{n+1} \neq 0$, \mathbf{C} possui posto $n + 1$, o espaço gerado pelas colunas de \mathbf{C} coincide com \mathbb{C}^{n+1} e o espaço gerado pelas linhas de \mathbf{C} , seu espaço nulo, será vazio. Neste caso, o conjunto de equações em (4.14) é incompatível por não existir nenhum vetor \mathbf{b} que satisfaz $\mathbf{C} \begin{bmatrix} -1 \\ \mathbf{b} \end{bmatrix} = \mathbf{0}$. Portanto, para obter a solução, o posto de \mathbf{C} deverá ser reduzido para n , implicando em um espaço nulo não vazio.

Pelo teorema de *Eckart-Young-Mirsky* a melhor aproximação TLS com posto n para \mathbf{C} é dada por

$$\tilde{\mathbf{C}} = \mathbf{U}_c \tilde{\mathbf{\Sigma}}_c \mathbf{V}_c^H \quad \text{com} \quad \tilde{\mathbf{\Sigma}}_c = \text{diag}(\sigma_1, \dots, \sigma_n, 0). \quad (4.17)$$

Neste caso, a correção TLS mínima será

$$\min \left\| \mathbf{C} - \tilde{\mathbf{C}} \right\|_F = \sigma_{n+1}, \quad (4.18)$$

e o conjunto de equações compatíveis será dado por

$$\tilde{\mathbf{C}} \begin{bmatrix} -1 \\ \mathbf{b} \end{bmatrix} = \mathbf{0}. \quad (4.19)$$

A solução para (4.19) será dada, portanto, pelo único vetor que pertence ao espaço nulo de $\tilde{\mathbf{C}}$, ou seja \mathbf{v}_{n+1} , última coluna de \mathbf{V}_c . A solução TLS é então obtida “escalando” \mathbf{v}_{n+1} para que sua primeira componente seja igual a -1 . Assim, teremos

$$\begin{bmatrix} -1 \\ \mathbf{b}_{TLS} \end{bmatrix} = \frac{-1}{v_{1,n+1}} \mathbf{v}_{n+1}, \quad \text{se} \quad v_{n+1,n+1} \neq 0. \quad (4.20)$$

Portanto,

$$\mathbf{b}_{TLS} = -\frac{1}{v_{1,n+1}} [v_{1,n+1}, \dots, v_{n,n+1}]^T. \quad (4.21)$$

Caso $v_{n+1,n+1} = 0$, o problema TLS falha, não apresentando solução. Problemas que caem nesta condição são denominados de não genéricos. Em alguns casos, colunas da matriz \mathbf{W} podem ser retiradas até que a sub-matriz restante possua posto cheio [VanHuffel91].

Podemos notar que para $\sigma_{n+1} = 0$, nenhuma aproximação será necessária e a solução, sem correção, também será dada por (4.21).

CASO II) A matriz \mathbf{C} possui valores singulares repetidos:

Neste caso a solução TLS não será única. A solução TLS, tratada aqui, fornecerá a solução de menor norma *Euclidiana* dentre as infinitas soluções possíveis.

Se σ_{n+1} for um valor singular repetido de \mathbf{C} , isto é, se $\sigma_1 > \sigma_2 \cdots > \sigma_r > \sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_{n+1}$, o procedimento TLS fornecerá mais de uma solução. Esse caso também compreende as situações nas quais temos os valores singulares repetidos iguais a zero.

Neste caso, a solução também apresentará a forma $\mathbf{b}_{TLS} = -\mathbf{y}/\alpha$, como em (4.21), onde α é um escalar definido adequadamente e o vetor \mathbf{y} será dado por uma combinação linear dos vetores de \mathbf{V}_c associados aos $n - r + 1$ menores valores singulares em Σ_c , tal que \mathbf{b}_{TLS} apresente a menor norma *Euclidiana* possível.

Pode-se demonstrar que a solução de norma-mínima para o problema TLS pode ser então obtida definindo-se $\mathbf{V}_2 = [\mathbf{v}_{r+1}, \dots, \mathbf{v}_{n+1}]$ e computando a matriz de *Householder*, \mathbf{Q} , tal que

$$\mathbf{V}_2 \mathbf{Q} = \begin{bmatrix} \boxtimes & \mathbf{y} \\ 0 \cdots 0 & \alpha \end{bmatrix}, \quad (4.22)$$

onde \boxtimes é uma matriz qualquer com dimensão apropriada.

Pode ser demonstrado que o vetor da forma \mathbf{y}/α , obtido a partir de (4.22), será dado pela combinação linear de \mathbf{v}_i , para $i = r+1, \dots, n+1$, de menor norma *Euclidiana*, conforme desejado. Assim, para $\alpha \neq 0$, teremos a solução TLS dada por $\mathbf{b}_{TLS} = -\mathbf{y}/\alpha$.

Se o resultado da decomposição em (4.22) apresentar $\alpha = 0$, então o problema não possui solução TLS, a não ser que o valor de r seja reduzido até se obter um valor diferente de zero para α .

Pode ser notado que para determinar \mathbf{b}_{TLS} a partir de (4.22), apenas a última coluna de \mathbf{Q} é necessária, reduzindo assim o esforço computacional requerido nesse procedimento.

Para explicitar a solução para esse caso, é sugerida a seguinte partição para \mathbf{V}_2

$$\mathbf{V}_2 = \begin{bmatrix} \mathbf{V}'_2 \\ \nu_2 \end{bmatrix}, \quad (4.23)$$

na qual ν_2 representa a última linha de \mathbf{V}_2 e \mathbf{V}'_2 é uma sub-matriz de \mathbf{V}_2 , com dimensões $(n) \times (n - r + 1)$. Dessa forma, demonstra-se que \mathbf{b}_{TLS} será expresso como

$$\mathbf{b}_{TLS} = -\frac{\mathbf{V}'_2 \nu_2^H}{\nu_2^H \nu_2}, \quad (4.24)$$

ou, de forma explícita,

$$\mathbf{b}_{TLS} = -\frac{1}{\sum_{i=r+1}^{n+1} |v_{i,n+1}|^2} \sum_{k=r+1}^{n+1} (v_{k,n+1})^* \mathbf{v}'_k, \quad (4.25)$$

na qual

$$\mathbf{v}_k = \begin{bmatrix} \mathbf{v}'_k \\ v_{k,n+1} \end{bmatrix}.$$

Comparando-se (4.25) com (4.12) podemos constatar que a solução TLS é obtida através da combinação linear dos vetores singulares associados aos menores valores singulares, enquanto que a solução LS com posto reduzido é obtida através da combinação linear dos vetores singulares associados aos maiores valores singulares.

A solução TLS obtida através da combinação linear dos vetores singulares associados aos maiores valores singulares também é possível. Particionando a matriz \mathbf{V}_c conforme a seguir

$$\mathbf{V}_c = [\mathbf{V}_1 | \mathbf{V}_2] = \begin{bmatrix} \mathbf{V}'_1 & \mathbf{V}'_2 \\ \nu_1 & \nu_2 \end{bmatrix}, \quad (4.26)$$

e lembrando que $\mathbf{V}_c \mathbf{V}_c^H = \mathbf{I}$, podemos estabelecer as seguintes relações

$$\begin{aligned} \nu_2^H \nu_2 &= 1 - \nu_1^H \nu_1 \\ -\mathbf{V}'_2 \nu_2^H &= \mathbf{V}'_1 \nu_1^H. \end{aligned} \quad (4.27)$$

Substituindo (4.27) em (4.24), obtemos

$$\mathbf{b}_{TLS} = \frac{\mathbf{V}'_1 \nu_1^H}{1 - \nu_1^H \nu_1}. \quad (4.28)$$

Devemos observar, ainda, que para $m < n + 1$, a matriz \mathbf{C} , de dimensões $m \times (n + 1)$, terá posto $p = \min(m, n + 1) \leq m$. Conseqüentemente, teremos $\sigma_{p+1} = \dots = \sigma_{n+1} = 0$ e, portanto, poderemos também considerar a solução do **CASO II** para essa situação.

4.4 Problema TLS Equivalente

A formulação para o problema de minimização apresentada nesta seção se aplica nos casos em que estamos interessados apenas na relação linear entre as colunas da matriz $\mathbf{C} = [\mathbf{w} : \mathbf{W}]$, não importando qual coluna de \mathbf{C} está no lado direito da equação em (4.13).

Considerando o conjunto genérico de equações do tipo

$$\mathbf{C}_{p \times q} \mathbf{c} \approx 0, \text{ com } p > q, \quad (4.29)$$

na qual a matriz \mathbf{C} possui posto q e todos os seus valores singulares distintos, o problema proposto aqui consiste em obter a matriz aproximada $\tilde{\mathbf{C}}$ que satisfaz

$$\min \left\| \mathbf{C} - \tilde{\mathbf{C}} \right\|_F \text{ sujeito a } \tilde{\mathbf{C}} \mathbf{c} = 0 \text{ e } \mathbf{c}^H \mathbf{c} = 1, \quad (4.30)$$

na qual $\mathbf{c} = [c_0, c_1, \dots, c_q]^T$ e $\tilde{\mathbf{C}}$ é uma aproximação de \mathbf{C} com posto reduzido de uma unidade, segundo o teorema de *Eckart-Young-Mirsky*. Observe que as restrições impostas em (4.30) têm por objetivo garantir a existência de uma solução útil.

Utilizando uma abordagem através de *multiplicadores de Lagrange*, pode ser mostrado que a solução para o problema proposto em (4.30) será dada por [VanHuffel91]

$$\begin{aligned}\tilde{\mathbf{C}} &= \mathbf{C} - \mathbf{u}_q \sigma_q \mathbf{v}_q^H, \\ \mathbf{c} &= \mathbf{v}_q\end{aligned}\quad (4.31)$$

onde σ_q é o menor valor singular de \mathbf{C} , e \mathbf{u}_q e \mathbf{v}_q são seus vetores singulares correspondentes. Neste caso a correção será

$$\min \left\| \mathbf{C} - \tilde{\mathbf{C}} \right\|_F = \sigma_q. \quad (4.32)$$

A restrição quadrática $\mathbf{c}^H \mathbf{c} = 1$ em (4.30) pode ser substituída por uma restrição linear sobre os elementos de \mathbf{c} , como por exemplo $c_0 = -1$. Podemos observar que a solução dada por \mathbf{v}_q , sem nenhuma normalização particular, é suficiente uma vez que $\mathbf{v}_q^H \mathbf{v}_q = 1$ ou podemos escalonar \mathbf{c} livremente. Podemos notar, ainda, que a solução para (4.30) é dada pelo vetor singular de \mathbf{C} associado ao seu menor valor singular e, portanto, coincide com a solução TLS correspondente a menos de um fator de escala.

Um problema de minimização equivalente a (4.30) consiste em obter

$$\min \quad \|\mathbf{C}\mathbf{c}\|_2^2 \quad \text{sujeito a } \mathbf{c}^H \mathbf{c} = 1, \quad (4.33)$$

na qual podemos interpretar $\mathbf{C}\mathbf{c} = \mathbf{e}$ como um erro residual. A restrição imposta em (4.33) é necessária para evitar a solução trivial $\mathbf{c} = \mathbf{0}$.

Para demonstrar a equivalência entre os problemas propostos em (4.33) e (4.30), podemos constatar, a partir da decomposição de \mathbf{C} em valores singulares, que

$$\frac{\|\mathbf{C}\mathbf{c}\|_2^2}{\mathbf{c}^H \mathbf{c}} \geq \sigma_q^2; \quad \text{com } \mathbf{c} \neq \mathbf{0}, \quad (4.34)$$

na qual o vetor unitário $\mathbf{c} = \mathbf{v}_q$ satisfaz a igualdade. Considerando o resultado em (4.32), constatamos a equivalência.

Devemos ainda ressaltar que o vetor \mathbf{c} que satisfaz a igualdade em (4.34) pode ser escalonado por qualquer fator sem alterar o resultado. Assim, a equivalência também é válida se a restrição de norma $\mathbf{c}^H \mathbf{c} = 1$ for substituída por uma restrição linear.

4.4.1 Solução TLS Alternativa

Considerando o problema equivalente ao problema TLS, podemos obter sua solução a partir da SVD da matriz \mathbf{C} , conforme apresentado na Seção 4.3.

Apresentaremos, aqui, uma forma de obtenção da solução para o problema em (4.33), utilizando uma restrição linear para a não-trivialidade da solução.

Para obter uma solução para (4.33) prosseguimos com a derivada de (4.33) em relação a \mathbf{c} e igualando o seu resultado a zero, de onde obtemos

$$\frac{\partial \mathbf{c}^H \mathbf{C}^H \mathbf{C} \mathbf{c}}{\partial \mathbf{c}} = \mathbf{C}^H \mathbf{C} \mathbf{c} = \mathbf{0}. \quad (4.35)$$

Apresentaremos, a seguir, uma implementação alternativa para a solução da equação normal em (4.35), através da decomposição **Q-R** da matriz $\mathbf{C}^H\mathbf{C}$. Nessa implementação, para evitar a solução trivial, faremos uso de uma restrição explícita sob os elementos de \mathbf{c} .

Seja c_i o elemento da i 'ésima linha do vetor \mathbf{c} . Para evitar a solução trivial tomemos a restrição linear na qual impomos $c_i = 1$ durante o processo de minimização em (4.33). Façamos, agora, \mathbf{L} ser uma matriz construída a partir da matriz $\mathbf{C}^H\mathbf{C}$ através da retirada da sua i 'ésima coluna, e \mathbf{r} o vetor formado pela i 'ésima coluna retirada de $\mathbf{C}^H\mathbf{C}$. Além disso, façamos μ representar \mathbf{c} com o seu i 'ésimo elemento removido.

Após essas considerações, podemos substituir o problema em (4.33), sujeito à restrição linear alternativa, pelo problema de minimização sem restrições a seguir

$$\min_{\mu} \|\mathbf{r} + \mathbf{L}\mu\|_2^2. \quad (4.36)$$

Fatorando \mathbf{L} através da decomposição **Q-R** de forma que $\mathbf{L} = \mathbf{Q}\mathbf{R}$, onde \mathbf{R} é uma matriz quadrada triangular superior e $\mathbf{Q}^H\mathbf{Q} = \mathbf{I}$, podemos mostrar que (4.36) é equivalente a

$$\min_{\mu} \|\mathbf{Q}^H\mathbf{r} + \mathbf{R}\mu\|_2^2. \quad (4.37)$$

A solução para o problema em (4.37) é naturalmente obtida quando

$$-\mathbf{R}\mu = \mathbf{Q}^H\mathbf{r}. \quad (4.38)$$

Como \mathbf{R} é uma matriz triangular superior, (4.38) pode ser resolvida facilmente para μ . Uma vez obtido μ , o vetor \mathbf{c} poderá ser então construído. Tal procedimento se mostra mais eficiente que o procedimento de solução direta da equação normal.

Essa solução também pode ser obtida através da utilização de *multiplicadores de Lagrange* incorporando uma restrição linear em \mathbf{c} para evitar a solução trivial.

Concluimos, portanto, que a solução TLS via SVD oferece a solução de norma-mínima, na qual a restrição de não-trivialidade é considerada de forma implícita. Uma solução alternativa para um problema TLS equivalente pode ser obtida através da decomposição **Q-R**, quando implementamos, de forma explícita, uma restrição linear sob os elementos de \mathbf{c} . Nessas condições, a obtenção da solução através do uso da decomposição **Q-R** da matriz $\mathbf{C}^H\mathbf{C}$ se mostra mais eficiente que o uso da SVD. Observamos, ainda, que tal implementação, considerando uma restrição linear específica com $c_0 = 1$ para o problema TLS, apresenta a mesma estrutura de um problema LS. Dessa forma, a implementação da solução via decomposição **Q-R** torna-se também uma alternativa que pode ser aplicada em problemas de minimização do tipo LS.

4.5 Conclusão

Apesar dos procedimentos LS e TLS apresentados neste capítulo se caracterizarem como procedimentos de otimização para a obtenção de estimativas de parâmetros lineares, os mesmos podem ser aplicados na busca de parâmetros não-lineares (DOA) através de uma reparametrização

não-linear do problema original. Nessa reparametrização, a busca dos ângulos DOA é substituída pela obtenção dos elementos de um vetor \mathbf{b} . Posteriormente, os ângulos DOA são obtidos a partir da posição angular das raízes do polinômio formado pelos elementos do vetor \mathbf{b} , através de um mapeamento não-linear.

No caso particular em que o sistema de equações lineares do tipo $\mathbf{W}\mathbf{b} = \mathbf{w}$ origina-se de um modelo de predição linear, a matriz $\mathbf{C} = \begin{bmatrix} \mathbf{W} & \mathbf{w} \end{bmatrix}$ possuirá uma estrutura particular do tipo *Hankel*, *Toeplitz* ou *Hankel-Toeplitz*, dependendo da estrutura do filtro utilizado. Neste caso, pode ser desejado a mesma estrutura para a aproximação $\tilde{\mathbf{C}}$ em (4.30). A solução para esse tipo de problema não será mais dada pela SVD da matriz \mathbf{C} . A manutenção da estrutura da matriz original implicará na adição de restrições ao problema TLS clássico, dando origem aos problemas denominados de TLS com restrições (CTLS) ou TLS estruturado (STLS). Uma outra variação para o problema TLS com restrições, aplicado no contexto da predição linear, consiste em restringir o erro a ter uma característica de ruído branco. Neste caso é a estrutura do erro que deve ser preservada, dando origem aos problemas denominados de TLS ponderado ou “branqueado” (WTLS). Tais procedimentos com restrições serão abordados no capítulo seguinte.

A introdução das variáveis \mathbf{W} e \mathbf{w} foi intencional para enfatizar o caráter genérico dos procedimentos LS e TLS. No caso particular da predição linear apresentada no capítulo anterior, as variáveis correspondentes serão \mathbf{D} e \mathbf{d} , respectivamente.

Problemas do tipo $\min \left\| \mathbf{C} - \tilde{\mathbf{C}} \right\|_F$ formam a base para os procedimentos totalmente baseados em subespaços, fazendo, portanto, uma ponte entre o critério TLS e o critério “Subspace fitting” (SF).

Capítulo 5

Adicionando Informações ao Problema de Otimização

5.1 Introdução

Variações para os problemas de otimização, apresentados no capítulo anterior, podem ser obtidas através da incorporação de informações adicionais ao problema original. A obtenção de tais informações geralmente depende de um conhecimento profundo do problema e dos objetivos de interesse, conforme ficará claro no decorrer do capítulo. Essas informações podem ser incorporadas ao problema de otimização de uma forma implícita ou através de restrições matemáticas explícitas. Em geral, a incorporação de restrições matemáticas reduz a dimensão do problema original e, conseqüentemente, reduz o esforço computacional necessário para atingir sua solução [DeMoor93b].

No capítulo anterior, as restrições eram intrínsecas ao problema e tinham como objetivo a obtenção de uma solução não trivial. Portanto, diferem das restrições tratadas neste capítulo que são impostas ao modelo de otimização.

Podemos classificar tipos distintos de restrições. Um primeiro tipo consiste em impor restrição à estrutura do modelo utilizado. Esse tipo de restrição se aplica, por exemplo, no caso em que partimos de um modelo de predição linear cuja matriz de dados possui uma estrutura particular do tipo *Hankel-Toeplitz*, sendo, portanto, interessante obtermos uma solução otimizada que leve à mesma estrutura. Um segundo tipo de restrição consiste na imposição de uma característica de ruído do tipo branco ao erro de estimação. Esse tipo de restrição também se aplica no contexto da predição linear, na qual temos o filtro de erro de predição atuando como um filtro “branqueador”, quando otimizado para os dados de entrada. Podemos ainda considerar um terceiro tipo que não se caracteriza como um tipo isolado de restrição, mas sim como um procedimento adicional. Nesse último são utilizadas informações resultantes da decomposição da matriz de dados em subespaços, o que permite a extração da porção ruidosa através da supressão do subespaço de ruído.

Os algoritmos mais comuns existentes para tratar problemas com restrições serão também apresentados. Parte de tais algoritmos consiste de variações do procedimento TLS apresentado no capítulo anterior. Os algoritmos TLS com restrições (CTLS) [Abatzoglou91] e TLS estruturado (STLS) [DeMoor94] se aplicam quando a forma de restrição imposta compreende a estrutura da matriz de dados. Já o algoritmo TLS ponderado ou “branqueado” (WTLS) se

aplica no caso em que impomos a condição de ruído branco para o erro de predição [Hua90].

Para o tipo de restrição em que utilizamos informações dos subespaços associados aos dados do problema, nenhum algoritmo específico será apresentado, por não se tratar de uma forma isolada de procedimento. Para esse tipo, alguns exemplos em que encontramos a sua aplicação serão citados.

Apesar de não estarem restritas ao contexto da predição linear, tais formas de restrição serão apresentadas apenas nesse contexto por satisfazer nosso interesse. As relações entre as diversas formas de restrição também serão apresentadas, possibilitando a utilização de formas combinadas de restrições em pesquisas futuras.

5.2 Estruturando os Dados

De acordo com o modelo de predição apresentado no Capítulo 3, podemos observar que os dados disponíveis são arranjados de uma forma sistemática para a formação da matriz de dados \mathbf{D} . Nessa forma, as componentes de dados estão relacionadas linearmente, bem como as componentes de ruído presente, levando a contribuições repetidas. Na solução TLS para o preditor ótimo apresentada no Capítulo 4, tal estrutura não foi levada em consideração.

No caso de um filtro preditor progressivo, a matriz de dados possui a forma *Toeplitz*. Já para um filtro regressivo, a estrutura será *Hankel* e, conseqüentemente, para um filtro preditor progressivo-regressivo, a matriz de dados possuirá uma estrutura *Hankel-Toeplitz* [Antunes92]. Essas informação podem ser levadas em consideração no processo de otimização desses preditores conforme apresentaremos.

Mais especificamente, considerando o vetor de dados \mathbf{d} e a matriz de dados \mathbf{D} , no contexto da predição linear apresentada no Capítulo 3, e os coeficientes do filtro preditor dados pelos elementos do vetor \mathbf{b} , a imposição da restrição quanto à estrutura transforma o problema de otimização TLS, dado em (4.30), no seguinte problema:

$$\begin{aligned} \min_{\Delta\mathbf{C}, \mathbf{b}} \|\Delta\mathbf{C}\|_F \quad \text{sujeito a } (\mathbf{C} + \Delta\mathbf{C}) \begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix} = 0; \quad \left\| \begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix} \right\|_2 \neq 0 \\ \text{com} \\ \Delta\mathbf{C} = \left[\mathbf{F}_1\mathbf{z} : \mathbf{F}_2\mathbf{z} : \cdots : \mathbf{F}_{L+1}\mathbf{z} \right], \end{aligned} \quad (5.1)$$

onde $\mathbf{C} = \begin{bmatrix} \mathbf{d} : \mathbf{D} \end{bmatrix}$, $\Delta\mathbf{C} = \begin{bmatrix} \mathbf{e} : \mathbf{E} \end{bmatrix}$ e $\mathbf{z} = [z_0, z_1, \dots, z_{N-1}]^T$ é o vetor de perturbações com a dimensão mínima necessária. As matrizes \mathbf{F}_l , para $l = 1, \dots, L + 1$, possuem dimensões $(N - L) \times N$ e são compostas por zeros e uns, de modo a estabelecer as relações desejadas entre as componentes da matriz $\Delta\mathbf{C}$ e, portanto, dependem da estrutura do filtro utilizado [Abatzoglou91]. O vetor \mathbf{z} de perturbações deve possuir uma quantidade de elementos distintos igual à quantidade de amostras de dados distintos em \mathbf{C} . Portanto, deve possuir a mesma dimensão dos dados disponíveis, $N \times 1$.

O algoritmo que propicia uma solução do tipo TLS que incorpora a dependência entre os elementos da matriz \mathbf{C} conforme (5.1) é denominado de TLS com restrições (CTLS) e será apresentado na próxima seção.

De uma outra forma, podemos também reformular o problema de otimização com restrição em (5.1) no seguinte problema equivalente:

$$\min_{\Delta \mathbf{C}, \mathbf{b}} \|\Delta \mathbf{C}\|_F \quad \text{sujeito a } (\mathbf{C} + \Delta \mathbf{C}) \begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix} = \mathbf{0}; \quad \left\| \begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix} \right\|_2 \neq 0$$

com

$$\mathbf{C} = \sum_{n=0}^{N-1} x_n \mathbf{S}_n \quad e \quad (\mathbf{C} + \Delta \mathbf{C}) = \sum_{n=0}^{N-1} t_n \mathbf{S}_n, \quad (5.2)$$

na qual x_n são os elementos do vetor de amostras $\mathbf{x} = [x_0, x_1, \dots, x_{N-1}]$, $t_n = (x_n + z_n)$ e \mathbf{S}_n são matrizes compostas por zeros e uns de forma adequada [DeMoor94]. Podemos observar que o problema em (5.2) difere de (5.1) apenas na forma de se impor a restrição sobre a estrutura dos dados. Em (5.1) apenas $\Delta \mathbf{C}$ é estruturado, enquanto que em (5.2) é imposta a mesma estrutura de \mathbf{C} para $(\mathbf{C} + \Delta \mathbf{C})$.

O algoritmo que propicia uma solução do tipo TLS que incorpora a dependência entre os elementos de \mathbf{C} conforme (5.2) é denominado de TLS estruturado (STLS) e também será apresentado posteriormente ao algoritmo CTLS.

5.2.1 Algoritmo CTLS

Esse algoritmo implementa a solução para o problema em (5.1) considerando $b_0 = -1$. Podemos constatar que

$$\Delta \mathbf{C} \begin{pmatrix} -1 \\ \mathbf{b} \end{pmatrix} = \left[\mathbf{F}_1 \mathbf{z} : \mathbf{F}_2 \mathbf{z} : \dots : \mathbf{F}_{L+1} \mathbf{z} \right] \begin{pmatrix} -1 \\ \mathbf{b} \end{pmatrix} = \left(-\mathbf{F}_1 + \sum_{l=2}^{L+1} \mathbf{F}_l b_{l-1} \right) \mathbf{z}, \quad (5.3)$$

onde $\mathbf{b}^T = [b_1, \dots, b_L]$.

Definindo

$$\mathbf{G}^H = -\mathbf{F}_1 + \sum_{l=2}^{L+1} \mathbf{F}_l b_{l-1}; \quad \mathbf{G}^H \in C^{(N-L) \times (L+1)}, \quad (5.4)$$

teremos

$$(\mathbf{C} + \Delta \mathbf{C}) \begin{pmatrix} -1 \\ \mathbf{b} \end{pmatrix} = \mathbf{C} \begin{pmatrix} -1 \\ \mathbf{b} \end{pmatrix} + \mathbf{G}^H \mathbf{z}. \quad (5.5)$$

Podemos ainda notar que

$$\|\Delta \mathbf{C}\|_F^2 = \sum_{n=0}^{N-1} \|z_n\|_2^2 p_n = \mathbf{z}^H \boldsymbol{\Omega} \mathbf{z}, \quad (5.6)$$

onde $\boldsymbol{\Omega} = \text{diag}[p_0, \dots, p_{N-1}]$ com $p_n =$ número de ocorrência da componente z_n em $\Delta \mathbf{C}$.

Desta forma, o problema em (5.1) pode ser reescrito como

$$\min_{\mathbf{z}, \mathbf{b}} \left\{ \mathbf{z}^H \boldsymbol{\Omega} \mathbf{z} \right\} \quad \text{sujeito a } \mathbf{C} \begin{pmatrix} -1 \\ \mathbf{b} \end{pmatrix} + \mathbf{G}^H \mathbf{z} = \mathbf{0}, \quad (5.7)$$

que consiste em um problema de minimização quadrática ponderado, com uma restrição quadrática envolvendo duas variáveis: \mathbf{z} e \mathbf{b} . No problema em (5.7), a matriz de perturbações original $\Delta \mathbf{C}$ é então substituída por uma combinação linear do vetor \mathbf{z} de dimensão menor.

O procedimento CTLS transforma ainda o problema (5.7) em um problema sem restrições quando o posto de \mathbf{G}^H é cheio. Para atingir esse resultado, podemos reescrever o problema em (5.7) como

$$\min_{\mathbf{z}, \mathbf{b}} \{ \mathbf{z}^H \boldsymbol{\Omega} \mathbf{z} \} = \min_{\mathbf{b}} \left\{ \min_{\mathbf{z}} \mathbf{z}^H \boldsymbol{\Omega} \mathbf{z} \right\}, \quad (5.8)$$

ou seja, deveremos encontrar o mínimo para um dado vetor \mathbf{b} e depois minimizar para todos os \mathbf{b} .

Fazendo $\mathbf{J} = \mathbf{z}^H \boldsymbol{\Omega} \mathbf{z}$, devemos primeiramente resolver o problema

$$\min_{\mathbf{z}} \{ \mathbf{J} \} \quad \text{sujeito a } \mathbf{C} \begin{pmatrix} -1 \\ \mathbf{b} \end{pmatrix} + \mathbf{G}^H \mathbf{z} = \mathbf{0}; \text{ com } \mathbf{b} \text{ fixo}, \quad (5.9)$$

que é um problema de minimização quadrática com uma restrição linear em apenas uma variável. Tal problema pode ser facilmente resolvido através de *multiplicadores de Lagrange*, de onde obtemos

$$\mathbf{z} = -\boldsymbol{\Omega}^{-1} \mathbf{G} \left(\mathbf{G}^H \boldsymbol{\Omega}^{-1} \mathbf{G} \right)^\dagger \mathbf{C} \begin{pmatrix} -1 \\ \mathbf{b} \end{pmatrix}. \quad (5.10)$$

Substituindo a solução obtida para \mathbf{z} dada em (5.10) na restrição de (5.9), teremos a restrição atualizada como

$$\left[\mathbf{I} - \left(\mathbf{G}^H \boldsymbol{\Omega}^{-1} \mathbf{G} \right) \left(\mathbf{G}^H \boldsymbol{\Omega}^{-1} \mathbf{G} \right)^\dagger \right] \mathbf{C} \begin{pmatrix} -1 \\ \mathbf{b} \end{pmatrix} = \mathbf{0}. \quad (5.11)$$

Considerando a existência de $\boldsymbol{\Omega}^{-1}$, se \mathbf{G}^H tiver posto cheio, então $\mathbf{G}^H \boldsymbol{\Omega}^{-1} \mathbf{G}$ também possuirá inversa. Neste caso temos $\left(\mathbf{G}^H \boldsymbol{\Omega}^{-1} \mathbf{G} \right)^\dagger = \left(\mathbf{G}^H \boldsymbol{\Omega}^{-1} \mathbf{G} \right)^{-1}$ e a restrição em (5.11) sempre será satisfeita para qualquer \mathbf{b} , não havendo necessidade de incorporá-la no processo de otimização. Podemos ainda atestar que \mathbf{G}^H possuirá sempre posto cheio se as matrizes \mathbf{F}_l , para $l = 1, 2, \dots, L + 1$, forem formadas considerando a estrutura de um filtro predictor.

Assim, a solução do problema em (5.7) através do algoritmo CTLS reduz-se a resolver o problema de minimização quadrática sem restrição da forma

$$\min_{\mathbf{b}} \mathbf{z}^H \boldsymbol{\Omega} \mathbf{z} = \min_{\mathbf{b}} \left\{ \begin{pmatrix} -1 \\ \mathbf{b} \end{pmatrix}^H \mathbf{C}^H \left(\mathbf{G}^H \boldsymbol{\Omega}^{-1} \mathbf{G} \right)^{-1} \mathbf{C} \begin{pmatrix} -1 \\ \mathbf{b} \end{pmatrix} \right\}, \quad (5.12)$$

desde que \mathbf{G}^H tenha posto cheio.

Podemos notar que se considerarmos $\boldsymbol{\Omega} = \mathbf{I}$ em (5.12), não estaremos mais considerando as contribuições repetidas das perturbações devido à estrutura. Se considerarmos, ainda, $\mathbf{G}^H \boldsymbol{\Omega}^{-1} \mathbf{G} = \mathbf{I}$ em (5.12), não estaremos mais considerando a estrutura dos dados e, conseqüentemente, teremos um problema sem restrições. Podemos então concluir que o procedimento CTLS, para a obtenção da solução estruturada otimizada sob o critério TLS para o problema $\mathbf{D}\mathbf{b} \approx \mathbf{d}$, pode ser visto como uma generalização do procedimento TLS.

Embora tenhamos reduzido o problema original em (5.7) no problema em (5.12), sua solução não é simples. Em geral é necessário computar as derivadas parciais em relação a \mathbf{b} . Já que a condição necessária para sua minimização é termos todas as derivadas nulas, deveremos resolver um sistema com L equações e L variáveis complexas. Algoritmos numéricos, como o método de Newton, podem ser utilizados para esse propósito [Lueberger73].

5.2.2 Algoritmo STLS

O algoritmo denominado de mínimos quadrados total estruturado (STLS) consiste na busca da solução para o problema de mínimos quadrados no qual a matriz $(\mathbf{C} + \Delta\mathbf{C})$ original é substituída por uma matriz aproximada, com a mesma estrutura da matriz \mathbf{C} , enquanto a norma *Euclidiana* do erro dessa aproximação é minimizada.

Pode ser mostrado que a solução para tal problema pode ser obtida através da ferramenta de decomposição em valores singulares generalizada não-linear, a qual, por sua vez, pode ser resolvida através de um algoritmo baseado na iteração inversa [DeMoor93a], [DeMoor94].

O algoritmo STLS genérico propõe a solução do problema em (5.2), ou seja

$$\min_{\mathbf{t}, \mathbf{b}} \sum_{n=0}^{N-1} p_n (x_n - t_n)^2 \quad \text{sujeito a } (\mathbf{C} + \Delta\mathbf{C}) \begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix} = 0, \quad (5.13)$$

com

$$\begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix}^H \begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix} = 1,$$

onde $\mathbf{t} = [t_0, t_1, \dots, t_{N-1}]$, e p_n são pesos que indicam o número de ocorrência de x_n e t_n em \mathbf{C} e $(\mathbf{C} + \Delta\mathbf{C})$, respectivamente. Definiremos, então, $\mathbf{S}(\mathbf{t}) = \mathbf{S}_0 + \sum_{n=1}^{N-1} t_n \mathbf{S}_n \in R^{p \times q}$ (com $p \geq q$) como uma matriz função dos componentes t_n do vetor dos parâmetros $\mathbf{t} \in R^N$, onde $\mathbf{S}_n, n = 0, 1, \dots, N-1$, são fixos para a matriz dada.

Consideraremos a solução através de *multiplicadores de Lagrange*. Após obter as derivadas parciais e igualando-as a zero, tal procedimento estabelece as seguintes relações:

$$\sum_{n=0}^{N-1} \frac{1}{p_n} \mathbf{S}_n^H \mathbf{u} (\mathbf{u}^H \mathbf{S}_n) = \mathbf{D}_u \quad (5.14)$$

e

$$\sum_{n=0}^{N-1} \frac{1}{p_n} \mathbf{S}_n \mathbf{v} (\mathbf{v}^H \mathbf{S}_n^H) = \mathbf{D}_v, \quad (5.15)$$

onde $\mathbf{u} \in C^{N-L}$, $\mathbf{v} \in C^{L+1}$, $\sigma \in R$ são os vetores singulares e valores singulares obtidos a partir da decomposição da matriz \mathbf{C} . As matrizes \mathbf{D}_u e \mathbf{D}_v são ambas simétricas e definidas positivas ou semi-definidas positivas, com elementos que são quadráticos em suas componentes \mathbf{u} e \mathbf{v} .

Retomando o problema inicial em (5.13), pode ser demonstrado que

$$\min_{\mathbf{t}, \mathbf{b}} \sum_{n=0}^{N-1} p_n (x_n - t_n)^2 = \min \left\{ \mathbf{u}^H \mathbf{D}_v \mathbf{u} \sigma^2 \right\} \quad (5.16)$$

convertendo o problema inicial em (5.13) na obtenção dos vetores \mathbf{u} e \mathbf{v} . Podemos observar que o problema em (5.16) é de quarta ordem, uma vez que \mathbf{D}_v é uma função de segunda ordem em \mathbf{v} .

Pode ser demonstrado que resolver o problema em (5.16) é equivalente a encontrar $(\mathbf{u}, \sigma, \mathbf{v})$ correspondente ao menor escalar σ que satisfaz

$$\begin{aligned} \mathbf{C}\mathbf{v} &= \mathbf{D}_v \mathbf{u} \sigma, & \mathbf{u}^H \mathbf{D}_v \mathbf{u} &= 1 \\ \mathbf{C}^H \mathbf{u} &= \mathbf{D}_u \mathbf{v} \sigma, & \mathbf{v}^H \mathbf{D}_u \mathbf{v} &= 1 \end{aligned} \quad (5.17)$$

Sob a hipótese de que \mathbf{D}_u e \mathbf{D}_v são constantes, (5.17) pode ser vista como uma generalização do problema de autovalores. Em (5.17), como \mathbf{D}_u e \mathbf{D}_v são matrizes dependentes de \mathbf{u} e \mathbf{v} , então o menor autovalor pode ser computado através do algoritmo de iteração inversa, onde a cada iteração \mathbf{D}_u e \mathbf{D}_v são constantes para valores fixos de \mathbf{u} e \mathbf{v} e em seguida atualizadas. Utilizando este algoritmo, o cálculo dos vetores \mathbf{u} e \mathbf{v} pode ser efetuado recursivamente a partir das equações em (5.17).

Resumidamente, o algoritmo parte de valores iniciais $\mathbf{u}^{(0)}$, $\mathbf{v}^{(0)}$ e $\sigma^{(0)}$, seguindo com o cálculo de $\mathbf{u}^{(k+1)} = (\mathbf{D}_v^{(k)})^{-1} \mathbf{C} \mathbf{v}^{(k)} / \sigma^{(k)}$ e, posteriormente, de $\mathbf{v}^{(k+1)} = (\mathbf{D}_u^{(k)})^{-1} \mathbf{C} \mathbf{u}^{(k)} / \sigma^{(k)}$, impondo suas respectivas restrições, até a sua convergência. O algoritmo STLS segue então com a obtenção do vetor $\begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix}$ dado por $\begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix} = \mathbf{v} / \|\mathbf{v}\|_2$.

As componentes do vetor \mathbf{t} , presente em (5.16), são obtidas através de

$$t_n = x_n - \mathbf{u}^H \mathbf{S}_n \mathbf{v} \sigma, \quad n = 0, \dots, N-1. \quad (5.18)$$

Podemos notar ainda que utilizando as relações (5.17) em (5.16), teremos o seguinte problema equivalente

$$\min_{\mathbf{v}} \left\{ \mathbf{v}^H \mathbf{C}^H (\mathbf{D}_v)^{-1} \mathbf{C} \mathbf{v} \sigma^2 \right\}, \quad (5.19)$$

o qual consiste em um problema de minimização de quarta ordem, uma vez que \mathbf{D}_v é uma função quadrática de \mathbf{v} .

5.2.3 Equivalência entre os Algoritmos CTLS e STLS

A seguir apresentaremos a equivalência existente nos procedimentos CTLS e STLS [VanHuffel93], [Lemmerlin96], [Lemos97].

Inicialmente podemos escrever

$$\Delta \mathbf{C} = \left[\mathbf{F}_1 \mathbf{z} : \mathbf{F}_2 \mathbf{z} : \dots : \mathbf{F}_{L+1} \mathbf{z} \right] \quad \text{e} \quad \mathbf{C} = \sum_{n=0}^{N-1} x_n \mathbf{S}_n,$$

assim como podemos escrever a l 'ésima coluna de \mathbf{C} como

$$\mathbf{C}_l = x_0 \mathbf{S}_{0,l} + x_1 \mathbf{S}_{1,l} + \dots + x_{N-1} \mathbf{S}_{N-1,l} = \mathbf{F}_l \mathbf{x},$$

onde $\mathbf{S}_{n,l}$ representa a l 'ésima coluna de \mathbf{S}_n , da qual concluímos que $\mathbf{F}_l = \left[\mathbf{S}_{0,l} : \mathbf{S}_{1,l} : \dots : \mathbf{S}_{N-1,l} \right]$.

A partir de (5.4) e dos resultados anteriores, teremos

$$\mathbf{G}^H = \sum_{l=0}^L b_l \left[\mathbf{S}_{0,l+1} : \mathbf{S}_{1,l+1} : \dots : \mathbf{S}_{N-1,l+1} \right] = \left[\mathbf{S}_0 \begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix} : \mathbf{S}_1 \begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix} : \dots : \mathbf{S}_{N-1} \begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix} \right].$$

Conseqüentemente, teremos $\mathbf{G}^H \mathbf{G} = \sum_{n=1}^N \left[\mathbf{S}_n \begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix} \right] \left[\mathbf{S}_n \begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix} \right]^H$ e, utilizando (5.15), podemos concluir que $\mathbf{G}^H \mathbf{\Omega}^{-1} \mathbf{G} = \mathbf{D}_v / \|\mathbf{v}\|_2^2$.

Escrevendo (5.16) em função de $\begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix}$ através da substituição $\begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix} = \mathbf{v} / \|\mathbf{v}\|_2$, teremos o seguinte problema equivalente:

$$\min_{\mathbf{b}} \left\{ \begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix}^H \mathbf{C}^H (\mathbf{G}^H \boldsymbol{\Omega}^{-1} \mathbf{G})^{-1} \mathbf{C} \begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix} \right\}. \quad (5.20)$$

Se utilizarmos a restrição $b_0 = -1$, teremos então a equivalência com o CTLS. Devemos notar que as mesmas considerações de invertibilidade impostas à $\mathbf{G}^H \boldsymbol{\Omega}^{-1} \mathbf{G}$ devem valer para \mathbf{D}_v .

Concluimos, portanto, que os dois algoritmos procuram minimizar a mesma função custo sujeita a restrições equivalentes.

5.3 Branqueando o Erro de Predição

Essa forma de restrição consiste em impor uma característica de ruído branco ao resultado do erro de predição. A motivação para tal restrição surge da característica do filtro de erro de predição como um filtro “branqueador”, quando otimizado para os dados de entrada. Esse comportamento, apesar de conhecido, não foi utilizado pelos procedimentos de otimização apresentados no capítulo anterior.

Sendo mais específico, considerando o equacionamento da ação do filtro de erro de predição, dado por (3.20) do Capítulo 3, temos

$$\mathbf{e} = \mathbf{C} \begin{pmatrix} -1 \\ \mathbf{b} \end{pmatrix} = \mathbf{D}' \mathbf{b}'. \quad (5.21)$$

Reescrevendo (5.21) através da reparametrização de \mathbf{b} , apresentada no Capítulo 3, teremos

$$\mathbf{e} = \mathbf{B}^H \mathbf{y} = \mathbf{B}^H (\mathbf{x} + \mathbf{n}), \quad (5.22)$$

onde \mathbf{B}^H é dada por (3.29), $\mathbf{n} = [n_0, n_1, \dots, n_{N-1}]^T$ e $\mathbf{y} = [y_0, y_1, \dots, y_{N-1}]^T$.

Pelo Teorema 2 apresentado no Capítulo 3, considerando os dados livres de ruído, teremos o polinômio $b(z)$, de ordem $L \geq M$, com M de suas raízes localizadas sobre a CRU nas posições angulares dada por ϕ_m , $m = 1, 2, \dots, M$. Nessa condição teremos $\mathbf{B}^H \mathbf{x} = 0$ e, conseqüentemente, $\mathbf{e} = \mathbf{B}^H \mathbf{n}$. Através da suposição de que \mathbf{n} é um ruído tipo branco, e sabendo que para o filtro preditor operar em sua condição ótima devemos ter em sua saída um sinal tipo ruído branco, deveríamos esperar que o sinal de erro \mathbf{e} também fosse branco. No entanto, podemos notar que o sinal \mathbf{e} não é branco pois $\mathcal{E}\{\mathbf{e}\mathbf{e}^H\} = \sigma^2 \mathbf{B}^H \mathbf{B}$, onde $\mathcal{E}\{\mathbf{n}\mathbf{n}^H\} = \sigma^2 \mathbf{I}$, com $\mathcal{E}\{\mathbf{n}\} = 0$. Portanto, é natural branquearmos o sinal \mathbf{e} fazendo

$$\mathbf{e}_w = (\mathbf{B}^H \mathbf{B})^{-1/2} \mathbf{B}^H \mathbf{y} = (\mathbf{B}^H \mathbf{B})^{-1/2} \mathbf{C} \begin{pmatrix} -1 \\ \mathbf{b} \end{pmatrix}. \quad (5.23)$$

O procedimento adotado em (5.23) é perfeitamente aplicável a todas as situações na qual desejamos branquear um sinal qualquer.

Retornando ao problema original, devemos agora encontrar o vetor de norma-mínima \mathbf{b} que minimiza a norma do erro em (5.23), ou seja

$$\begin{aligned} \min \{ \mathbf{e}_w^H \mathbf{e}_w \} &= \min_{\mathbf{b}} \left\{ \begin{pmatrix} -1 \\ \mathbf{b} \end{pmatrix}^H \mathbf{C}^H \left[(\mathbf{B}^H \mathbf{B})^{-1/2} \right]^H (\mathbf{B}^H \mathbf{B})^{-1/2} \mathbf{C} \begin{pmatrix} -1 \\ \mathbf{b} \end{pmatrix} \right\} \\ &= \min_{\mathbf{b}} \left\{ \begin{pmatrix} -1 \\ \mathbf{b} \end{pmatrix}^H \mathbf{C}^H (\mathbf{B}^H \mathbf{B})^{-1} \mathbf{C} \begin{pmatrix} -1 \\ \mathbf{b} \end{pmatrix} \right\}. \end{aligned} \quad (5.24)$$

O algoritmo que propicia a solução TLS que incorpora tal restrição sobre o erro de predição é denominado de TLS ponderado ou “branqueado” (WTLS) e será apresentado a seguir [Hua90].

5.3.1 Algoritmo WTLS

Nesse algoritmo o vetor de norma-mínima \mathbf{b} é obtido iterativamente, otimizando o vetor $\mathbf{b}^{(k+1)}$ no espaço nulo definido pela matriz perturbada $(\mathbf{R}_w^{(k)} + \Delta \mathbf{R}_w^{(k)})$, onde k representa o passo da iteração. Mais especificamente, devemos ter $\|\Delta \mathbf{R}_w^{(k)}\|_F$ mínimo e $\mathbf{R}_w^{(k)} \begin{pmatrix} -1 \\ \mathbf{b}^{(k+1)} \end{pmatrix} = 0$, onde $\mathbf{R}_w^{(k)} = \mathbf{C}^H (\mathbf{B}_{(k)}^H \mathbf{B}_{(k)})^{-1} \mathbf{C}$. No passo k da iteração, teremos $(\mathbf{B}_{(k)}^H \mathbf{B}_{(k)})$ fixo e com elementos dado pelo passo anterior. Dessa forma podemos aplicar o procedimentos TLS de minimização a cada passo.

Podemos notar a necessidade da obtenção da inversa da matriz $(\mathbf{B}_{(k)}^H \mathbf{B}_{(k)})$ a cada iteração. Para tanto, algumas alterações no algoritmo podem ser necessárias para contornar um possível mal condicionamento dessa matriz ou apenas para reduzir o esforço de tal operação. Essa observação também é válida para a matriz $(\mathbf{G}^H \mathbf{\Omega}^{-1} \mathbf{G})^{-1}$ presente no procedimentos STLS e CTLS.

Podemos observar que, equivalentemente aos algoritmos STLS e CTLS, o problema de minimização de quarta ordem, oriundo do branqueamento do erro de predição em (5.24), pode ser substituído pela solução iterativa de um problema de minimização de segunda ordem no contexto TLS, através do algoritmo WTLS.

Podemos notar, ainda, que para a primeira iteração, se escolhermos $\mathbf{B}^H \mathbf{B} = \mathbf{I}$, teremos $\mathbf{R}_w = \mathbf{C}^H \mathbf{C}$ e, portanto, a solução coincidirá com aquela obtida pelo procedimento TLS sem restrições, aplicado à predição linear, tratado no capítulo anterior.

5.4 Relação entre WTLS, CTLS e STLS

Podemos notar, comparando (5.24) com (5.12) ou (5.20), que o procedimento de branquear o sinal de erro de predição adotado no algoritmo WTLS leva a um problema de minimização semelhante ao encontrado nos procedimentos de estruturação dos dados, representados pelos algoritmos CTLS e STLS. Uma vez que podemos constatar a igualdade $\mathbf{B}^H = \mathbf{G}^H$, e desprezando as diferentes formas de implementação da restrição de não-trivialidade da solução para \mathbf{b} , podemos notar que a única diferença é a ausência da matriz de ponderação $\mathbf{\Omega}$ no algoritmo WTLS.

Portanto, os três algoritmos procuram tratar as dependências lineares entre as componentes de ruído da matriz de dados. No caso do algoritmo WTLS tal ruído é branqueado para descorrelacionar suas amostras e, conseqüentemente, eliminar a dependência linear existente entre as mesmas.

Originalmente, cada algoritmo apresenta uma estratégia própria para a obtenção da solução da equação de quarta ordem associada ao problema. Podemos notar que tais estratégias podem

ser aplicadas a todos os algoritmos indistintamente, dada a semelhança entre suas formulações.

5.5 Incorporando Informações sobre os Subespaços

Essa terceira forma de incorporar informações adicionais ao problema de minimização está baseada na possibilidade de distinção entre o subespaço de sinal e o subespaço de ruído, definidos pela matriz de dados disponíveis, através de uma decomposição espectral. Uma vez estimado os valores singulares e vetores singulares associados a esses subespaços, torna-se possível e adequado desconsiderar a sua porção de ruído na obtenção das estimativas dos parâmetros desejados.

Tal consideração pode ser realizada antes do processo de otimização, durante o processo ou, ainda, imposta na solução final. Exemplos da sua aplicação serão apresentados no capítulo seguinte, quando abordaremos os métodos de estimação do parâmetro DOA.

Tal forma, assim como as demais, não está restrita ao contexto da predição linear.

5.6 Generalização dos Procedimentos de Otimização

$$\begin{array}{l} \min_{\mathbf{b}} \|\mathbf{H}'\mathbf{b}'\|^2 \\ \min_{\mathbf{b}} \|\Delta\mathbf{H}'\|^2 \\ \min_{\mathbf{b}} \|\mathbf{H}' - \tilde{\mathbf{H}}'\|^2 \end{array}, \quad \tilde{\mathbf{H}}' = \mathbf{H}' - \Delta\mathbf{H}'.$$

Sem Restrições:

$$\min_{\mathbf{b}} \|\mathbf{H}'\mathbf{b}'\|^2 = \min_{\mathbf{b}} \|\mathbf{b}'^H (\mathbf{H}'^H \mathbf{H}') \mathbf{b}'\|, \quad (\mathbf{H}'^H \mathbf{H}') = (\mathbf{Y}'^H \mathbf{Y}')$$

Com Restrições:

$$\min_{\mathbf{b}} \|\mathbf{H}'\mathbf{b}'\|^2 = \min_{\mathbf{b}} \|\mathbf{b}'^H (\mathbf{H}'^H \mathbf{H}') \mathbf{b}'\|, \quad (\mathbf{H}'^H \mathbf{H}') = (\mathbf{Y}'^H [\mathbf{G}^H \boldsymbol{\Omega}^{-1} \mathbf{G}]^{-1} \mathbf{Y}')$$

5.7 Conclusão

Os procedimentos de otimização que impõem restrições ao modelo linear associado aos dados do problema, seja na sua estrutura ou na característica do erro, levam a um problema de minimização de quarta ordem, quando abordados no contexto da predição linear.

Embora a motivação para a restrição imposta na estrutura e na característica do erro sejam diferentes, as mesmas levam a problemas de minimização semelhantes no contexto da predição linear.

Todos os algoritmos existentes para obter a solução dos problemas de minimização com restrições tratados aqui consistem em procedimentos iterativos. Nesses procedimentos, obtém-se a solução para um problema de segunda ordem a cada iteração. Através desse artifício torna-se possível estender a aplicação dos mínimos quadrados apresentada no capítulo anterior em problemas de quarta ordem.

Concluimos que o algoritmo STLS é equivalente ao CTLS, assim como podemos considerá-los também equivalentes ao algoritmo WTLS a menos da existência de uma matriz de ponderação, ausente no algoritmo WTLS. Esse resultado será de grande utilidade quando apresentarmos os métodos para estimação do parâmetro DOA no próximo capítulo.

Uma outra forma de incorporar informações adicionais ao problema de minimização, na qual utilizamos a decomposição em subespaços da matriz dos dados disponíveis, também foi considerada superficialmente.

O intuito deste capítulo é, portanto, apresentar as formas mais comuns de restrições associadas aos problemas de otimização enfrentados pelos diversos métodos de estimação do parâmetro DOA e alternativas para a sua solução.

Capítulo 6

Métodos para Estimar DOA

6.1 Introdução

Este capítulo tem por objetivo apresentar os métodos mais comuns existentes na literatura para estimar o parâmetro DOA através de arranjo de sensores, objeto de estudo no Capítulo 2. Para tanto, faremos uso de todo o conjunto de resultados obtidos nos capítulos anteriores.

Conforme apresentado no Capítulo 3, os procedimentos de estimação baseados no critério da *máxima-verossimilhança* (ML) mostram uma grande capacidade de resolução e nenhuma dificuldade particular na aplicação em arranjos de qualquer geometria ou para o caso multidimensional. Porém, tais procedimentos exigem um esforço computacional proibitivo [Stoica90b], [Krim96], [Kay93] e [Lemos97]. Assim, novos procedimentos surgiram para superar a grande carga computacional e ao mesmo tempo manter níveis próximos de desempenho.

Dentre esses novos procedimentos destacam-se, pela simplicidade computacional, aqueles baseados na predição linear. Tais procedimentos realizam a estimação do parâmetro DOA através de um mapeamento adequado entre ângulo elétrico e ângulo espacial. As desvantagens desses procedimentos, quando comparados com aqueles baseados no ML, consistem na limitação em aplicações com arranjos lineares e uniformes e na baixa capacidade de resolução quando aplicados em sinais fortemente correlacionados e em baixas relações sinal-ruído [Lemos97], [Antunes92], [Tufts82] e [Colares97].

Os procedimentos baseados em subespaços, também apresentados no Capítulo 3, são uma alternativa complementar aos procedimentos baseados na predição linear ou no critério ML. Tais procedimentos realizam a estimação do parâmetro DOA através do uso da decomposição dos dados disponíveis em subespaços de sinal e de ruído, bem como da propriedade de ortogonalidade entre os mesmos [Jansson99], [Cadzow90], [Reddi79] e [Pisarenko73].

Uma outra classe de procedimentos baseados no critério da “distância” entre subespaços (“Subspace Fitting-SF”) se destaca como competidora do critério ML por apresentar uma resolução equivalente [Ottersten92]. Esse critério consiste em obter uma melhor aproximação possível para um conjunto de dados ruidosos que possui a menor “distância” em relação a seu subespaço de sinal [Viberg91a], [Viberg91b] e [Krim96].

Uma outra forma de contornar o esforço computacional exigido pelo critério ML surge também através do uso de algoritmos numéricos mais eficientes [Bresler86] e [Stoica90a]. Por serem alternativas na implementação do critério ML, tais procedimentos apresentam desempenho equivalente ao ML. Uma das desvantagens dessa abordagem é que tais algoritmos são

geralmente iterativos e sofrem de problemas de convergência numérica [Stoica97].

Nem todos os métodos apresentados aqui são consagrados para aplicação em estimação do parâmetro DOA. No entanto, além de possibilitarem tal aplicação nos permitem uma visão mais ampla das várias alternativas.

Um dos pontos em comum entre todos esses procedimentos de estimação consiste na obtenção das estimativas através da minimização de uma função custo adequada. Por esse motivo, destacamos as alternativas de otimização nos Capítulos 4 e 5. Embora essas técnicas estejam restritas a modelos lineares, nos quais os parâmetros desejados estão relacionados linearmente com os dados disponíveis, também podem ser adotadas em problemas que envolvem uma relação não-linear mediante uma reparametrização adequada do problema original.

Embora as descrições de todos esses métodos estejam disponíveis na literatura, não se encontram em uma formulação unificada. Assim, é também objetivo desse capítulo apresentá-los através de uma descrição unificada, que permita estabelecer relações entre os mesmos e também explorar alternativas inéditas resultante destas relações. Essa abordagem possibilita uma melhor compreensão das características e desempenho de cada um dos métodos descritos. Nessa descrição apresentada aqui está uma das maiores contribuições desse trabalho.

O problema central que os métodos, a seguirem, pretendem resolver é estimar os M parâmetros DOA dos sinais que chegam em um arranjo receptor N -dimensional, baseando-se em um conjunto finito de amostras $\{y_0(k), y_1(k), \dots, y_{N-1}(k)\}$ contaminadas por ruído, para cada experimento k , num total de K experimentos, e considerando conhecido o número M de fontes de sinal.

6.2 Métodos Baseados no Critério ML

Os métodos apresentados nessa seção fazem uso do critério ML, implementando a busca multidimensional exigida pela sua minimização através de alternativas de menor esforço computacional.

6.2.1 Método IQML

O método da iteração quadrática para o ML (IQML), pode ser visto como uma alternativa ao procedimento de minimização do critério ML determinístico (MLD) apresentado no Capítulo 3. Esse método procura resolver o problema de minimização de quarta ordem do critério ML através da solução de um problema de segunda ordem, utilizando um algoritmo iterativo [Bresler86].

Considerando a função custo do critério ML determinístico em (3.5), esse método faz uso da reparametrização do projetor $\mathbf{\Pi}_{\mathbf{A}}^{\perp}$ através da relação em (3.33), utilizando a matriz \mathbf{B}^H de (3.29) com $L = M$. Desta forma, o critério ML pode ser reparametrizado para os coeficientes b_k , e as estimativas dos coeficientes podem ser calculadas minimizando-se a função custo apresentada em (3.34) e rerepresentada a seguir

$$J_{IQML}(\mathbf{b}') = Tr \left\{ \left[\mathbf{B}(\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H \right] \hat{\mathbf{R}} \right\}, \text{ com } L = M. \quad (6.1)$$

De uma forma mais genérica, o algoritmo IQML procura minimizar (6.1) resolvendo, em um primeiro passo, o seguinte problema mais simples

$$\hat{\mathbf{b}}' = \arg \left\{ \min_{\mathbf{b}'} Tr \left\{ \mathbf{B} \mathbf{B}^H \hat{\mathbf{R}} \right\} \right\}, \quad (6.2)$$

seguindo com a formação da matriz $\hat{\mathbf{B}}$ e do termo $(\hat{\mathbf{B}}^H \hat{\mathbf{B}})^{-1}$, presentes em (6.1), a partir das estimativas iniciais dadas por $\hat{\mathbf{b}}'$ em (6.2). Em um segundo passo, novas estimativas são obtidas, desta vez a partir de (6.1) com a substituição do termo $(\mathbf{B}^H \mathbf{B})^{-1}$ por sua estimativa $(\hat{\mathbf{B}}^H \hat{\mathbf{B}})^{-1}$ e considerando a simetria complexa conjugada dos coeficientes b_k , fazendo $b_m = b_{M-m}^*$ para $m = 0, 1, \dots, M$. Tal restrição aos coeficientes b_k será discutida posteriormente. Esse segundo passo continua iterativamente até atingir a convergência. O método tem como último passo o cálculo das raízes de $b(z)$ para a obtenção das estimativas $\hat{\phi}_m$, $m = 1, 2, \dots, M$.

Entrando em alguns detalhes da implementação, o método é construído através da igualdade proveniente da comutatividade da operação de convolução, apresentada abaixo

$$\mathbf{B}^H \mathbf{y}(k) = \mathbf{Y}(k) \mathbf{b}', \text{ para } k = 1, 2, \dots, K, \quad (6.3)$$

onde K representa o número total de experimentos (“snapshots”), e a matriz $\mathbf{Y}(k)$ é definida como

$$\mathbf{Y}(k) \triangleq \begin{bmatrix} y_M(k) & y_{M-1}(k) & \cdots & y_0(k) \\ y_{M+1}(k) & y_M(k) & \cdots & y_1(k) \\ \vdots & \vdots & & \vdots \\ y_{N-1}(k) & y_{N-2}(k) & \cdots & y_{N-M-1}(k) \end{bmatrix}_{N-M \times M+1}, \text{ para } k = 1, 2, \dots, K, \quad (6.4)$$

com $N > M$.

Substituindo $\hat{\mathbf{R}} \triangleq \frac{1}{K} \sum_{k=1}^K \mathbf{y}(k) \mathbf{y}^H(k)$ em (6.1) obtemos

$$J_{IQML}(\mathbf{b}') = Tr \left\{ \frac{1}{K} \sum_{k=1}^K \mathbf{B} (\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H \mathbf{y}(k) \mathbf{y}^H(k) \right\}. \quad (6.5)$$

Utilizando (6.3), considerando que $Tr \left\{ \left(\mathbf{B} (\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H \mathbf{y}(k) \right) \mathbf{y}^H(k) \right\} = \mathbf{y}^H(k) \left(\mathbf{B} (\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H \mathbf{y}(k) \right)$, e ainda desconsiderando o termo K à esquerda, teremos a função custo equivalente a (6.1) dada por

$$J_{IQML}(\mathbf{b}') = (\mathbf{b}')^H \left[\sum_{k=1}^K \mathbf{Y}^H(k) (\mathbf{B}^H \mathbf{B})^{-1} \mathbf{Y}(k) \right] (\mathbf{b}'), \quad (6.6)$$

onde Ψ representa o conjunto adequado de restrições impostas a estimativa $\hat{\mathbf{b}}'$.

Podemos observar que o problema em (6.6) é um problema de minimização de quarta ordem com restrições dadas por Ψ . No entanto, quando $\mathbf{B}^H \mathbf{B}$ é constante e independente de \mathbf{b}' , tal problema torna-se de segunda ordem. Neste caso, é possível uma minimização iterativa, em que cada passo consiste na solução de um problema de minimização de segunda ordem, no qual o termo $(\mathbf{B}^H \mathbf{B})^{-1}$ é atualizado através da estimativa anterior de \mathbf{b}' .

O conjunto Ψ adequado de restrições é construído através da interseção de duas restrições

$$\Psi = \Psi_0 \cap \Psi_s. \quad (6.7)$$

A primeira restrição, denominada de restrição de “não-trivialidade”, é definida como

$$\Psi_0 \triangleq \{\mathbf{b}' : \|\mathbf{b}'\|_2 \neq 0\}, \quad (6.8)$$

e consiste em evitar a solução trivial estabelecendo uma correspondência única entre b_m e ϕ_m , além de garantir a invertibilidade de $(\mathbf{B}^H \mathbf{B})$. A implementação de tal restrição pode ser através de uma relação linear fazendo algum elemento de \mathbf{b}' diferente de zero, considerando sua parte real e imaginária de forma distinta, ou através de uma relação quadrática na qual impomos um valor diferente de zero para a norma do vetor \mathbf{b}' .

A segunda restrição Ψ_s , escolhida de acordo com o modelo do sinal considerado, é essencial para a obtenção da solução ótima. Para o modelo de sinal composto pela soma de M exponenciais complexas não-amortecidas, apresentado no Capítulo 2, teremos as raízes do polinômio $b(z)$ sobre a CRU para o caso ótimo, ou seja

$$\Psi_s \triangleq \{\mathbf{b}' : b(z) \text{ possui raízes sobre a circunferência de raio unitário}\}.$$

Uma possibilidade de implementação para essa restrição consiste em estabelecer uma restrição de simetria complexa conjugada sobre os coeficientes b_k , fazendo $b_m = b_{M-m}^*$, $m = 0, \dots, M$. Apesar de restringir as possibilidades de escolha para os coeficientes b_k que satisfazem Ψ_s , tal implementação apresenta a vantagem de reduzir pela metade o número de parâmetros a serem obtidos. No entanto, não pode ser utilizada juntamente com a implementação da restrição de não trivialidade na qual $b_0 = 1$, pois um polinômio da forma $b(z) = b_0 \prod_{m=1}^M (z - e^{j\phi_m})$ não satisfaz, em geral, a relação $b_m = b_{M-m}^*$ com $b_0 = 1$. Neste caso, outra alternativa para implementação da restrição de não-trivialidade deve ser utilizada.

Como consequência da implementação anterior, podemos propor uma outra possibilidade de implementação da restrição Ψ_s . Tal alternativa consiste em considerar o polinômio $b(z)$ formado por coeficientes puramente reais, fazendo $Im(\mathbf{b}') = 0$, onde $Im(\bullet)$ indica a parte imaginária. Essa implementação apesar de satisfazer a condição de raízes sobre a CRU e reduzir o número de parâmetros pela metade, só deve ser aplicada no caso de sinais senoidais puros, sendo portanto mais restritiva que a implementação anterior. Nessa implementação podemos impor $b_0 = 1$ para evitar a solução trivial.

O algoritmo para a realização do método IQML, que minimiza a função custo em (6.6), considera a implementação da restrição de simetria complexa conjugada de forma implícita, e para evitar a solução trivial, faz $Re(b_0) = 1$, onde $Re(\bullet)$ indica a parte real.

Algoritmo IQML:

- I) Inicialização: $i = 0$ e $\mathbf{B}_{(i)}^H \mathbf{B}_{(i)} = \mathbf{I}$, onde o índice (i) indica o passo da iteração.
- II) Obter a matriz $\Omega_{IQML}^{(i)}$ de dimensões $[(M+1) \times (M+1)]$:

$$\Omega_{IQML}^{(i)} = \sum_{k=1}^K \mathbf{Y}^H(k) (\mathbf{B}_{(i)}^H \mathbf{B}_{(i)})^{-1} \mathbf{Y}(k), \quad (6.9)$$

onde $\mathbf{B}_{(i)}$ é construído a partir de $\mathbf{b}^{(i)}$ para $i > 0$.

III) Resolver o problema de minimização quadrática:

$$\hat{\mathbf{b}}'_{(i+1)} = \arg \left\{ \min_{\mathbf{b}' \in \Psi} \mathbf{b}'^H \boldsymbol{\Omega}_{IQML}^{(i)} \mathbf{b}' \right\}. \quad (6.10)$$

Neste passo se faz necessário a implementação das restrições. A implementação implícita das restrições converte o problema original (6.10) em um problema sem restrições equivalente. Apresentaremos a seguir a implementação implícita das restrições de não-trivialidade e simetria complexa conjugada para o caso de um sinal composto por M exponenciais complexas não-amortecidas.

Tomando $M = 2q + 1$ (para M par a implementação é similar), particiona-se \mathbf{b}' e $\mathbf{Y}(k)$ tal que

$$[b_0 \cdots b_q | b_{q+1} \cdots b_{2q+1}]^T = [\mathbf{b}_1^T | \mathbf{b}_2^T]^T \quad (6.11)$$

e

$$\mathbf{Y}(k) \begin{bmatrix} b_0 \\ \mathbf{b} \end{bmatrix} = [\mathbf{Y}_1(k) | \mathbf{Y}_2(k)] \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} = \mathbf{Y}_1(k) \mathbf{b}_1 + \mathbf{Y}_2(k) \mathbf{b}_2. \quad (6.12)$$

Definindo a matriz de permutação reversa $\tilde{\mathbf{I}}$,

$$\tilde{\mathbf{I}} = \begin{bmatrix} 0 & & & 1 \\ & & \ddots & \\ & & 1 & \\ & \ddots & & \\ 1 & & & 0 \end{bmatrix}_{q+1 \times q+1}, \quad (6.13)$$

a relação de simetria complexa conjugada pode ser representada por $\mathbf{b}_2 = \tilde{\mathbf{I}} \mathbf{b}_1^*$.

Definindo agora

$$\mathbf{c}_{IQML} = [\text{Re}(\mathbf{b}_1^T), \text{Im}(\mathbf{b}_1^T)]^T \quad (6.14)$$

e

$$\tilde{\mathbf{Y}}(k) = \left[\mathbf{Y}_1(k) + \mathbf{Y}_2(k) \tilde{\mathbf{I}} \mid j(\mathbf{Y}_1(k) - \mathbf{Y}_2(k) \tilde{\mathbf{I}}) \right], \quad (6.15)$$

o problema em (6.10) torna-se

$$\hat{\mathbf{c}}_{IQML(i+1)} = \arg \left\{ \min \mathbf{c}_{IQML}^T \left[\sum_{k=1}^K \tilde{\mathbf{Y}}^H(k) (\mathbf{B}_{(i)}^H \mathbf{B}_{(i)})^{-1} \tilde{\mathbf{Y}}(k) \right] \mathbf{c}_{IQML} \right\}, \text{ com } \mathbf{c}_{IQML} \in \mathcal{R}^{M+1}. \quad (6.16)$$

A restrição de não-trivialidade para a solução \mathbf{c}_{IQML} do problema em (6.16) pode ser implementada fazendo o primeiro elemento do vetor \mathbf{c}_{IQML} igual a unidade ($c_{IQML_0} = 1$), ou seja $\text{Re}(b_0) = 1$ em (6.14). A partir de $\hat{\mathbf{c}}_{IQML(i+1)}$ obtem-se $\hat{\mathbf{b}}'_{(i+1)}$ através da relação em (6.14).

Lembrando que $(\mathbf{B}_{(i)}^H \mathbf{B}_{(i)})^{-1}$ é *Hermitiana* e que, portanto, $\tilde{\mathbf{Y}}^H(k)(\mathbf{B}_{(i)}^H \mathbf{B}_{(i)})^{-1} \tilde{\mathbf{Y}}(k)$ é também *Hermitiana*, pode-se demonstrar que o problema em (6.16) é equivalente ao problema [Bresler86]

$$\hat{\mathbf{c}}_{IQML(i+1)} = \arg \left\{ \min_{\mathbf{c}_{IQML_0}=1} \mathbf{c}_{IQML}^T \boldsymbol{\Omega}_{IQML} \mathbf{c}_{IQML} \right\} \quad (6.17)$$

com

$$\boldsymbol{\Omega}_{IQML} = \sum_{k=1}^K \text{Re} \left\{ \tilde{\mathbf{Y}}^H(k)(\mathbf{B}_{(i)}^H \mathbf{B}_{(i)})^{-1} \tilde{\mathbf{Y}}(k) \right\}. \quad (6.18)$$

Desta forma, o problema em (6.17) é equivalente ao problema em (6.10) a menos da redução do esforço computacional requerido e pelo fato de trazer implicitamente a implementação das restrições impostas ao problema inicial. Por se tratar de um problema de minimização de segunda ordem para $(\mathbf{B}_{(i)}^H \mathbf{B}_{(i)})^{-1}$ fixo, podemos fazer uso dos procedimentos de minimização tratados no Capítulo 4 para a obtenção de sua solução.

IV) Incrementar o passo de iteração: $i = i + 1$.

V) Checar convergência: $\left\| \hat{\mathbf{b}}'_{(i-1)} - \hat{\mathbf{b}}'_{(i)} \right\|_2 < \varepsilon$? Se sim, termina e vai para (VI); se não, retornar para (II).

VI) Obtém as estimativas dos parâmetros do sinal $(\hat{\phi}_m, m = 1, 2, \dots, M)$: Calcular as raízes do polinômio $b(z)$, formado a partir de $\hat{\mathbf{c}}_{IQML(i)}$, terminando assim o algoritmo.

Na prática, apesar de apresentar bom desempenho, o método IQML sofre com problemas de instabilidade numérica e dificuldade de convergência para um mínimo global [Bresler86], [Stoica90a] e [Stoica97].

A restrição de simetria complexa conjugada imposta aos coeficientes de $b(z)$, a escolha da ordem do modelo de reparametrização igual ao número de parâmetros ($L = M$), bem como o uso da matriz de correlação estimada, dada pelo produto explícito do vetor de amostras disponíveis, são características próprias do método IQML que serão contrastadas com outros métodos.

Podemos concluir que as duas idéias mais relevantes nesse método são o uso da reparametrização do critério ML e a utilização de uma forma iterativa para resolver um problema de minimização de quarta ordem com restrições a partir da solução de um problema de segunda ordem sem restrições.

6.2.2 Método MODE

O método MODE pretende superar as desvantagens observadas no método IQML. Tal método pode ser interpretado como um aprimoramento das idéias do método IQML, no qual se obtém melhores estimativas no passo inicial, levando a uma rápida convergência. Além disso, no seu desenvolvimento faz-se uso dos subespaços definidos pela matriz de correlação conforme a versão do critério ML baseada em subespaços, apresentada no Capítulo 3, Seção 3.2.3 [Stoica90a].

Mais especificamente, a idéia básica do MODE é minimizar a função custo

$$J(\mathbf{b}') = \underset{\mathbf{b}' \simeq \mathbf{b}'_0}{Tr} \left[\mathbf{B}(\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H \hat{\mathbf{R}} \right], \quad (6.19)$$

localmente em torno de \mathbf{b}'_o , onde \mathbf{b}' e \mathbf{b}'_o indicam, respectivamente, o parâmetro desconhecido e seu valor ótimo. Desta forma, o MODE pretende contornar os problemas de convergência e reduzir as iterações necessárias para alcançar as estimativas finais.

Considerando uma situação assintótica para o número de experimentos ($K \rightarrow \infty$) e a hipótese de ergodicidade para o sinal observado, poderemos substituir sua correlação estatística \mathbf{R} pela correlação das amostras, $\hat{\mathbf{R}}$. Partindo-se da decomposição em subespaços da matriz $\hat{\mathbf{R}}$, conforme realizado para \mathbf{R} em (2.24) e (2.25), observando ainda que $\hat{\mathbf{U}}_s \hat{\mathbf{U}}_s^H + \hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H = \mathbf{I}$, poderemos escrever a função custo (6.19) localmente como [Stoica90a]

$$J(\mathbf{b}') \underset{\mathbf{b}' \simeq \mathbf{b}'_o}{\approx} Tr \hat{\mathbf{\Lambda}}_n + Tr \left\{ \left[\hat{\mathbf{U}}_s^H \quad \mathbf{\Pi}_B \quad \hat{\mathbf{U}}_s \right] \left[\hat{\mathbf{\Lambda}}_s - \sigma^2 \mathbf{I} \right] \right\}, \quad (6.20)$$

com $\hat{\mathbf{\Lambda}}_n \approx \sigma^2 \mathbf{I}$, uma vez que $\hat{\mathbf{\Lambda}}_n$ é um estimador consistente para $\sigma^2 \mathbf{I}$.

Portanto, estamos interessados na obtenção da estimativa $\hat{\mathbf{b}}'$ considerando o comportamento local de $J(\mathbf{b}')$ em (6.20), na qual $\left\| \hat{\mathbf{b}}' - \mathbf{b}'_o \right\|_2 \approx 0$ para K grande. Isso implica que $\left\| \hat{\mathbf{B}} - \mathbf{B}_o \right\|_2 \approx 0$ e, conseqüentemente, $\hat{\mathbf{B}}^H \hat{\mathbf{U}}_s = \hat{\mathbf{U}}_s^H \hat{\mathbf{B}} = \mathbf{0}$ para K grande [Stoica89]. Desta forma, o estimador MODE será dado pela minimização da função custo equivalente a (6.20)

$$J(\mathbf{b}') = Tr \left\{ \left[\hat{\mathbf{U}}_s^H \quad \mathbf{\Pi}_B \quad \hat{\mathbf{U}}_s \right] \left[\hat{\mathbf{\Lambda}}_s - \hat{\sigma}^2 \mathbf{I} \right] \right\}, \quad (6.21)$$

com $\mathbf{\Pi}_B = \mathbf{B}(\hat{\mathbf{B}}^H \hat{\mathbf{B}})^{-1} \mathbf{B}^H$ e na qual $\hat{\sigma}^2$ e $\hat{\mathbf{B}}$ são estimativas consistentes de σ^2 e \mathbf{B} , respectivamente.

Uma estimativa consistente de σ^2 pode ser obtida através de

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{N-M} \left\{ Tr \left\{ \hat{\mathbf{R}} \right\} - Tr \left\{ \hat{\mathbf{\Lambda}}_s \right\} \right\} \\ &= \frac{1}{N-M} Tr[\hat{\mathbf{\Lambda}}_n], \end{aligned}$$

$\hat{\mathbf{B}}$ e $\hat{\mathbf{b}}'$, por sua vez, podem ser determinados através da minimização da função quadrática

$$J(\mathbf{b}') = Tr \left\{ \mathbf{B} \mathbf{B}^H \hat{\mathbf{U}}_s \mathbf{W}_{MODE} \hat{\mathbf{U}}_s^H \right\}, \quad (6.22)$$

na qual \mathbf{W}_{MODE} é uma matriz de ponderação escolhida adequadamente que generaliza o método MODE, dando origem a generalização da função custo em (6.21), dada por

$$J_{MODE}(\mathbf{b}') = Tr \left\{ \left[\hat{\mathbf{U}}_s^H \quad \mathbf{\Pi}_B \quad \hat{\mathbf{U}}_s \right] \mathbf{W}_{MODE} \right\}. \quad (6.23)$$

A consistência da estimativa obtida através de (6.23) se justifica pelo fato da expressão alcançar seu mínimo valor apenas quando $\mathbf{B}^H \hat{\mathbf{U}}_s = \mathbf{0}$ e para \mathbf{W}_{MODE} positiva definida.

Como $\hat{\mathbf{U}}_s$ é geralmente estimada com precisão, algumas possibilidades para \mathbf{W}_{MODE} são [Li98]

$$\mathbf{W}_{mode1} = \hat{\mathbf{\Lambda}}_s,$$

e

$$\mathbf{W}_{mode2} = \hat{\mathbf{\Lambda}}_s - \hat{\sigma}^2 \mathbf{I}.$$

Demonstra-se que a escolha ótima para \mathbf{W}_{MODE} , \mathbf{W}_o , sob o ponto de vista de minimização da variância do estimador, é dada por [Stoica90a], [Krim96] e [Li98]

$$\mathbf{W}_o = (\hat{\mathbf{\Lambda}}_s - \hat{\sigma}^2 \mathbf{I})^2 \hat{\mathbf{\Lambda}}_s^{-1}. \quad (6.24)$$

A vantagem da utilização de (6.24) para \mathbf{W}_{MODE} em (6.22) é que as estimativas se aproximam dos seus valores assintóticos, dando origem a um procedimento direto para obtenção da estimativa $\hat{\mathbf{b}}'$, em dois passos.

No primeiro passo resolve-se o problema quadrático:

$$\hat{\mathbf{b}}' = \arg \left\{ \min_{\mathbf{b}'} Tr \left\{ \mathbf{B} \mathbf{B}^H \hat{\mathbf{U}}_s \mathbf{W}_o \hat{\mathbf{U}}_s^H \right\} \right\}, \quad (6.25)$$

do qual obtem-se $\hat{\mathbf{B}}$.

No segundo passo resolve-se o problema quadrático:

$$\hat{\mathbf{b}}' = \arg \left\{ \min_{\mathbf{b}'} Tr \left\{ \hat{\mathbf{U}}_s^H \mathbf{B} (\hat{\mathbf{B}} \hat{\mathbf{B}})^{-1} \mathbf{B}^H \hat{\mathbf{U}}_s \mathbf{W}_o \right\} \right\}, \quad (6.26)$$

com $\hat{\mathbf{B}}$ dado pelo primeiro passo.

As estimativas finais, $\hat{\phi}_m$, $m = 1, 2, \dots, M$, são então obtidas através das raízes do polinômio $b(z)$ formado pelos elementos de $\hat{\mathbf{b}}'$ obtido no último passo.

Comparando as expressões (6.25) e (6.26), respectivamente, com (6.1) e (6.2), poderemos contrastar os métodos MODE e IQML. Inicialmente, devemos constatar que

$$Tr \left\{ \hat{\mathbf{U}}_s^H \mathbf{B} (\hat{\mathbf{B}} \hat{\mathbf{B}})^{-1} \mathbf{B}^H \hat{\mathbf{U}}_s \mathbf{W}_o \right\} = Tr \left\{ \mathbf{B} (\hat{\mathbf{B}} \hat{\mathbf{B}})^{-1} \mathbf{B}^H \hat{\mathbf{U}}_s \mathbf{W}_o \hat{\mathbf{U}}_s^H \right\},$$

e que $\hat{\mathbf{U}}_s \mathbf{W}_o \hat{\mathbf{U}}_s^H$ é uma estimativa da parcela de $\hat{\mathbf{R}}$ associada ao subespaço de sinal. Assim, a comparação direta entre as expressões citadas nos permite concluir que os métodos diferem pelo fato do método MODE operar sobre a parcela do subespaço de sinal de $\hat{\mathbf{R}}$, com o objetivo de eliminar os efeitos indesejáveis da presença do ruído. Dessa forma torna-se possível a obtenção de uma melhor estimativa inicial para \mathbf{b}' , que leva, em geral, a estimativa final no segundo passo.

Podemos ainda notar que a minimização em (6.25) corresponde ao primeiro passo do método IQML quando $(\hat{\mathbf{B}} \hat{\mathbf{B}})^{-1} = \mathbf{I}$ e $\hat{\mathbf{R}} \cong \hat{\mathbf{U}}_s \mathbf{W}_o \hat{\mathbf{U}}_s^H$. Nesse caso, a matriz de correlação não é mais obtida pelo produto explícito do vetor de amostras como no método IQML, mas dada pela sua versão decomposta em subespaço, da qual apenas a porção associada ao sinal é considerada.

Além disso, é utilizada uma estimativa mais próxima do valor ótimo para os autovalores associado ao subespaço de sinal, dada por \mathbf{W}_o . Dessa forma, o método MODE procura avaliar a função custo em (6.19) nas proximidades do seu valor ótimo, evitando assim os passos de iteração presentes no método IQML e seus problemas de convergência.

Embora o método MODE apresente uma redução no esforço computacional em relação ao IQML devido a eliminação das iterações, se faz necessário a obtenção de estimativas adicionais através da decomposição da matriz de correlação. Assim sendo, seu sucesso depende da qualidade dessas estimativas.

Concluimos, portanto, que a utilização da forma decomposta em subespaços da matriz de correlação e da distinção entre os subespaços de sinal e de ruído apresenta vantagens sobre a forma dada pelo produto explícito dos vetor de amostras, desde que esses subespaços possam ser corretamente estimados.

Por apresentar a melhor precisão assintótica possível e não exigir grande número de iterações, tal método é um forte candidato a melhor método para arranjos lineares uniformes.

O algoritmo de solução para o método MODE considera implementação implícita da restrição de simetria complexo conjugada sobre os coeficientes de $b(z)$, além da restrição de não-trivialidade fazendo $Re(b_0) = 1$.

Para a restrição de simetria complexo conjugada, o algoritmo MODE sugere sua implementação implícita assegurando que os elementos de \mathbf{b}' satisfaçam

$$b_m = b_{M-m}^*, \text{ para } m = 0, \dots, M. \quad (6.27)$$

Quanto à incompatibilidade da aplicação da restrição da forma em (6.27) com a restrição de não-trivialidade, fazendo $b_0 = 1$, valem os mesmos comentários apresentados para o método IQML.

Com base em (6.25) e (6.26), o algoritmo MODE estima os parâmetros ϕ_m , $m = 1, 2, \dots, M$, do sinal utilizando uma estimativa $\hat{\mathbf{U}}_s$, juntamente com uma estimativa dos \bar{r} ($\bar{r} = \min\{M, \text{posto}(\hat{\mathbf{P}})\}$) maiores autovalores de $\hat{\mathbf{R}}$ e considerando a estimativa da matriz \mathbf{W}_o dada por [Li98]

$$\hat{\mathbf{W}}_o = \text{diag} \left\{ \frac{(\hat{\lambda}_1 - \hat{\sigma}^2)^2}{\hat{\lambda}_1}, \frac{(\hat{\lambda}_2 - \hat{\sigma}^2)^2}{\hat{\lambda}_2}, \dots, \frac{(\hat{\lambda}_{\bar{r}} - \hat{\sigma}^2)^2}{\hat{\lambda}_{\bar{r}}} \right\} \quad (6.28)$$

com

$$\hat{\sigma}^2 = \frac{1}{N - \bar{r}} \left\{ \text{Tr}\{\hat{\mathbf{R}}\} - \sum_{i=1}^{\bar{r}} \hat{\lambda}_i \right\}. \quad (6.29)$$

Algoritmo MODE

O algoritmo MODE compreende, portanto, os seguintes passos:

- I) Obter $\hat{\mathbf{R}}$ e seus \bar{r} maiores autovalores e autovetores correspondentes. Formar $\hat{\mathbf{U}}_s$ e $\hat{\mathbf{W}}_o$.

II) Determinar uma estimativa inicial para o vetor \mathbf{b}' , dada por $\hat{\mathbf{b}}'_{(0)}$, minimizando-se a expressão em (6.25) impondo as restrições em (6.27).

III) Determinar a estimativa final $\hat{\mathbf{b}}'$, minimizando-se a expressão (6.26) sujeita as restrições em (6.27), na qual o termo $(\hat{\mathbf{B}}^H \hat{\mathbf{B}})^{-1}$ é formado através da estimativa $\hat{\mathbf{b}}'_{(0)}$ obtida no passo (II).

IV) Obter as estimativas $\hat{\phi}_m$, $m = 1, 2, \dots, M$, através das raízes do polinômio $b(z)$ formado pelos elementos de $\hat{\mathbf{b}}'$ obtido no passo (III).

Para obter o resultado da minimização em (6.26) de forma mais eficiente, o algoritmo MODE utiliza a forma compacta a seguir [Stoica90a] e [Li98]

$$J(\mathbf{b}')_{MODE} = \|\mathbf{H}\mathbf{b}'\|_2^2 \quad (6.30)$$

onde

$$\mathbf{H} = \begin{bmatrix} (\hat{\mathbf{B}}^H \hat{\mathbf{B}})^{-1/2} \tilde{\mathbf{U}}_1 \\ \vdots \\ (\hat{\mathbf{B}}^H \hat{\mathbf{B}})^{-1/2} \tilde{\mathbf{U}}_M \end{bmatrix} \quad (6.31)$$

com

$$\tilde{\mathbf{U}}_m = \begin{bmatrix} \tilde{u}_{M+1,m} & \cdots & \tilde{u}_{1,m} \\ \cdots & \cdots & \cdots \\ \tilde{u}_{N,m} & \cdots & \tilde{u}_{N-M,m} \end{bmatrix}, \quad m = 1, 2, \dots, \bar{r} \quad (6.32)$$

e

$$[\tilde{\mathbf{s}}_1 \cdots \tilde{\mathbf{s}}_{\bar{r}}] \triangleq \begin{bmatrix} \tilde{u}_{1,1} & \cdots & \tilde{u}_{1,\bar{r}} \\ \cdots & \cdots & \cdots \\ \tilde{u}_{N,1} & \cdots & \tilde{u}_{N,\bar{r}} \end{bmatrix} \triangleq \hat{\mathbf{U}}_s (\hat{\mathbf{W}}_o)^{1/2}. \quad (6.33)$$

Dessa forma, a estrutura especial da matriz $\hat{\mathbf{B}}^H \hat{\mathbf{B}}$ torna possível a utilização de algoritmos eficientes para computar $(\hat{\mathbf{B}}^H \hat{\mathbf{B}})^{-1/2}$ [Dickinson79]. Além disso, a estrutura triangular da matriz $(\hat{\mathbf{B}}^H \hat{\mathbf{B}})^{-1/2}$ reduz o esforço computacional para avaliar \mathbf{H} em (6.31).

Após a reparametrização em (6.30), a restrição de simetria complexa conjugada da forma em (6.27) é implementada implicitamente no problema de minimização através do uso da matriz $\tilde{\mathbf{I}}$ de (6.13) e do vetor

$$\beta = [b_0, b_1, \dots, b_q]^T, \quad (6.34)$$

no qual q é o maior inteiro menor ou igual a $[(M-1)/2]$, que nos permite expressar tal restrição como

$$\mathbf{b}' = \begin{bmatrix} \beta \\ \tilde{\mathbf{I}} \beta^* \end{bmatrix} \quad \text{para } M \text{ par}$$

ou

$$\mathbf{b}' = \begin{bmatrix} \beta \\ \mu \\ \tilde{\mathbf{I}} \beta^* \end{bmatrix} \quad \text{para } M \text{ ímpar,}$$
(6.35)

na qual μ é um escalar real. Podemos observar que essa implementação é a mesma adotado no método IQML, reconhecendo $\mathbf{b}_1 = \beta$ e $\mathbf{b}_2 = \tilde{\mathbf{I}} \beta^*$ em (6.11).

O algoritmo segue decompondo-se a matriz \mathbf{H} da forma

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 \end{bmatrix} \quad \text{para } M \text{ par}$$

ou

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{h} & \mathbf{H}_2 \end{bmatrix} \quad \text{para } M \text{ ímpar,}$$
(6.36)

nas quais \mathbf{H}_1 e \mathbf{H}_2 são ambas de dimensão $[M(N - M)] \times [M/2]$.

Agora, substituindo (6.36) e (6.35) em (6.30), teremos

$$J(\rho)_{MODE} = \left\| \begin{bmatrix} Re(\mathbf{H}_1) + Re(\mathbf{H}_2 \tilde{\mathbf{I}}) & Im(\mathbf{H}_2 \tilde{\mathbf{I}}) - Im(\mathbf{H}_1) \\ Im(\mathbf{H}_1) + Im(\mathbf{H}_2 \tilde{\mathbf{I}}) & Re(\mathbf{H}_1) - Re(\mathbf{H}_2 \tilde{\mathbf{I}}) \end{bmatrix} \begin{bmatrix} Re(\beta) \\ Im(\beta) \end{bmatrix} \right\|_2^2$$

$\triangleq \|\mathbf{F}\rho\|_2^2 \quad \text{para } M \text{ par}$

ou

$$J(\rho)_{MODE} = \left\| \begin{bmatrix} Re(\mathbf{H}_1) + Re(\mathbf{H}_2 \tilde{\mathbf{I}}) & Im(\mathbf{H}_2 \tilde{\mathbf{I}}) - Im(\mathbf{H}_1) & Re(\mathbf{h}) \\ Im(\mathbf{H}_1) + Im(\mathbf{H}_2 \tilde{\mathbf{I}}) & Re(\mathbf{H}_1) - Re(\mathbf{H}_2 \tilde{\mathbf{I}}) & Im(\mathbf{h}) \end{bmatrix} \begin{bmatrix} Re(\beta) \\ Im(\beta) \\ \mu \end{bmatrix} \right\|_2^2$$

$\triangleq \|\mathbf{F}\rho\|_2^2 \quad \text{para } M \text{ ímpar.}$

(6.37)

Desta forma, o problema de minimização quadrática com restrições em (6.26) é convertido em um problema equivalente sem restrições e dimensões menores em (6.37). O algoritmo MODE sugere ainda uma restrição ao primeiro elemento do vetor β , b_0 , da forma

$$Re(b_0) = 1 \quad \text{ou} \quad Im(b_0) = 1, \quad (6.38)$$

evitando assim a solução trivial para β e a incompatibilidade do uso $b_0 = 1$. Para o caso em que $Re(b_0) = 1$, como sugerido no método IQML, exclui-se as situações nas quais a soma dos parâmetros $\{\phi_m\}$ é um inteiro ímpar múltiplo de π . Nessas situações pode-se, portanto, utilizar $Im(b_0) = 1$, que por sua vez exclui as situações nas quais a soma dos parâmetros $\{\phi_m\}$ é um inteiro par múltiplo de π . Visto que tais formas não são válidas apenas em situações mutuamente exclusivas, o uso de uma ou outra forma de acordo com a situação será válida para todas as circunstâncias.

Para a minimização de (6.37), o algoritmo MODE sugere a utilização do critério LS implementado através da decomposição Q-R, apresentado no Capítulo 4, Seção 4.4.1 [Stoica90a]. Tal procedimento mostra-se mais eficiente que o procedimento de solução direta para o conjunto de equações derivado a partir de (6.37).

Deve ser notado que uma vez que a função custo associada ao problema de minimização em (6.10) encontrado pelo método IQML, a cada passo de iteração, pode ser escrito na forma compacta

$$J_{IQML}(\mathbf{b}') = \left\| (\boldsymbol{\Omega}_{IQML})^{1/2} \mathbf{b}' \right\|_2^2,$$

que é semelhante em sua formulação ao problema do método MODE em (6.30), os procedimentos de solução sugerido pelos dois algoritmos podem ser aplicados a ambos indistintamente.

Apesar de ser considerado um algoritmo não-iterativo, o algoritmo MODE está naturalmente preparado para ser iterado, levando, em geral, a estimativas mais próximas de seus valores ótimos [Stoica90a].

6.3 Métodos Baseados na Predição Linear

Os métodos apresentados nessa seção fazem uso de um preditor linear de ordem L para estimar M parâmetros desconhecidos de um sinal, utilizando um conjunto de N amostras a cada experimento. Embora a estrutura do preditor linear tenha sido derivada no Capítulo 3 para o caso de um único experimento, a sua extensão para aplicações com vários experimentos também é possível e será apresentada.

Os métodos baseados na predição linear se relacionam com os métodos baseados no critério ML determinístico através da reparametrização do critério ML, quando fazemos uso de $\boldsymbol{\Pi}_A^\perp = \mathbf{B}(\mathbf{B}^H\mathbf{B})^{-1}\mathbf{B}^H$, e através da relação (6.3), reescrita aqui

$$\mathbf{B}^H \mathbf{y}(k) = \mathbf{Y}(k) \begin{bmatrix} b_0 \\ \mathbf{b} \end{bmatrix}, \text{ para } k = 1, 2, \dots, K, \text{ com } \mathbf{Y}(k) \stackrel{L \times M}{=} \mathbf{D}'(k),$$

na qual $\mathbf{D}'(k)$ corresponde à matriz de dados aumentada \mathbf{D}' da predição linear, apresentada no Capítulo 3, em que explicitamos a dependência com k . Lembramos que na predição linear temos ainda a restrição $b_0 = -1$.

Devemos acrescentar que o critério ML, diferente do conceito de predição linear, não está limitado a aplicação em arranjos de sensores lineares e uniformes [Krim96].

6.3.1 Método de Prony

O método desenvolvido por Prony pode ser interpretado como um processo de predição linear no qual a ordem do filtro é igual ao número de fontes de sinais e o número de amostras disponíveis é igual ao dobro do número de fontes de sinal ($L = M$ e $N = 2M$) [Kumaresa84]. Assim, a matriz de dados \mathbf{D} será quadrada e terá posto cheio permitindo uma solução direta do conjunto de equações $\mathbf{D}\mathbf{b} = \mathbf{d}$ através de sua inversa, ou seja $\hat{\mathbf{b}} = \mathbf{D}^{-1}\mathbf{d}$, e com $b_0 = -1$ em $b(z)$.

O método de Prony oferece uma solução exata e igual à do ML para o caso de um sinal sem ruído [Bresler86]. No entanto, é muito sensível à presença de ruído, mesmo em pequena quantidade.

De uma forma geral, a função custo a ser minimizada nesse método é dada por (3.22), reescrita aqui como

$$J_{\text{Prony}}(\mathbf{b}) = \left\{ \begin{bmatrix} b_0 \\ \mathbf{b} \end{bmatrix}^H \begin{bmatrix} \mathbf{d}:\mathbf{D} \\ \mathbf{d}:\mathbf{D} \end{bmatrix} \begin{bmatrix} b_0 \\ \mathbf{b} \end{bmatrix} \right\}, \text{ com } b_0 = -1. \quad (6.39)$$

Por se tratar de um problema de minimização de segunda ordem podemos fazer uso do procedimento LS de minimização, de onde derivamos a solução geral a seguir

$$\hat{\mathbf{b}}_{\text{Prony}} = (\mathbf{D}^H \mathbf{D})^{-1} \mathbf{D}^H \mathbf{d} = \mathbf{D}^\dagger \mathbf{d},$$

com $L = M$.

Reconhecendo a estimativa da matriz de correlação $\hat{\mathbf{R}} = \begin{bmatrix} \mathbf{d}:\mathbf{D} \\ \mathbf{d}:\mathbf{D} \end{bmatrix}$ em (6.39), poderemos contrastar esse método com os métodos baseados na matriz de correlação estimada como o ML. Para o caso particular em que temos um único experimento ($K = 1$) e $L = M$, podemos constatar que $\begin{bmatrix} \mathbf{d}:\mathbf{D} \\ \mathbf{d}:\mathbf{D} \end{bmatrix} = \mathbf{Y}$, definido em (6.4). Assim, poderemos reescrever (6.39) como

$$J_{\text{Prony}}(\mathbf{b}) = \begin{bmatrix} b_0 \\ \mathbf{b} \end{bmatrix}^H \mathbf{Y}^H \mathbf{Y} \begin{bmatrix} b_0 \\ \mathbf{b} \end{bmatrix}. \quad (6.40)$$

A extensão do método de Prony para a situação em que dispomos de múltiplos experimentos (K “snapshots”) pode ser obtida naturalmente substituindo-se a função custo em (6.40) por sua generalização

$$J(\mathbf{b}) = \begin{bmatrix} b_0 \\ \mathbf{b} \end{bmatrix}^H \mathbf{K}_y \begin{bmatrix} b_0 \\ \mathbf{b} \end{bmatrix}, \quad (6.41)$$

onde $\mathbf{K}_y \triangleq \sum_{k=1}^K \mathbf{Y}(k)^H \mathbf{Y}(k)$ possui dimensões $(L+1) \times (L+1)$ e possui um caráter de matriz de correlação.

O uso de múltiplos experimentos para a formação da matriz \mathbf{K}_y permite combater as dificuldades de resolução desse método na presença de sinais coerentes [Shan85].

Podemos agora comparar (6.41) com a função custo associada ao método IQML em (6.6). Dessa comparação podemos concluir que a minimização da função custo da forma em (6.41) com a restrição $b_0 = -1$, não apresentará solução ótima, visto que a condição para uma solução ótima, apresentada pelo método IQML, exige que o polinômio $b(z)$ apresente suas raízes sobre a CRU, considerando um sinal composto por exponenciais complexas não-amortecidas. Além disso, notamos a ausência da matriz $(\mathbf{B}^H \mathbf{B})^{-1}$ em (6.41), presente na formulação do IQML e útil na compensação da presença de correlação entre as fontes de sinal. Dessa forma (6.41) se assemelha à função custo presente no primeiro passo do algoritmo IQML, quando temos $(\mathbf{B}^H \mathbf{B})^{-1} = \mathbf{I}$.

6.3.2 Métodos FLP, BLP e FBLP

Melhores resultados na presença de ruído, com relação ao método de Prony, podem ser obtidos através da extensão da ordem do filtro e do número de amostras disponíveis. Isto é, considerando $L \geq M$ e $N > 2M$, dando origem aos métodos FLP, BLP e FBLP. Nesses casos,

o sistema linear de equações $\mathbf{D}\mathbf{b} = \mathbf{d}$ não possuirá uma única solução levando à necessidade de uma restrição adicional para a escolha da solução ótima.

Os métodos FLP e BLP, precursores do FBLP, consistem na utilização de um filtro preditor com estrutura progressiva (“forward”) e regressiva (“backward”), respectivamente. Conseqüentemente, a composição da matriz de dados em cada método é diferente [Antunes92], [Colares97] e [Lemos97].

A diferença de ordem prática entre o método FLP e o método de Prony é que esse último limita a ordem do preditor utilizado, fazendo $L = M$. A utilização de ordens mais elevadas que o número de fonte de sinais, no intuito de melhorar o desempenho do método, pode ser justificada, visto que quanto maior a ordem do preditor maior será a quantidade de amostras utilizadas para estimar a amostra desejada. Além disso, o polinômio resultante possuirá $L - M$ raízes adicionais, fornecendo, assim, um maior grau de liberdade para acomodar as M raízes de interesse [Tufts82].

O método FBLP, por sua vez, utiliza um procedimento de predição na forma progressiva e regressiva para um mesmo conjunto de dados, otimizados conjuntamente. Este método fornece melhores resultados que os métodos FLP e BLP por fazer melhor uso dos dados disponíveis. Como desvantagem, esse método está limitado às aplicações na qual o sinal é composto por exponenciais complexas não amortecidas. Devemos observar que para o método FBLP, a matriz de dados possui dimensões $2(N - L) \times L$, enquanto que para os métodos FLP e BLP sua dimensão é $(N - L) \times L$ [Tufts82].

As expressões para o método FBLP são, portanto, as mesmas do método de Prony com a ordem do preditor estendida e a estrutura da matriz de dados alterada para comportar os dados na forma progressiva e regressiva. O filtro progressivo-regressivo de predição linear possui ordem que satisfaz a desigualdade $M \leq L \leq N - M/2$. Caso $(M/2)$ seja racional, este valor é arredondado para o valor inteiro mais próximo. Neste caso, o número de equações será igual a $2(N - L)$ e sua solução segundo o critério LS é tomada como aquela que apresenta a menor norma *Euclidiana*. Demonstra-se, experimentalmente, que tal método apresenta melhor desempenho quando $L \cong N/3$ [Tufts82].

A solução de norma mínima dada pelo procedimento LS para os métodos FLP, BLP e FBLP pode ser expressa através da *pseudo-inversa* da matriz de dados associada a estrutura de cada filtro, dada genericamente como

$$\hat{\mathbf{b}}_{LP} = \mathbf{D}_{LP}^\dagger \mathbf{d}_{LP} = \sum_{i=1}^L \frac{1}{\sigma_{LP(i)}} \mathbf{v}_{LP(i)} \left(\mathbf{u}_{LP(i)}^H \mathbf{d}_{LP} \right), \text{ com } b_0 = -1, \quad (6.42)$$

na qual \mathbf{D}_{LP} , \mathbf{d}_{LP} , σ_{LP} , \mathbf{v}_{LP} e \mathbf{u}_{LP} denotam, respectivamente, as versões genéricas da matriz de dados, do vetor de observações, dos valores singulares e vetores singulares associados à matriz de dados nas formas adequadas para cada método (FLP, BLP e FBLP).

Quando a ordem L é igual ao seu limite superior ($N - M$ para os métodos FLP e BLP; $N - M/2$ para o FBLP), obtemos um caso de especial interesse, devido à sua simplicidade computacional e desempenho satisfatório, denominado de caso Kumaresan-Prony (KP) [Colares97], [Tufts82] e [Antunes92].

A única vantagem aparente desses métodos sobre o IQML consiste na possibilidade de utilização de uma ordem $L > M$, que em geral melhora a qualidade das estimativas obtidas.

6.3.3 Método FBLP Modificado

O método FBLP modificado pode ser interpretado como uma versão baseada em subespaços do método FBLP.

Apesar de permitir o aumento da ordem do preditor em relação ao método de Prony, a utilização de ordens $L > M$ no método FBLP implica também na adição de uma maior quantidade de ruído para obtenção da solução, levando a uma limitação de ordem prática para a escolha de L [Colares98]. Essa influência do ruído pode ser notada através de (6.42), na qual todos os valores e vetores singulares, tanto de sinal como de ruído, da matriz de dados \mathbf{D}_{LP} participam da solução. Assim, dado que temos apenas M vetores associados ao sinal, então o aumento de L leva a um aumento do número de vetores de ruído na obtenção da solução.

O método FBLP modificado é, portanto, uma variação do método FBLP na qual é utilizada uma aproximação da matriz de dados por uma versão de posto reduzido obtida através da ferramenta de decomposição em valores singulares e considerações com respeito aos subespaços de sinal e ruído associados à essa matriz. A dimensão do novo posto é escolhida igual ao número de parâmetros do sinal observado (M). Essa versão de posto reduzido da matriz de dados é obtida através da supressão dos valores e vetores singulares associados ao subespaço de ruído, produzindo um efeito semelhante a um aumento da relação sinal-ruído. Dessa forma, tal método permite o aumento da ordem do preditor para valores além daqueles que são apropriados para método FBLP, melhorando assim sua capacidade de resolução. Pode ser demonstrado experimentalmente que tal método apresenta um melhor desempenho quando a ordem L do filtro é aproximadamente igual a $3N/4$ [Tufts82] e [Antunes92].

Mais uma vez, fazendo uso do procedimento LS de minimização, a solução para esse método é dada por

$$\hat{\mathbf{b}}_{FBLP-M} = \sum_{i=1}^M \frac{1}{\sigma_{FBLP(i)}} \mathbf{v}_{FBLP(i)} \left(\mathbf{u}_{FBLP(i)}^H \mathbf{d}_{FBLP} \right), \text{ com } b_0 = -1,$$

na qual σ_{FBLP} , \mathbf{v}_{FBLP} e \mathbf{u}_{FBLP} denotam, respectivamente, os valores singulares e vetores singulares associados à matriz de dados na forma progressiva-regressiva.

Podemos notar que para o caso em que $L = N - M/2$ tal método é equivalente ao método FBLP no caso KP, pois nesse caso a supressão do subespaço de ruído na solução FBLP é realizada automaticamente devido às ordens envolvidas.

6.3.4 Método TLS-LP

O método, ou conjunto de métodos, aqui denominado de TLS-LP consiste em obter a solução para o preditor linear através do procedimento TLS de minimização [DeMoor90] e [Lemos97].

A solução de norma mínima fornecida pelo procedimento TLS, apresentado no Capítulo 4, consiste em um vetor associado ao menor valor singular da matriz de dados aumentada. Se esse menor valor singular for um valor singular repetido, a solução será dada, então, pela combinação dos vetores singulares associados aos menores valores singulares repetidos. As soluções possíveis para o problema de predição linear através do procedimento TLS foram apresentadas no Capítulo 4 para uma matriz de dados aumentada, genérica, \mathbf{C} . No contexto da

predição linear teremos $\mathbf{C} = [\mathbf{d}:\mathbf{D}]$. Para permitir uma comparação com os métodos baseados no critério ML podemos observar que $[\mathbf{d}:\mathbf{D}] = \mathbf{Y}$ para $L = M$ e $K = 1$.

Além das diferenças na concepção da solução obtida pelos procedimentos LS e TLS, o método LS opera sobre a matriz de dados \mathbf{D} da predição linear, enquanto o TLS opera sobre essa matriz aumentada pelo vetor de observação \mathbf{d} . Se ainda considerarmos a decomposição em valores e vetores singulares da matriz aumentada, poderemos interpretar a solução TLS para o preditor linear como uma solução baseada no subespaço de ruído dessa matriz, enquanto que a solução LS com restrição no subespaço da matriz de dados original, apresentada pelo método FBLP Modificado, é baseada no subespaço de sinal. Assim, podemos notar que o método TLS-LP já incorpora em sua solução uma restrição no subespaço da matriz de dados aumentada.

Assim, como o procedimento LS aplicado sobre um predição linear progressiva, regressiva e progressiva-regressiva deu origem aos métodos FLP, BLP e FBLP, o procedimento TLS aplicado a essas formas de predição dará origem aos respectivos métodos TLS-FLP, TLS-BLP e TLS-FBLP.

Pela superioridade apresentada pela predição na forma progressiva-regressiva, apresentaremos o método TLS-FBLP na próxima seção, permitindo assim comparações com o método FBLP Modificado.

6.3.5 Método TLS-FBLP

O método aqui denominado de TLS-FBLP consiste em obter a solução para o preditor linear progressivo-regressivo através de um procedimento de minimização inspirado no TLS, inserindo novas considerações com relação aos subespaços [Lemos97] e [Dowling96].

O método TLS-FBLP propõe uma modificação na solução de norma mínima apresentada pelo procedimento TLS clássico. Nessa modificação é realizada uma restrição de subespaço adicional sobre a matriz de dados aumentada.

Inicialmente consideremos a decomposição da matriz de dados aumentada da forma

$$[\mathbf{d}:\mathbf{D}] = \mathbf{D}' = [\mathbf{D}_1:\mathbf{D}_2] = [\mathbf{U}_{D1} \mid \mathbf{U}_{D2}] \begin{bmatrix} [\boldsymbol{\Sigma}_{D1}]_{M \times M} & \mathbf{0} \\ \mathbf{0} & [\boldsymbol{\Sigma}_{D2}]_{(L-M) \times (L-M)} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{D1}^H \\ \mathbf{V}_{D2}^H \end{bmatrix}, \quad (6.43)$$

na qual $\mathbf{D}_1 = \mathbf{U}_{D1} (\boldsymbol{\Sigma}_{D1}) \mathbf{V}_{D1}^H$ representa a partição de \mathbf{D}' associada aos M maiores valores singulares presentes em $(\boldsymbol{\Sigma}_{D1})$ e $\mathbf{D}_2 = \mathbf{U}_{D2} (\boldsymbol{\Sigma}_{D2}) \mathbf{V}_{D2}^H$ representa sua porção complementar associada aos $L-M$ menores valores singulares presentes em $(\boldsymbol{\Sigma}_{D2})$. Podemos, assim, associar essas duas partições aos subespaço de sinal e de ruído da matriz \mathbf{D}' , respectivamente.

Em uma situação prática, na qual o sinal é ruidoso, teremos os valores singulares em $\boldsymbol{\Sigma}_{D2}$ distintos. Nesse caso, a solução TLS clássica será dada pelo vetor singular associado ao menor valor singular de \mathbf{D}' . Na ausência de ruído, os valores singulares presentes em $\boldsymbol{\Sigma}_{D2}$ serão nulos.

A modificação sugerida pelo método TLS-FBLP sobre a solução TLS clássica, no caso de sinal com ruído, se inspira no comportamento teórico desse caso em que os $L - M$ menores valores singulares em $\boldsymbol{\Sigma}_{D2}$ são repetidos e iguais ao desvio padrão do ruído branco presente no sinal. Esse resultado nos permite desconsiderar as discrepâncias entre a magnitude desses valores singulares observadas em uma situação prática.

Portanto, a solução sugerida para esse método será equivalente à solução TLS clássica para o caso em que os $L - M$ menores valores singulares de \mathbf{D}' são repetidos, mesmo que na prática possuam diferenças em suas magnitudes. Considerando (6.43) tal solução é expressa por

$$\mathbf{b} = -\frac{\mathbf{y}}{\alpha} = -\frac{\mathbf{V}'_{D_2} \mathbf{p}^*}{\mathbf{p}^H \mathbf{p}}, \quad (6.44)$$

na qual utilizamos a seguinte partição $\mathbf{V}_{D_2} = \begin{bmatrix} \mathbf{p}^T \\ \mathbf{V}'_{D_2} \end{bmatrix}$.

Além das diferenças na concepção da solução obtida pelos procedimentos LS e TLS, os mesmos também diferem no ponto de vista computacional. O procedimento LS utilizado no método FBLP modificado requer a obtenção de M vetores singulares da matriz de dados \mathbf{D}_{FBLP} , com dimensões $2(N - L) \times L$, onde M determina o número de parâmetros desconhecidos do sinal. Já o método TLS-FBLP opera sobre a matriz de dados aumentada $[\mathbf{d}; \mathbf{D}_{FBLP}]$, de dimensões $2(N - L) \times (L + 1)$, e requer a obtenção dos $L - M$ vetores singulares dessa matriz associados aos seus menores valores singulares. Portanto, se M (dimensão do subespaço de sinal) for menor que $L - M$ (dimensão do subespaço de ruído), o procedimento LS será mais eficiente computacionalmente.

6.3.6 Método WTLS-LP

O método WTLS-LP foi desenvolvido para estimar freqüências através do procedimento TLS de minimização aplicado sobre um preditor, de forma distinta dos métodos TLS-LP [Makhoul75] com uma solução através do procedimento TLS [Rahman87]. O termo “W” em sua denominação denota “Whitened”, do inglês “Branqueado”, ou “Weighted”, do inglês “Ponderado”, e consiste em uma modificação do procedimento de otimização TLS clássico. Tal modificação procura aproveitar a característica do ruído branco presente no sinal, impondo-se uma restrição adequada ao problema original, conforme tratado no Capítulo 5, a qual leva a melhores resultados em termos de desempenho com relação ao método TLS-LP [Hua90].

Reportando-se à Seção 5.3 do Capítulo 5 e aplicando o procedimento WTLS de minimização sobre os dados da predição linear, através da substituição da matriz \mathbf{C} em (5.24) pela matriz de dados aumentada \mathbf{D}' , o método WTLS-LP procura uma solução que minimiza a função custo

$$J_{WTLS-LP}(\mathbf{b}) = \begin{pmatrix} -1 \\ \mathbf{b} \end{pmatrix}^H (\mathbf{D}')^H (\mathbf{B}^H \mathbf{B})^{-1} \mathbf{D}' \begin{pmatrix} -1 \\ \mathbf{b} \end{pmatrix}, \quad (6.45)$$

onde \mathbf{B}^H é dado em (3.29).

Considerando a seguinte decomposição $(\mathbf{B}^H \mathbf{B})^{-1} = \left((\mathbf{B}^H \mathbf{B})^{-1/2} \right)^H (\mathbf{B}^H \mathbf{B})^{-1/2}$, podemos reescrever (6.45) na forma compacta

$$J_{WTLS-LP}(\mathbf{b}') = \|\mathbf{D}(\mathbf{B})\mathbf{b}'\|_2^2, \quad (6.46)$$

na qual $\mathbf{b}' = \begin{pmatrix} -1 \\ \mathbf{b} \end{pmatrix}$ e $\mathbf{D}(\mathbf{B}) = \left[(\mathbf{B}^H \mathbf{B})^{-1/2} \mathbf{D}' \right]$. Desde que a matriz $\mathbf{D}(\mathbf{B})$ seja fixa, a minimização de (6.46) poderá ser obtida através do procedimento TLS. Podemos notar que se

fizermos $\mathbf{B}^H \mathbf{B} = \mathbf{I}$ em (6.46) teremos a mesma função custo proposta para os métodos TLS-LP. Dessa forma, por apresentar um problema de minimização de quarta ordem, o método WTLS-LP pode ser visto como uma generalização dos métodos baseados em predição linear.

O algoritmo iterativo que implementa a solução para o método WTLS-LP pelo procedimento TLS, considerando $\mathbf{D}(\mathbf{B})$ fixa a cada iteração, compreende os seguintes passos:

I) Inicialização: fazer $i = 0$ e $\mathbf{b}'_{(0)} = \begin{pmatrix} -1 \\ \mathbf{b}_{(0)} \end{pmatrix}$, com $\mathbf{b}_{(0)} = [0, \dots, 0]^T$, onde o subscrito (i) indica o índice da iteração. Outros valores para $\mathbf{b}_{(0)}$, mais próximos de seu valor ótimo, provenientes de uma estimativa inicial grosseira, também podem ser utilizados. Quanto melhor a inicialização, mais rápida será a convergência. Para o caso sem ruído, qualquer valor inicial para \mathbf{b}' deverá propiciar uma convergência após a primeira iteração.

Uma inicialização para a matriz $\mathbf{B}_{(0)}^H \mathbf{B}_{(0)}$ é, então, obtida a partir dos valores de $\mathbf{b}'_{(0)}$.

II) Utilizando o procedimento TLS, obter uma nova estimativa para \mathbf{b}' , com $b_0 = -1$, dada por $\mathbf{b}'_{(i+1)}$, que minimiza (6.46) para a matriz $\mathbf{D}(\mathbf{B})_{(i)}$ formada pelos valores mais recentes de \mathbf{b}' , $\mathbf{b}'_{(i)}$. A solução para $\mathbf{b}'_{(i+1)}$ será, portanto, o vetor de norma-mínima dado pelo vetor singular associado ao menor valor singular da decomposição da matriz $\mathbf{D}(\mathbf{B})_{(i)}$.

III) Se $\|\mathbf{b}'_{(i+1)} - \mathbf{b}'_{(i)}\| < \varepsilon$, seguir para o próximo passo. Caso contrário, fazer $i = i + 1$ e voltar para o passo anterior, utilizando os valores para \mathbf{b}' e para a matriz $\mathbf{D}(\mathbf{B})$.

IV) Obter as L raízes do polinômio $b(z)$, formado a partir de \mathbf{b} e com $b_0 = -1$. As estimativas dos parâmetros ϕ_m , $m = 1, 2, \dots, M$, do sinal são então tomadas pela posição angular dos M zeros mais próximos da CRU, os quais estão associados aos parâmetros do sinal. Através de simulações constata-se que os $L - M$ zeros restantes estão localizados fora da CRU, facilitando assim a distinção entre esses e os M zeros de sinal [Hua90].

A partir de (6.45) é possível contrastar a função custo do método WTLS-LP com a do método IQML em (6.6). Para o caso de um único experimento $K = 1$ e $L = M$, teremos $\mathbf{D}' = \mathbf{Y}$ e nesse caso as funções diferem apenas na restrição imposta sobre \mathbf{b}' , com $b_0 = -1$ no método WTLS-LP. Como foi enfatizado anteriormente, tal restrição não é utilizada no método IQML por ser incompatível com a restrição de simetria complexo conjugada adotada para obtenção da solução ótima nesse método. No método WTLS-LP a condição $b_0 = -1$ é oriunda da estrutura de um suposto filtro de erro de predição associado ao sinal, o qual é a base de sua concepção. Embora o método WTLS-LP não adicione restrições associadas às características do sinal, a condição de otimização é garantida através do Teorema 1 da predição linear apresentado no Capítulo 3, Seção 3.1. Portanto, apesar de possuírem formulações semelhantes, o conjunto de soluções apresentado pelos dois métodos será totalmente distinto, embora sejam ótimos sob cada critério. A formulação semelhante permite ainda afirmar que os métodos WTLS-LP e IQML possuem teoricamente o mesmo potencial de resolução quando aplicados na estimação de frequências, com uma vantagem do método WTLS-LP sobre o IQML, por permitir utilizar $L > M$ em suas formulações.

Para permitir uma comparação com os métodos baseados na matriz de correlação, podemos considerar $\hat{\mathbf{R}}_w = [\mathbf{D}(\mathbf{B})]^H \mathbf{D}(\mathbf{B})$ como uma matriz de correlação associada aos dados branqueados da matriz \mathbf{D}' original, e reescrever (6.45) como

$$J_{WTLS-LP}(\mathbf{b}) = \begin{pmatrix} -1 \\ \mathbf{b} \end{pmatrix}^H \hat{\mathbf{R}}_w \begin{pmatrix} -1 \\ \mathbf{b} \end{pmatrix}. \quad (6.47)$$

A partir de (6.47) podemos estender a aplicação do método WTLS-LP quando dispomos de mais de um experimento ($K > 1$), tomando a estimativa $\hat{\mathbf{R}}_w$ sobre todos os experimentos.

Concluimos, ainda, que o algoritmo apresentado para o método WTLS-LP nesta seção obtém a solução de um problema de quarta ordem a partir da solução iterativa de um problema de segunda ordem, com solução através do procedimento TLS de minimização. Por esse motivo, tal algoritmo também pode ser utilizado para obter a solução do método IQML ou do MODE.

Dentre as diferenças observadas entre esses algoritmos podemos citar a forma de inicialização, que ao invés de utilizar $\mathbf{B}^H \mathbf{B} = \mathbf{I}$ como nos algoritmos do IQML e MODE, o algoritmo WTLS-LP utilizou uma inicialização particular para \mathbf{b} , e a partir dessa formou a matriz $\mathbf{B}^H \mathbf{B}$. Uma outra diferença, consiste na utilização do procedimento TLS em lugar do LS sugerido para os algoritmos IQML e MODE.

Resultados de simulações mostram que o método WTLS-LP possui maior precisão, tanto em termos de variância como em polarização, quando comparado com o método TLS-LP para todas as ordens L , sobretudo para ordens mais baixas. Quando comparado com o algoritmo IQML, o método WTLS-LP não apresentou bons resultados para nenhum valor de L , sendo equivalente para $L = M$ [Hua90] e [Lemos97].

6.3.7 Métodos CTLS-LP e STLS-LP para DOA

Assim como o procedimento de minimização WTLS deu origem ao método de estimação WTLS-LP, considerando a equivalência entre o WTLS, CTLS e STLS apresentada no Capítulo 5, Seção 5.2.3, podemos também considerar a existência natural dos métodos CTLS-LP e STLS-LP. Notando que a diferença entre esses procedimentos e o WTLS, consiste na ausência de uma matriz de ponderação, que leva em consideração a contagem repetida dos elementos de ruído na matriz de dados estruturada.

Considerando, portanto, o procedimento de minimização que generaliza os procedimentos CTLS e STLS, aplicado sobre um preditor linear, teremos o seguinte problema de minimização

$$\min_{\mathbf{b}} \left\{ \begin{pmatrix} -1 \\ \mathbf{b} \end{pmatrix}^H (\mathbf{D}')^H (\mathbf{B}^H \boldsymbol{\Omega}^{-1} \mathbf{B})^{-1} \mathbf{D}' \begin{pmatrix} -1 \\ \mathbf{b} \end{pmatrix} \right\}, \quad (6.48)$$

onde $\boldsymbol{\Omega}$ é uma matriz diagonal, formada por elementos inteiros, que realiza uma ponderação adequada relacionada à estrutura da matriz de dados \mathbf{D}' . Se não considerarmos a ponderação realizada pela matriz $\boldsymbol{\Omega}$, fazendo $\boldsymbol{\Omega} = \mathbf{I}$ em (6.48), teremos o mesmo problema enfrentado pelo método WTLS-LP. Portanto, o procedimento em (6.48) é mais geral que o procedimento WTLS.

A presença da matriz de ponderações $\boldsymbol{\Omega}$ torna o procedimento de minimização em (6.48) mais adequado quando aplicado em matrizes de dados estruturadas, como as oriundas da predição linear através de um filtro progressivo, regressivo ou progressivo-regressivo.

A solução para o problema em (6.48) pode ser, então, obtida pelos algoritmos CTLS ou STLS apresentados no Capítulo 5.

6.4 Métodos Baseados em Subespaços

Os métodos que descreveremos a seguir fazem uso dos subespaços definidos pela decomposição da matriz de correlação do sinal observado, bem como das propriedades associadas aos mesmos. A decomposição, considerada aqui, é a mesma apresentada no Capítulo 2 através de (2.23) e (2.25), as quais são válidas quando a matriz da correlação entre as amplitudes complexas de cada fonte de sinal, dada por \mathbf{P} , possui posto cheio.

Embora alguns métodos baseados na predição linear tenham feito uso de subespaços em suas formulações, diferem desses por terem suas origens no conceito de predição linear e por fazerem uso da decomposição em subespaços da matriz de dados e não da matriz de correlação do sinal. Por exemplo, o método FBLP Modificado e o método TLS-FBLP fazem considerações aos subespaços associados aos dados do problema, e tal informação é utilizada apenas para obter um aprimoramento no processo de estimação através de um preditor linear. Por esse motivo optamos por sua classificação junto aos métodos baseados na predição linear.

Os métodos baseados no ML, que fazem uso dos subespaços associados à matriz de correlação do sinal, como é o caso do método MODE, tem como origem o critério ML e por esse motivo optamos por sua classificação junto aos métodos baseados no ML.

Os métodos apresentados nesta seção que permitem uma interpretação baseada no conceito de predição linear ou no critério ML serão destacados.

6.4.1 Método de Pisarenko

O método de Pisarenko utiliza as propriedades dos subespaços associados à matriz de correlação para estimar os parâmetros do sinal observado [Pisarenko73]. O método considera a dimensão do arranjo de sensores igual ao número de fontes de sinais mais uma unidade ($N = M + 1$), caracterizando assim o subespaço de ruído através do menor autovalor e correspondente autovetor da matriz de correlação.

Considerando um sinal composto pela soma de M exponenciais complexas não-amortecidas, demonstra-se que para o caso desse sinal sem ruído, a seguinte equação é satisfeita [Kay93]

$$\mathbf{R}\mathbf{b}' = \mathbf{0}, \quad (6.49)$$

na qual \mathbf{R} é a matriz de correlação estatística do sinal sem ruído e \mathbf{b}' é um vetor de dimensões compatíveis, cujos elementos são os coeficientes de um polinômio $b(z)$ de ordem L que possui raízes dentro ou sobre a CRU.

Considerando a forma decomposta da matriz \mathbf{R} com dimensões $N \times N$, apresentada em (2.23) e (2.25), teremos

$$\mathbf{R} = \mathbf{A}\mathbf{P}\mathbf{A}^H + \sigma^2\mathbf{I} = \mathbf{U}_s\mathbf{\Lambda}_s\mathbf{U}_s^H + \mathbf{U}_n\mathbf{\Lambda}_n\mathbf{U}_n^H, \quad (6.50)$$

quando a matriz \mathbf{P} possui posto cheio. Nessa condição, podemos definir o subespaço de sinal da matriz \mathbf{R} através de seus autovetores associados aos maiores autovalores, representados pela porção $\mathbf{U}_s\mathbf{\Lambda}_s\mathbf{U}_s^H$, e seu subespaço de ruído através dos autovetores associados aos menores autovalores, representados pela porção $\mathbf{U}_n\mathbf{\Lambda}_n\mathbf{U}_n^H$.

No caso de um sinal composto por M fontes distintas de sinal, se tomarmos a ordem do polinômio $b(z)$ igual ao número de amostras disponíveis menos a unidade ($L = N - 1$), teremos a partir de (6.49) e utilizando (6.50)

$$\mathbf{R}\mathbf{b}' = \sigma^2\mathbf{b}' \quad (6.51)$$

e

$$\mathbf{A}^H\mathbf{b}' = \mathbf{0}. \quad (6.52)$$

A partir de (6.51) podemos notar que para uma situação sem ruído teremos a igualdade em (6.49) satisfeita. Além disso, tal equação será satisfeita se, e somente se, \mathbf{b}' for um autovetor do subespaço de ruído de \mathbf{R} .

Considerando a equação em (6.52), teremos as raízes do polinômio $b(z)$, associado ao vetor \mathbf{b}' , sobre a CRU, nas posições angulares indicadas pelas frequências do sinal.

O método de Pisarenko consiste na busca do vetor \mathbf{b}' no subespaço nulo de \mathbf{R} que melhor satisfaça (6.49). Considerando as dimensões envolvidas em (6.51), teremos apenas um único autovetor no espaço nulo de \mathbf{R} , o qual está associado ao menor autovalor de \mathbf{R} .

Nesse método, com $L = M$, o critério a ser minimizado é, portanto, dado por

$$\min_{\mathbf{b}'} \frac{(\mathbf{b}')^H \hat{\mathbf{R}} (\mathbf{b}')}{\|\mathbf{b}'\|_2^2}, \quad (6.53)$$

na qual a matriz de correlação estimada é dada por $\hat{\mathbf{R}} = \sum_{k=1}^K \mathbf{y}(k)\mathbf{y}^H(k)$, e possui dimensões $(M+1) \times (M+1)$.

Podemos observar que no caso de não haver restrições para a escolha de \mathbf{b}' em (6.53), a solução será dada pelo autovetor de $\hat{\mathbf{R}}$ correspondente ao seu menor autovalor, o qual corresponde ao autovetor do subespaço de ruído. Assim, a partir da relação em (6.52), podemos obter estimativas para os parâmetros do sinal através das raízes do polinômio $b(z)$ formado pelos elementos do vetor \mathbf{b} que satisfaz (6.53).

O método de Pisarenko é ótimo em uma situação assintótica com relação ao número de experimentos ($K \rightarrow \infty$) e, nesse caso, coincide com o resultado obtido através do critério ML [Bresler86].

Considerando a disponibilidade de um único experimento ($K = 1$), poderemos relacionar o método de Pisarenko com os métodos baseados na predição linear. Considerando a relação $\hat{\mathbf{R}} = (\mathbf{D}')^H \mathbf{D}'$ e substituindo em (6.53), teremos

$$\min_{\mathbf{b}} \frac{\begin{bmatrix} b_0 \\ \mathbf{b} \end{bmatrix}^H (\mathbf{D}')^H \mathbf{D}' \begin{bmatrix} b_0 \\ \mathbf{b} \end{bmatrix}}{\left\| \begin{bmatrix} b_0 \\ \mathbf{b} \end{bmatrix} \right\|_2^2} = \min_{\mathbf{b}'} \frac{\|\mathbf{D}'\mathbf{b}'\|_2^2}{\|\mathbf{b}'\|_2^2}. \quad (6.54)$$

Através de (6.54), com $b_0 = -1$, podemos interpretar o método de Pisarenko como um caso particular do critério adotado para os métodos baseados na predição linear, quando a ordem do filtro é igual ao número de amostras disponíveis menos a unidade ($L = N - 1$) e, ainda, $L = M$. Nessa comparação constatamos que a interpretação de predição apresenta a desvantagem de restringir a escolha do vetor \mathbf{b}' impondo $b_0 = -1$. Por outro lado, sob o ponto de vista da predição, temos como vantagem a possibilidade de utilizar $L > M$, desde que $N > L$. Podemos notar ainda que o problema em (6.54) corresponde a um problema TLS equivalente, como em (4.33), cuja solução é dada pelo vetor singular da matriz \mathbf{D}' , associado ao seu menor valor singular. Neste caso $\|\mathbf{D}'\mathbf{b}'\|_2^2$ pode ser interpretado como a energia do erro de predição obtida

através da aplicação do filtro preditor com ganhos dados pelos elementos de \mathbf{b}' sobre as amostras de sinal dadas por \mathbf{D}' .

O método FBLP, assim como sua versão com restrição de subespaço (FBLP Modificado), descritos anteriormente, também são denominados, respectivamente, de Pisarenko Melhorado (“Improved” Pisarenko (IP)) e Pisarenko Modificado (“Modified” Pisarenko (MP)). Tal denominação se justifica por serem considerados variações do método de Pisarenko interpretado através da predição linear [Kay93]. Nesses métodos, a matriz de correlação utilizada no método Pisarenko é substituída pela matriz do auto-produto da matriz de dados, como em (6.54). No caso do método Pisarenko Modificado são adicionadas considerações com relação aos subespaços associados à matriz de dados. Tais evoluções são, portanto, formas alternativas de se incorporar informação adicional ao problema, seja explicitando a matriz de dados ou sua forma decomposta em subespaços.

Para permitir a comparação com os métodos baseado no critério ML, consideraremos a mesma matriz de reparametrização \mathbf{B}^H , dada em (3.29), com $N = M + 1$ e $L = M$. Nesse caso, teremos $\mathbf{D}' = \mathbf{Y}$, $\mathbf{B}^H = [\mathbf{b} \quad b_0]$ e poderemos transformar o problema em (6.53) no seguinte problema equivalente

$$\min_{\mathbf{b}'} Tr \left\{ \mathbf{B}(\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H \hat{\mathbf{R}} \right\}, \quad (6.55)$$

que por sua vez é equivalente ao problema encontrado para os métodos baseados no ML, considerando a mesma reparametrização. Portanto, o problema enfrentado pelo método de Pisarenko coincide com o problema proposto para o método IQML no caso particular em que $N = M + 1$. Além dessa limitação quanto ao número de amostras por experimentos (dimensão do arranjo de sensores), o método de Pisarenko não sugere nenhuma restrição associada às características particulares do sinal aplicada sobre os elementos de \mathbf{b}' que satisfaça a condição ótima, como nos métodos IQML e MODE. Por esse motivo, observa-se que para um número finito de amostras, o polinômio $b(z)$ obtido pelo método de Pisarenko não possuirá seus zeros nas posições ótimas, mesmo para um sinal sem ruído. Para contornar essa limitação, o método de Pisarenko clássico deve ser, então, modificado implementando-se restrições apropriadas. Para aplicação na qual o modelo do sinal é composto por exponenciais complexas não-amortecidas, a mesma restrição de simetria complexa-conjugada adotada para os métodos IQML e MODE pode ser estabelecida através de uma restrição quadrática e incorporada através de *multiplicadores de Lagrange*, resultando em um problema geral de autovalores, ou ainda de forma implícita como implementada nos algoritmos IQML e MODE.

A interpretação da matriz de correlação $\hat{\mathbf{R}}$, ora como o auto-produto da matriz de dados, ora em forma decomposta em subespaços, permite diferenciar os métodos FLP, BLP e FBLP, daqueles com restrição de subespaços, como é o caso dos métodos FBLP Modificado, TLS-LP e TLS-FBLP. Essa diferença é essencialmente a mesma que diferencia os métodos IQML e MODE.

Devido à possibilidade de utilização de uma mesma matriz de reparametrização \mathbf{B}^H , a solução para o método de Pisarenko pode ser interpretada sob o ponto de vista da predição linear ou através do critério ML, estabelecendo as restrições apropriadas sobre o vetor solução \mathbf{b}' em (6.53).

6.4.2 Método MUSIC

O método MUSIC faz uso explícito dos subespaços definidos a partir da decomposição da matriz de correlação estimada. Nesse método não existe a limitação sobre a dimensão do arranjo de sensores, como no método de Pisarenko [Cadzow90] e [Reddi79].

O método MUSIC original tem por objetivo formar uma função espectral definida como

$$P_{MUSIC}(\phi) = \frac{1}{\|\mathbf{U}_n^H \mathbf{a}(\phi)\|_2^2}, \quad (6.56)$$

na qual $\mathbf{a}(\phi)$ é o vetor de direção, dado em (2.13), e \mathbf{U}_n denota a matriz cujas colunas são os autovetores da matriz de correlação associados aos seus $N - M$ menores autovalores, definida no Capítulo 2. Embora (6.56) não seja um espectro realmente, indica o grau de ortogonalidade entre os subespaços de ruído, formado por \mathbf{U}_n , e sinal na direção ϕ . Dessa forma, a função em (6.56) apresenta picos nas proximidades dos parâmetros ϕ_m do sinal observado, para $m = 1, 2, \dots, M$, possibilitando a obtenção de estimativas para esses parâmetros. Tal forma de obtenção dos parâmetros DOA dá origem ao método MUSIC Espectral, ou método MUSIC baseado no espectro.

Uma forma alternativa e mais vantajosa para a obtenção de estimativas dos parâmetros ϕ_m do sinal consiste na busca das raízes de um polinômio associado ao denominador da função $P_{MUSIC}(\phi_m)$ em (6.56), de tal forma que as raízes desse polinômio estejam relacionadas com os picos da função espectral. Tal polinômio pode, por sua vez, ser interpretado como o polinômio resultante de um filtro preditor de ordem $L = N$, otimizado para o sinal de entrada. Essa forma paramétrica de obtenção dos parâmetros DOA dá origem ao método root-MUSIC, ou método MUSIC baseado nas raízes.

Mais especificamente, o método root-MUSIC tem como objetivo encontrar as raízes da função

$$b(z) = z^{N-1} \mathbf{b}^H(z) \mathbf{U}_n \mathbf{U}_n^H \mathbf{b}(z), \quad (6.57)$$

na qual $\mathbf{b}(z) = [1, z, \dots, z^{N-1}]$. O polinômio resultante é dado, portanto, pela combinação linear dos autovetores associados ao subespaço de ruído da matriz de correlação do sinal.

Uma outra variação para o método MUSIC resulta da utilização da forma decomposta da matriz de correlação $\mathbf{R} = \mathbf{A}(\phi) \mathbf{P} \mathbf{A}(\phi)^H + \sigma^2 \mathbf{I} = \mathbf{U}_s \mathbf{\Lambda}_s \mathbf{U}_s^H + \mathbf{U}_n \mathbf{\Lambda}_n \mathbf{U}_n^H$ e do fato de termos os autovetores do subespaço de ruído, dados pelas colunas de \mathbf{U}_n , ortogonais às linhas da matriz de $\mathbf{A}^H(\phi)$. Através dessas considerações, teremos

$$\mathbf{A}^H(\phi) \mathbf{U}_n = \mathbf{U}_n^H \mathbf{A}(\phi) = 0; \text{ quando } \phi = \phi_o, \quad (6.58)$$

na qual ϕ_o denota os valores ótimos dos parâmetros. Enfatizamos que a igualdade em (6.58) é válida apenas quando a matriz $\mathbf{P} = \mathcal{E}\{\mathbf{s}(\mathbf{k})\mathbf{s}(\mathbf{k})^H\}$ possui posto cheio, implicando em um cenário no qual os sinais $s_m(k)$ não são coerentes. Neste caso, os valores ótimos ϕ_1, \dots, ϕ_M são as únicas possíveis soluções para a relação (6.58), das quais extraímos os parâmetros DOA.

Portanto, dada uma estimativa $\hat{\mathbf{U}}_n$ de \mathbf{U}_n , devemos encontrar o conjunto de parâmetros estimados, $\hat{\phi}_m$, $m = 1, 2, \dots, M$, que minimiza o seguinte critério quadrático:

$$\left\| \mathbf{A}^H(\phi) \hat{\mathbf{U}}_n \right\|_F^2 = \sum_{m=1}^M \left\| \hat{\mathbf{U}}_n^H \mathbf{a}(\phi_m) \right\|_2^2. \quad (6.59)$$

Embora o método MUSIC clássico permita a obtenção dos parâmetros $\hat{\phi}_m$, para $m = 1, 2, \dots, M$, de forma independente, o critério a ser minimizado em (6.59) considera a obtenção simultânea dos parâmetros ótimos e por esse motivo dá origem ao método denominado de MUSIC multidimensional (MD-MUSIC).

Utilizando a definição da norma de *Frobenius*, o estimador MUSIC baseado no critério em (6.59) reduz-se à minimização da seguinte função custo:

$$\begin{aligned} J_{MUSIC}(\phi) &= Tr \left\{ \mathbf{A}^H(\phi) \left(\hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H \right) \mathbf{A}(\phi) \right\}, \text{ com } \phi = \mathcal{R}^M, \\ &= Tr \left\{ \mathbf{A}(\phi) \mathbf{A}^H(\phi) \left(\hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H \right) \right\} \end{aligned} \quad (6.60)$$

A função custo derivada em (6.60) será de grande utilidade na comparação com os métodos de estimação baseados no critério ML ou na predição linear, que utilizam os subespaços da matriz de correlação em suas formulações. Uma característica marcante do método MUSIC que pode ser observada em (6.60) consiste na utilização do subespaço de ruído para a obtenção de sua solução.

Uma extensão do MUSIC, denominada de MUSIC Ponderado, faz uso de uma matriz de ponderação, \mathbf{W}_{MUSIC} de dimensões $(N - M) \times (N - M)$, que pondera a influência de cada um dos autovetores de forma distinta, levando à composição do critério

$$J_{MUSIC_W}(\phi) = Tr \left\{ \mathbf{A}^H(\phi) \left(\hat{\mathbf{U}}_n [\mathbf{W}_{MUSIC}] \hat{\mathbf{U}}_n^H \right) \mathbf{A}(\phi) \right\}. \quad (6.61)$$

Podemos notar que desprezar a ponderação sugerida pela matriz \mathbf{W}_{MUSIC} em (6.61) é equivalente ao caso $\mathbf{W}_{MUSIC} = \mathbf{I}$, para qual temos a mesma função custo derivada em (6.60). Por esse motivo, a função custo em (6.61) generaliza o método MUSIC.

Embora o método MUSIC original apresente uma menor variância assintótica que a sua versão ponderada, em situações em que o número de amostras é pequeno, a relação sinal-ruído é baixa e os sinais são correlacionados, podemos obter uma melhor capacidade de resolução escolhendo adequadamente os elementos da matriz \mathbf{W}_{MUSIC} [Krim96].

Reescrevendo o critério $J_{MUSIC_W}(\phi)$ em termos de uma outra matriz de ponderações, \mathbf{V}_{MUSIC} de dimensões $M \times M$, podemos interpretar o método MUSIC como o resultado do processo de minimização da função

$$J_{MUSIC_V}(\phi) = Tr \left\{ \mathbf{A}^H(\phi) \hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H \mathbf{A}(\phi) [\mathbf{V}_{MUSIC}] \right\}, \quad (6.62)$$

na qual \mathbf{V}_{MUSIC} é uma matriz positiva semi-definida de ponderação, cuja escolha $\mathbf{V}_{MUSIC} = \mathbf{I}$ também leva à função custo em (6.60).

Pode ser demonstrado que a escolha para \mathbf{V}_{MUSIC} em (6.62) da forma [Krim96]

$$\mathbf{V}_{MUSIC} = \mathbf{V}_o = \mathbf{A}^\dagger(\phi_o) \hat{\mathbf{U}}_s \mathbf{W} \hat{\mathbf{U}}_s^H [\mathbf{A}^\dagger(\phi_o)]^H, \quad (6.63)$$

onde \mathbf{V}_o e ϕ_o denota seus valores ótimos, leva à obtenção de estimativas assintoticamente iguais às obtidas pela minimização da função custo em (6.61).

Poderemos, ainda, considerar a presença de duas matrizes de ponderações \mathbf{W}_{MUSIC} e \mathbf{V}_{MUSIC} na função custo em (6.60), dando origem à seguinte função custo

$$J_{MUSIC_VW}(\phi) = Tr \left\{ \mathbf{A}(\phi) [\mathbf{V}_{MUSIC}] \mathbf{A}^H(\phi) \left(\hat{\mathbf{U}}_n [\mathbf{W}_{MUSIC}] \hat{\mathbf{U}}_n^H \right) \right\}, \quad (6.64)$$

a qual reduz-se à função custo em (6.60), quando \mathbf{V}_{MUSIC} e \mathbf{W}_{MUSIC} são feitas iguais à matriz identidade de dimensões apropriadas. Embora potencialmente a função custo em (6.64) possibilite melhores resultados quando aplicada em sinais correlacionados, apresentada a dificuldade adicional da escolha das matrizes de ponderações mais adequadas.

Para permitir comparação com os métodos baseados no critério ML, procederemos com manipulações algébricas e novas interpretações sobre a expressão para a função custo derivada em (6.64).

Podemos observar que fazendo $\mathbf{V}_{MUSIC} = [\mathbf{A}^H(\phi)\mathbf{A}(\phi)]^{-1}$ e $\mathbf{W}_{MUSIC} = \hat{\mathbf{\Lambda}}_n$ em (6.64), teremos

$$J_{MUSIC_VW}(\phi) = Tr \left\{ \mathbf{\Pi}_{\mathbf{A}}(\phi) \hat{\mathbf{R}}_n \right\}, \quad (6.65)$$

na qual reconhecemos $\mathbf{\Pi}_{\mathbf{A}} \triangleq \mathbf{A}(\phi) [\mathbf{A}^H(\phi)\mathbf{A}(\phi)]^{-1} \mathbf{A}^H(\phi) = \mathbf{U}_s \mathbf{U}_s^H$ e $\hat{\mathbf{R}}_n = \hat{\mathbf{U}}_n \hat{\mathbf{\Lambda}}_n \hat{\mathbf{U}}_n^H$ é a matriz associada a porção ruidosa da matriz de correlação estimada na forma decomposta em subespaços.

A partir da função custo derivada para o critério ML, dada por $J_{ML} = Tr \left\{ \mathbf{\Pi}_{\mathbf{A}}^\perp(\phi) \hat{\mathbf{R}} \right\}$, considerando que $\hat{\mathbf{R}} = \hat{\mathbf{R}}_s + \hat{\mathbf{R}}_n$, com $\hat{\mathbf{R}}_s = \hat{\mathbf{U}}_s \hat{\mathbf{\Lambda}}_s \hat{\mathbf{U}}_s^H$, poderemos reescrever a função custo, equivalente, do critério ML baseado em subespaço como

$$J_{ML}(\phi) = Tr \left\{ \mathbf{\Pi}_{\mathbf{A}}^\perp(\phi) \hat{\mathbf{R}}_s \right\}, \quad (6.66)$$

na qual $\mathbf{\Pi}_{\mathbf{A}}^\perp(\phi) = \mathbf{U}_n \mathbf{U}_n^H$.

Podemos, então, concluir, através da comparação entre (6.65) e (6.66), que ambas buscam os parâmetros ϕ_m , $m = 1, 2, \dots, M$, que melhor satisfazem a ortogonalidade entre os subespaços de sinal, representado pelo projetor $\mathbf{\Pi}_{\mathbf{A}}(\phi)$ em (6.65) e por $\hat{\mathbf{R}}_s$ no critério ML, e o subespaço de ruído, representado por $\hat{\mathbf{R}}_n$ em (6.65) e pelo projetor ortogonal $\mathbf{\Pi}_{\mathbf{A}}^\perp(\phi)$ no critério ML. A diferença básica entre as formas de obtenção das estimativas é que a função custo associada ao método MUSIC é baseada no subespaço de ruído da matriz de correlação estimada, enquanto que o critério ML baseia-se no subespaço de sinal.

Para estender a comparação entre o método MUSIC e os métodos baseados no critério ML, como IQML e MODE, devemos considerar a reparametrização $\mathbf{\Pi}_{\mathbf{A}} = \mathbf{\Pi}_{\mathbf{B}}^\perp$. Para esse caso, valem os mesmos algoritmos e considerações colocados anteriormente para os métodos IQML e MODE, uma vez que a função custo em (6.65) também é de quarta ordem.

Para estabelecer comparações com os métodos baseados na predição linear deveremos considerar o método root-MUSIC em (6.57), no qual L corresponde à ordem do preditor utilizado.

Essa comparação também pode ser realizada a partir da função custo em (6.65), considerando a reparametrização $\mathbf{\Pi}_A = \mathbf{\Pi}_B^\perp$, para $L = M$.

Uma das limitações desse método é a incapacidade de resolver sinais próximos entre si, sobretudo em baixas relações sinal-ruído [Stoica89] e [Stoica90d]. No caso de sinais coerentes, a propriedade de posto cheio da matriz \mathbf{P} e, conseqüentemente, de \mathbf{APA}^H é violada e o método falha na obtenção de estimativas consistentes. Uma outra limitação vem do fato da matriz de ponderações ótima, \mathbf{V}_o em (6.63), depender de ϕ_o , exigindo um procedimento em duas etapas.

Apesar dessas restrições, o desempenho do método MUSIC cresce assintoticamente com o aumento do número de amostras disponíveis para alta relação sinal-ruído [Stoica89] e [Stoica90d]. Como vantagem, temos a baixa complexidade e baixo esforço computacional de sua implementação se considerarmos sua formulação original em (6.56).

6.4.3 Método da Norma-mínima

A método da Norma-mínima faz uso das relações entre os subespaços da mesma forma que no método MUSIC. Por esse motivo, poderemos interpretar o método da Norma-mínima como um caso particular do método MUSIC. Mais precisamente, quando consideramos a versão do método MUSIC baseado nas raízes (root-MUSIC) e impomos que o vetor associado à sua solução apresente a menor norma *Euclidiana*. Por sua vez, o resultado desse procedimento é equivalente à escolha particular para a matriz de ponderações existente no método MUSIC Ponderado.

O princípio do método da Norma-mínima, aplicado em arranjos uniformes lineares, é procurar estimativas dos parâmetros DOA através da localização dos mínimos da seguinte função espectral [Krim96] e [Kaveh86]

$$P_{NM}(\phi) = \frac{1}{\|\mathbf{U}_n^H \mathbf{b}(z)\|_2^2}, \quad (6.67)$$

na qual $\mathbf{b}(z) = [1, z, \dots, z^{N-1}]^T$, com $z = e^{j\phi}$. Podemos observar que para $z = e^{j\phi_m}$, teremos $\mathbf{b}(z) = \mathbf{a}(\phi_m)$ e, portanto, (6.67) é equivalente a (6.56), apresentando picos espectrais em $\phi = \phi_m$, para $m = 1, 2, \dots, M$.

A forma paramétrica de implementação da solução para o método da Norma-mínima consiste na obtenção de um vetor \mathbf{b}' N -dimensional, cujos elementos são os coeficientes do polinômio formado pelo denominador da função em (6.67), dado por

$$b(z) = (\mathbf{b}')^T \mathbf{b}(z) \quad (6.68)$$

com M de suas raízes dadas por $z = e^{j\phi_m}$. Uma vez que $N > M$, tal condição não será satisfeita por um único vetor \mathbf{b}' e neste caso o método da Norma-mínima fornecerá como solução aquele com a menor norma *Euclidiana*, dando origem à sua denominação.

De acordo com (6.67) e (6.68), o vetor \mathbf{b}' deve pertence ao subespaço de ruído e dado, portanto, por uma combinação linear das colunas da matriz \mathbf{U}_n , da forma

$$\mathbf{b}' = \mathbf{U}_n \mathbf{v}, \quad (6.69)$$

onde \mathbf{v} é um vetor de ponderações adequado, tendo em conta a menor norma *Euclidiana* para \mathbf{b}' .

Conseqüentemente, \mathbf{b}' deve ser ortogonal ao subespaço de sinal, gerado pelas colunas de \mathbf{U}_s e, portanto, satisfaz a relação

$$\mathbf{U}_s^H \mathbf{b}' = \mathbf{0}. \quad (6.70)$$

Para o caso em que dispomos apenas de uma estimativa de \mathbf{U}_n e \mathbf{U}_s , a relação em (6.70) deve ser satisfeita de forma aproximada considerando algum critério de erro.

A solução para esse método é, portanto, obtida através do seguinte problema de otimização

$$\mathbf{b}' = \arg \left\{ \min \|\mathbf{b}'\|_2^2 \right\} \text{ tal que } \hat{\mathbf{U}}_s^H \mathbf{b}' = \mathbf{0} \text{ e } b_0 = 1. \quad (6.71)$$

Incorporando as restrições sobre $\mathbf{b}' = [b_0 \ \mathbf{b}]^T$, teremos a seguinte função custo a ser minimizada, associada ao problema em (6.71)

$$J_{NM}(\mathbf{b}') = \left\| \hat{\mathbf{U}}_s^H \begin{bmatrix} 1 \\ \mathbf{b} \end{bmatrix} \right\|_2^2. \quad (6.72)$$

A solução fechada para o vetor \mathbf{b} que minimiza a função custo em (6.72) e possui a menor norma *Euclidiana* pode ser obtida através do procedimento LS de otimização e é dada por

$$\mathbf{b} = \left(\hat{\mathbf{U}}_s' \right)^\dagger \mathbf{g} = \frac{\hat{\mathbf{U}}_s' \mathbf{g}}{1 - \mathbf{g}^H \mathbf{g}}, \quad (6.73)$$

na qual

$$\hat{\mathbf{U}}_s' = \begin{bmatrix} \mathbf{g}^T \\ \hat{\mathbf{U}}_s' \end{bmatrix}.$$

Após obter o vetor solução \mathbf{b}' , o método da Norma-mínima pode obter as estimativas para o parâmetro ϕ_m do sinal, para $m = 1, 2, \dots, M$, através dos picos da função em (6.67) ou, preferencialmente, através das raízes do polinômio $b(z)$ em (6.68). A característica responsável pela capacidade de resolução desse método é que a solução da norma-mínima leva M dos N zeros de $b(z)$ sobre a circunferência de raio unitário nas posições angulares dadas pelos parâmetros ϕ_m do sinal, e ao mesmo tempo reduz a magnitude dos $N - M$ zeros restantes ao valor mínimo possível.

Dessa forma, o método root-MUSIC também pode ser interpretado como uma generalização do método de Norma-mínima para os casos em que o polinômio solução em (6.57) é escolhido a partir dos elementos de um vetor no subespaço de ruído sem a restrição norma *Euclidiana* mínima.

Para permitir comparação com o método MUSIC Ponderado podemos redefinir a função custo em (6.72), considerando a seguinte relação equivalente a (6.70), dada por

$$\mathbf{A}(\phi)^H \mathbf{b}' = \mathbf{0}, \text{ quando } \phi = \phi_o, \text{ para } m = 1, 2, \dots, M, \quad (6.74)$$

e utilizando (6.69), obtendo

$$J_{NM}(\phi) = \left\| \mathbf{A}(\phi)^H \mathbf{b}' \right\|_2^2 = \left\| \mathbf{A}(\phi)^H \hat{\mathbf{U}}_n \mathbf{v} \right\|_2^2, \quad (6.75)$$

na qual as estimativas do parâmetro do sinal são obtidas a partir do conhecimento *a priori* de \mathbf{v} .

Poderemos, ainda, reescrever a função custo em (6.75) como

$$J_{NM}(\phi) = \left[\mathbf{A}(\phi)^H \hat{\mathbf{U}}_n [\mathbf{v}\mathbf{v}^H] \hat{\mathbf{U}}_n^H \mathbf{A}(\phi) \right]. \quad (6.76)$$

Comparando (6.76) com (6.61) podemos interpretar o método da Norma-mínima como um caso particular do método MUSIC Ponderado quando fazemos $\mathbf{W}_{MUSIC} = \mathbf{v}\mathbf{v}^H$. Pode ser demonstrado que a escolha para \mathbf{v} que satisfaz o método da Norma-mínima é dada por $\mathbf{v} = \hat{\mathbf{U}}_n^H \mathbf{e}_1$, onde $\mathbf{e}_1 = [1, 0, 0, \dots, 0]^T$ [Ermolaev94].

Apesar do método da Norma-mínima ter sua origem na aplicação de estimação de parâmetros de sinal no caso espacial, através do uso subespaços e utilizando arranjos uniformes lineares bem como uma reparametrização adequada, é possível estabelecer sua relação com os métodos baseados na predição linear, conforme apresentaremos a seguir.

A relação com os métodos baseados na predição linear se faz interpretando o vetor $\mathbf{b}' = [-b_0 \quad -\mathbf{b}]^T$, com $b_0 = 1$, como o vetor dos coeficientes de um filtro preditor de ordem $L = N$.

Através de (6.71) e (6.72), podemos construir o seguinte problema equivalente:

$$\min J(\mathbf{b}) = \min \left\{ \left\| \hat{\mathbf{U}}_s^H \begin{bmatrix} -1 \\ \mathbf{b} \end{bmatrix} \right\|_2^2 \right\}, \text{ sujeito a } \min \{ \mathbf{b}^H \mathbf{b} \}. \quad (6.77)$$

Desenvolvendo o produto em (6.77) obtemos

$$\min J(\mathbf{b}) = \min \left\{ \begin{bmatrix} -1 \\ \mathbf{b} \end{bmatrix}^H \hat{\mathbf{U}}_s \hat{\mathbf{U}}_s^H \begin{bmatrix} -1 \\ \mathbf{b} \end{bmatrix} \right\}, \text{ sujeito a } \min \{ \mathbf{b}^H \mathbf{b} \}. \quad (6.78)$$

Se considerarmos a versão decomposta da matriz de autocorrelação estimada como $\hat{\mathbf{R}} = \hat{\mathbf{R}}_s + \hat{\mathbf{R}}_n$, poderemos interpretar $\tilde{\mathbf{R}}_s = \hat{\mathbf{U}}_s \hat{\mathbf{U}}_s^H$ como uma versão da matriz $\hat{\mathbf{R}}_s$ na qual $\hat{\mathbf{A}}_s = \mathbf{I}$. Tal estimativa para o subespaço de sinal é válida para os casos em que as fontes de sinais são descorrelacionadas e possuem mesma potência.

Reescrevendo (6.78), após essas considerações, teremos

$$\min J(\mathbf{b}) = \min \left\{ \begin{bmatrix} -1 \\ \mathbf{b} \end{bmatrix}^H \tilde{\mathbf{R}}_s \begin{bmatrix} -1 \\ \mathbf{b} \end{bmatrix} \right\}, \text{ sujeito a } \min \{ \mathbf{b}^H \mathbf{b} \}. \quad (6.79)$$

Através de (6.79) podemos, então, interpretar o método da Norma-mínima como um procedimento de predição linear de ordem $L = N$ na qual a matriz de correlação estimada, dada anteriormente pelo produto $\mathbf{Y}^H \mathbf{Y}$ em (6.40), é substituída por uma versão truncada, baseada apenas no subespaço de sinal, dada por $\tilde{\mathbf{R}}_s$. Deve ser enfatizado que esse procedimento de restrição de subespaço, observado na função custo em (6.70), é distinto daquele apresentado para o método FBLP Modificado, onde tal restrição é imposta após o processo de otimização.

Para a solução do problema em (6.79) podemos utilizar o procedimento LS ou TLS de minimização, apresentados no Capítulo 4, garantindo de forma implícita a condição $\min \{ \mathbf{b}^H \mathbf{b} \}$.

6.5 Métodos Baseados na “Distância” entre Subespaços (“Subspace Fitting”)

A utilização das informações associadas aos subespaços oferece um aumento significativo no desempenho dos métodos convencionais baseados no critério ML ou na predição linear. O método MUSIC, por exemplo, fornece estimativas com uma precisão equivalente ao da implementação do critério ML determinístico, desde que as fontes de sinais sejam descorrelacionadas. No entanto, tais métodos geralmente exibem uma grande polarização em suas estimativas para um número finito de amostras, levando a problemas de resolução. Tal dificuldade de resolução é, então, agravada para fontes de sinais correlacionadas [Krim96], [Stoica89], [Stoica90d] e [Cadzow90].

Os métodos apresentados aqui fazem uso dos subespaços definidos pela matriz de correlação de uma forma distinta daquela utilizada pelos métodos baseados em subespaço descritos na seção anterior. Tal forma diferenciada permite melhorar o desempenho em situações que envolvem um número finito de amostras e possíveis correlações entre as fontes de sinais, equiparando-se aos melhores métodos. Por esse motivo foram classificados separadamente.

Nesses métodos o critério a ser minimizado consiste em uma medida de “distância” entre o subespaço de interesse e sua aproximação, que é obtida a partir de uma reparametrização adequada. Tal medida é geralmente tomada como um erro de aproximação entre duas matrizes [Viberg91a] e [Viberg91b].

O primeiro método a ser apresentado, denominado de ESPRIT, surgiu antes do conceito de “distância” entre subespaços. Optamos por sua classificação como um método baseado nesse critério devido à sua formulação semelhante [Paulraj86], [Stoica91] e [Roy89].

A utilização de arranjos lineares e uniforme permite relacionar os métodos descritos aqui com os métodos baseados no critério ML, IQML e MODE, bem como com os métodos baseados na predição linear.

6.5.1 Método ESPRIT

O método ESPRIT pode ser interpretado como a utilização de dois arranjos de sensores, lineares e uniformemente espaçados, separados por uma distância igual ao espaçamento entre seus sensores. Na prática não existe a necessidade de utilizar dois arranjos fisicamente separados, pois através da geometria do arranjo os elementos podem ser selecionados conforme essa propriedade [Krim96], [Paulraj86] e [Roy89].

O método usa a estrutura da matriz de *Vandermonde* \mathbf{A} para definir as submatrizes \mathbf{A}_1 e \mathbf{A}_2 , respectivamente, através da extração da primeira e última linha de \mathbf{A} , isto é

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \text{última linha} \end{bmatrix}_{N \times M} = \begin{bmatrix} \text{primeira linha} \\ \mathbf{A}_2 \end{bmatrix}_{N \times M}. \quad (6.80)$$

Dessa forma, um arranjo linear uniforme com N elementos espaçados entre si por uma distância d pode ser visto como dois arranjos com $N - 1$ elementos. Um dos arranjos seria composto pelos primeiros $N - 1$ elementos do arranjo inicial, enquanto que o outro seria formado pelos $N - 1$ últimos elementos. Dessa forma, teremos pares de sensores compostos pelo primeiro e o segundo elemento do arranjo inicial, o segundo e o terceiro elemento e assim sucessivamente.

Se considerarmos os sinais recebidos de uma fonte na direção θ_m pelo n 'ésimo par, descritos aqui como $x_{1,n}(t)$ e $x_{2,n}(t)$, a diferença entre esses dois sinais, segundo a expressão (2.6), será dada por

$$x_{2,n}(t) = x_{1,n}(t)e^{j2\pi\Delta_o\cos\theta_m} \quad (6.81)$$

onde $\Delta_o = d/\lambda_c$, representa a distância entre os dois elementos, medida em comprimentos de onda. Conseqüentemente, considerando todos os pares e a presença de ruído, os sinais fornecido pelos dois arranjos, de acordo com (2.16), são dados por

$$\mathbf{y}_1(t) = \mathbf{A}_1\mathbf{s}(t) + \mathbf{n}_1(t) \quad (6.82)$$

e

$$\mathbf{y}_2(t) = \mathbf{A}_1\mathbf{\Phi}\mathbf{s}(t) + \mathbf{n}_2(t), \quad (6.83)$$

onde \mathbf{A}_1 é uma matriz de *Vandermonde* correspondente ao primeiro subarranjo e $\mathbf{\Phi}$ uma matriz diagonal com elementos dados por

$$\Phi_{mm} = e^{j2\pi\Delta_o\cos\theta_m}, \quad m = 1, 2, \dots, M. \quad (6.84)$$

Comparando os sinais $\mathbf{y}_1(t)$ e $\mathbf{y}_2(t)$, podemos concluir que a matriz dos vetores de direção para o segundo subarranjo é dada por

$$\mathbf{A}_2 = \mathbf{A}_1\mathbf{\Phi}. \quad (6.85)$$

Considerando \mathbf{U}_{y_1} e \mathbf{U}_{y_2} como as matrizes cujas colunas representam os M autovetores associados ao subespaço de sinal das respectivas matrizes de correlação dos sinais $\mathbf{y}_1(t)$ e $\mathbf{y}_2(t)$, e notando que esses dois conjuntos de autovetores geram o mesmo espaço de sinal M -dimensional, podemos estabelecer as seguintes transformações

$$\mathbf{U}_{y_1} = \mathbf{A}_1\mathbf{T}; \quad \mathbf{U}_{y_2} = \mathbf{A}_2\mathbf{T}, \quad (6.86)$$

onde \mathbf{T} é uma matriz de transformação não singular e com dimensões $M \times M$.

Combinando (6.86) e (6.85) obtemos a seguinte relação

$$\mathbf{U}_{y_1}\mathbf{\Psi} = \mathbf{U}_{y_2}, \quad (6.87)$$

na qual $\mathbf{\Psi}$ é uma matriz de transformação não-singular única, definida como $\mathbf{\Psi} = \mathbf{T}^{-1}\mathbf{\Phi}\mathbf{T}$. Dessa forma as matrizes $\mathbf{\Psi}$ e $\mathbf{\Phi}$ estão relacionadas por uma transformação de similaridade.

Considerando, agora, $\hat{\mathbf{U}}_{y_1}$ e $\hat{\mathbf{U}}_{y_2}$ como as matrizes associadas ao subespaço de sinal das respectivas matrizes de correlação estimadas, teremos a relação aproximada

$$\hat{\mathbf{U}}_{y_1}\mathbf{\Psi} \approx \hat{\mathbf{U}}_{y_2}. \quad (6.88)$$

O método ESPRIT consistirá, portanto, na busca dos parâmetros em $\mathbf{\Psi}$ que melhor satisfaça (6.88). Isto é, devemos minimizar a norma do erro da relação aproximada em (6.88) através do seguinte problema:

$$\min_{\hat{\Psi}} \left\| \hat{\mathbf{U}}_{y_1} \hat{\Psi} - \hat{\mathbf{U}}_{y_2} \right\|_F^2. \quad (6.89)$$

A expressão em (6.89) pode ainda ser reescrita como

$$\min_{\tilde{\mathbf{C}}} \left\| \tilde{\mathbf{C}} - \mathbf{C} \right\|_F^2, \quad (6.90)$$

na qual procuramos a melhor aproximação para o subespaço definido pela matriz $\mathbf{C} = \hat{\mathbf{U}}_{y_2}$, dada por $\tilde{\mathbf{C}}$, através de um outro subespaço definido pela matriz $\hat{\mathbf{U}}_{y_1}$, semelhante ao problema TLS equivalente, apresentado no Capítulo 4, Seção 4.4.

A solução para o problema em (6.89) pode ser obtida através dos seguintes passos:

- I) Obter, a partir dos dados disponíveis, as versões decompostas em subespaços das matrizes de correlação estimadas, $\hat{\mathbf{R}}_1$ e $\hat{\mathbf{R}}_2$, associadas, respectivamente, aos sinais $\mathbf{y}_1(t)$ e $\mathbf{y}_2(t)$.
- II) Formar as matrizes $\hat{\mathbf{U}}_{y_1}$ e $\hat{\mathbf{U}}_{y_2}$ a partir dos M autovetores associados aos seus M maiores autovalores das matrizes $\hat{\mathbf{R}}_1$ e $\hat{\mathbf{R}}_2$, respectivamente,
- III) Resolver o problema de minimização em (6.89) através do procedimento LS ou TLS, dando origem aos respectivos métodos LS-ESPRIT e TLS-ESPRIT.
- IV) Obter os parâmetros DOA a partir dos autovalores de $\hat{\Psi}$.

Demonstra-se que os métodos LS-ESPRIT e TLS-ESPRIT apresentam uma precisão assintótica idêntica. Para um pequeno número de amostras, o método TLS-ESPRIT apresenta uma superioridade, esperada, visto que nesse método a natureza ruidosa de ambas as matrizes $\hat{\mathbf{U}}_{y_1}$ e $\hat{\mathbf{U}}_{y_2}$ é levada em consideração no procedimento TLS de minimização.

Devido à flexibilidade na partição do arranjo através da matriz \mathbf{A} , outras variações para o método ESPRIT são possíveis.

6.5.2 Método “Signal Subspace Fitting - SSF”

O método SSF consiste na utilização do critério de “distância” considerando uma aproximação baseada no subespaço de sinal da matriz de correlação [Krim96] e [Viberg91a].

Mais especificamente, o método SSF consiste em resolver o problema de minimização

$$\min_{\phi, \mathbf{T}} \left\| \hat{\mathbf{U}}_s - \mathbf{A}(\phi)\mathbf{T} \right\|_F^2, \quad (6.91)$$

com $\mathbf{A}(\phi)\mathbf{T} = \mathbf{U}_s$, para $\phi \in \mathcal{R}^M$, e $\mathbf{T} = \mathbf{P}\mathbf{A}^H\mathbf{U}_s(\Lambda_s - \sigma^2\mathbf{I})^{-1}$, conforme apresentado em (3.40).

Podemos notar que o problema em (6.91) possui a mesma estrutura do problema em (6.89) e, portanto, pode ser resolvido seguindo os mesmos passos.

Podemos demonstrar que se utilizarmos a solução para o parâmetro $\hat{\mathbf{T}}$, dada por $\hat{\mathbf{T}} = \mathbf{A}^\dagger(\phi) \hat{\mathbf{U}}_s$, e substituirmos em (6.91), teremos o problema equivalente

$$\min_{\phi} Tr \left\{ (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger) \hat{\mathbf{U}}_s \hat{\mathbf{U}}_s^H \right\} = \min_{\phi} Tr \left\{ \mathbf{\Pi}_A^\perp \hat{\mathbf{U}}_s \hat{\mathbf{U}}_s^H \right\}. \quad (6.92)$$

Podemos notar através de (6.92) que se tomarmos a versão truncada da matriz de correlação estimada como $\tilde{\mathbf{R}}_s = \hat{\mathbf{U}}_s \hat{\mathbf{U}}_s^H$, teremos um problema equivalente ao encontrado para os métodos baseados no critério ML que utilizam subespaços em suas formulações, como no caso do MODE se considerarmos a reparametrização $\mathbf{\Pi}_A^\perp = \mathbf{\Pi}_B$. A diferença entre eles consiste em uma melhor estimativa para a porção da matriz de correlação associada ao subespaço de sinal, dada pelo método MODE através da presença da matriz de ponderação \mathbf{W}_{MODE} , fornecendo $\tilde{\mathbf{R}}_s = \hat{\mathbf{U}}_s \mathbf{W}_{MODE} \hat{\mathbf{U}}_s^H$, além da reparametrização $\mathbf{\Pi}_A^\perp = \mathbf{\Pi}_B$. Uma limitação desse método, também observada nos métodos baseados no ML, consiste na exigência de posto cheio para a matriz \mathbf{P} e, conseqüentemente, $\mathbf{A}\mathbf{P}\mathbf{A}^H$ para permitir a decomposição da matriz de correlação conforme em (6.50).

6.5.3 Método “Noise Subspace Fitting - NSF”

O método NSF consiste na implementação do critério de “distância” considerando uma aproximação baseada no subespaço de ruído [Krim96].

Mais especificamente, o método NSF consiste em resolver o problema de minimização

$$\min_{\phi} Tr \left\{ \mathbf{\Pi}_A \hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H \right\}, \quad (6.93)$$

conforme proposto em (3.43), na qual podemos reconhecer $\hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H = \hat{\mathbf{\Pi}}_A^\perp$. Podemos notar, ainda, que o problema em (6.93) é equivalente ao problema encontrado para o método MUSIC Ponderado no caso particular apresentado em (6.65).

Observa-se que o método NSF, assim como o método SSF, apresentam estimativas pouco confiáveis em um cenário em que as fontes de sinais são coerentes.

6.5.4 Método “Subspace Fitting” ponderado (WSF)

O critério utilizado no método SSF considera a hipótese que as matrizes \mathbf{A} e \mathbf{U}_s geram o mesmo espaço. Isso é verdade apenas quando as fontes de sinais que incidem sobre o arranjo são descorrelacionadas e quando a matriz \mathbf{P} possui posto cheio e igual a M . Para um cenário em que \mathbf{P} possui posto $M' < M$, as colunas da matriz \mathbf{U}_s geram um subespaço M' -dimensional da matriz \mathbf{A} . Essa observação também se aplica ao método NSF quando consideramos a hipótese que as colunas da matriz \mathbf{U}_n geram o espaço ortogonal ao gerado pelas colunas de \mathbf{U}_s , denominado de espaço nulo da matriz \mathbf{A} .

Uma generalização para o problema em (6.92) consiste em adicionar uma matriz positiva definida de ponderação \mathbf{W}_{SSF} , levando ao seguinte problema de minimização [Krim96] e [Viberg91b]:

$$\min_{\phi} Tr \left\{ \mathbf{\Pi}_A^\perp \hat{\mathbf{U}}_s \mathbf{W}_{SSF} \hat{\mathbf{U}}_s^H \right\}. \quad (6.94)$$

Podemos notar que, para uma escolha adequada da matriz \mathbf{W}_{SSF} , o problema de minimização em (6.94) é equivalente ao problema encontrado para os métodos baseados no ML, sobretudo o método MODE se considerarmos a reparametrização $\mathbf{\Pi}_A^\perp = \mathbf{\Pi}_B$ [Ottersten92]. Por esse motivo as considerações com respeito a melhor escolha para a matriz \mathbf{W}_{SSF} seguem os resultados apresentados para a matriz de ponderações \mathbf{W}_{MODE} do método MODE.

Quando utilizamos a relação $\mathbf{\Pi}_A^\perp = \mathbf{\Pi}_B$ através da reparametrização do problema em (6.94) surge o método denominado de root-WSF. Tal método consiste na solução do método WSF através da busca das raízes do polinômio $b(z)$ formado pelos elementos da matriz de reparametrização \mathbf{B}^H . Neste caso, a utilização do algoritmo de solução proposto para o método MODE se aplica diretamente, fornecendo os mesmos resultados.

A consideração de uma matriz de ponderação pode também ser estendida para o método NSF, dando origem ao problema de minimização

$$\min_{\phi} Tr \left\{ \mathbf{\Pi}_A \hat{\mathbf{U}}_n \mathbf{W}_{NSF} \hat{\mathbf{U}}_n^H \right\}. \quad (6.95)$$

Podemos notar que, para uma escolha adequada da matriz \mathbf{W}_{NSF} , o problema de minimização em (6.95) é equivalente ao problema encontrado para o método MUSIC Ponderado da forma em (6.64) com $\mathbf{W}_{MUSIC} = \mathbf{W}_{NSF}$.

Podemos, então, concluir que os métodos baseados em subespaços podem ser descritos através de uma mesma formulação mais geral e que essa formulação, por sua vez, permite explicitar comparações com os métodos baseados no critério ML. A comparação dessas duas classes de métodos, baseados em subespaço e no critério ML, com a classe de métodos baseados na predição linear é realizada através da reparametrização do problema original na busca das raízes de um polinômio associado a um filtro de erro de predição.

Como alternativa para a obtenção da solução ML exata através de um problema de busca M -dimensional foi observado em todos os métodos a utilização dos procedimentos de otimização LS e TLS, e suas variações WTLS, CTLS e STLS. Nesses procedimentos uma solução aproximada, geralmente de norma-mínima, é obtida para o problema original.

6.6 Resumo Comparativo

Métodos Baseados no ML

$$\min Tr \left\{ \left[\mathbf{B}(\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H \right] \hat{\mathbf{R}} \right\}, \text{ com } L = M.$$

IQML:

$$\hat{\mathbf{R}} = \sum_{k=1}^K \mathbf{Y}^H(k) \mathbf{Y}(k) \quad \text{e} \quad \mathbf{B}^H \mathbf{y}(k) = \mathbf{Y}(k) \mathbf{b}', \text{ para } k = 1, 2, \dots, K.$$

$$\min_{\mathbf{b}'} \left\{ (\mathbf{b}')^H \sum_{k=1}^K \mathbf{Y}^H(k) (\mathbf{B}^H \mathbf{B})^{-1} \mathbf{Y}(k) (\mathbf{b}') \right\}, \text{ com } L = M.$$

MODE:

$$\hat{\mathbf{R}} = \left(\hat{\mathbf{U}}_s \left[\hat{\mathbf{\Lambda}}_s - \hat{\mathbf{\Lambda}}_n \right] \hat{\mathbf{U}}_s^H \right) \mathbf{e} = \hat{\sigma}^2 \mathbf{I}.$$

$$\min Tr \left\{ \mathbf{B} \left(\hat{\mathbf{B}}^H \hat{\mathbf{B}} \right)^{-1} \mathbf{B}^H \left(\hat{\mathbf{U}}_s \left[\hat{\mathbf{\Lambda}}_s - \hat{\sigma}^2 \mathbf{I} \right] \hat{\mathbf{U}}_s^H \right) \right\}, \text{ com } L = M.$$

Métodos Baseados na Predição Linear

$$\min Tr \left\{ \left[\mathbf{B} \mathbf{B}^H \right] \hat{\mathbf{R}} \right\}, L \geq M.$$

Prony - FLP, BLP, FBLP - FBLP Modificado:

$$\min_{\mathbf{b}'} \left\{ (\mathbf{b}')^H \left[(\mathbf{Y}')^H \mathbf{Y}' \right] (\mathbf{b}') \right\}, L \geq M \text{ e } K = 1.$$

TLS-LP - TLS-FBLP - WTLS-LP:

$$\min_{\mathbf{b}'} \left\{ (\mathbf{b}')^H \left[(\mathbf{Y}')^H (\mathbf{B}^H \mathbf{B})^{-1} \mathbf{Y}' \right] (\mathbf{b}') \right\}, L \geq M \text{ e } K = 1.$$

Métodos Baseados em Subespaços

Ortogonalidade:

MUSIC Espectral:

$$\min_{\phi} Tr \left\{ \mathbf{A}(\phi) \mathbf{A}^H(\phi) \left(\hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H \right) \right\}.$$

Root-MUSIC:

$$\min Tr \left\{ \mathbf{B} \mathbf{B}^H \left(\hat{\mathbf{U}}_s \hat{\mathbf{U}}_s^H \right) \right\}, L = M.$$

Norma-mínima:

$$\min Tr \left\{ \mathbf{B} \mathbf{B}^H \left(\hat{\mathbf{U}}_s \hat{\mathbf{U}}_s^H \right) \right\}, \min \|\mathbf{b}\| \text{ e } L = N.$$

“Distância”:

$$\min_{\phi, \mathbf{T}} \left\| \hat{\mathbf{U}}_s - \mathbf{A}(\phi)\mathbf{T} \right\|_F^2 = \min_{\phi} Tr \left\{ (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger) \hat{\mathbf{U}}_s \hat{\mathbf{U}}_s^H \right\}$$

ESPRIT:

$$\min_{\Psi} \left\| \hat{\mathbf{U}}_{y_1} \Psi - \hat{\mathbf{U}}_{y_2} \right\|_E^2$$

SSF:

$$\min_{\phi} Tr \left\{ \Pi_{\mathbf{A}}^\perp(\phi) \hat{\mathbf{U}}_s \hat{\mathbf{U}}_s^H \right\}, L = M.$$

NSF:

$$\min_{\phi} Tr \left\{ \Pi_{\mathbf{A}}(\phi) \hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H \right\}, L = M.$$

WSF:

$$\begin{cases} \min_{\phi} Tr \left\{ \Pi_{\mathbf{A}}^\perp(\phi) \hat{\mathbf{U}}_s \mathbf{W}_{SSF} \hat{\mathbf{U}}_s^H \right\} \\ \min_{\phi} Tr \left\{ \Pi_{\mathbf{A}}(\phi) \hat{\mathbf{U}}_n \mathbf{W}_{NSF} \hat{\mathbf{U}}_n^H \right\} \end{cases}, L = M.$$

6.7 Generalização dos Métodos

$$\begin{cases} \min Tr \left\{ \Pi_s^\perp(\phi) \hat{\Pi}_s \right\} \\ \min Tr \left\{ \Pi_n(\phi) \Pi_s^\perp \right\} \\ \min Tr \left\{ \Pi_s(\phi) \hat{\Pi}_n \right\} \\ \min Tr \left\{ \Pi_n^\perp(\phi) \hat{\Pi}_n \right\} \end{cases}$$

$$\begin{aligned} \hat{\Pi}_s &= \hat{\mathbf{U}}_s \hat{\mathbf{U}}_s^H \cong \hat{\mathbf{R}} \\ \hat{\Pi}_n &= \hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H \\ \Pi_s(\phi) &= \Pi_n^\perp(\phi) = \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H = \mathbf{I} - \mathbf{B}(\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H \\ \Pi_n(\phi) &= \Pi_s^\perp(\phi) = \mathbf{I} - \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H = \mathbf{B}(\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H \end{aligned}$$

6.8 Conclusão

A necessidade de alta-resolução das aplicações atuais de estimação do parâmetro DOA exige métodos com capacidade de estimar parâmetros próximos entre si mesmos em baixas relações sinal-ruído. O método da máxima-verossimilhança (ML), apesar de perfeitamente aplicável ao problema, apresenta esforço computacional proibitivo. Os métodos que surgiram mais recentemente como ESPRIT, MUSIC, Norma-Mínima, IQML e MODE procuram ser computacionalmente viáveis, sem contudo comprometer seus desempenhos quando comparados com o método ML.

Apesar desses métodos possuírem origens e hipóteses distintas para as suas concepções, os mesmos foram aqui interpretados como a busca das estimativas que minimizam critérios equivalentes ao ML.

A diferença entre o método ML e os métodos IQML e MODE pode ser vista como um procedimento de reparametrização visando a redução do esforço computacional requerido pelo critério ML. Particularmente, o método MODE pode ser interpretado como um método direto derivado do algoritmo IQML, no qual as iterações foram suprimidas através do uso de estimativas iniciais adequadas.

Por sua vez, os métodos MUSIC e Norma-mínima, historicamente concebidos anteriormente aos métodos IQML e MODE, puderam ser descritos como alternativas para o critério ML baseado em sub-espacos. Particularmente, o método Norma-mínima foi interpretado como um caso particular do método MUSIC, no qual é utilizada uma ponderação adequada na composição do critério de minimização. Através dessa descrição foi possível compreender que, apesar da redução do esforço computacional oferecida por esses métodos, quando comparados com os demais, os mesmos não foram concebidos para um cenário no qual os sinais são coerentes, principalmente em baixa relação sinal-ruído.

Foi ainda possível estabelecer uma relação entre os métodos MUSIC, IQML e MODE, na qual o método MUSIC pode ser visto como o resultado do primeiro passo do processo de minimização comum aos métodos IQML e MODE, constatando-se assim a superioridade desses últimos.

Os métodos baseados em predição linear tratados aqui compreendem os métodos de estimação de parâmetros através de modelos lineares de representação para o sinal. O problema de se estimar a frequência de um sinal composto por exponenciais complexas ou a determinação da direção de chegada de ondas planas foram considerados indistintamente através da possibilidade do mapeamento entre esses dois parâmetros.

As formulações tradicionais utilizadas nos métodos de predição linear foram apresentadas, bem como suas formulações baseadas na distinção entre os subespaço de sinal e de ruído. Dessa forma foi possível relacionar não apenas métodos tradicionais de estimação de parâmetros como também os métodos modernos, baseados em subespaços, seja na propriedade de ortogonalidade ou através do critério de “distância” entre os subespaços.

Foi constatado que o conceito de predição linear pode ser estendido aos métodos de estimação de parâmetros baseados em subespaços, através da decomposição em valores singulares da matriz de dados associada ao problema. Nessa interpretação, os métodos que fazem uso do

princípio da ortogonalidade entre os subespaços de sinal e de ruído, como MUSIC e Norma-mínima, foram associados ao conceito de um filtro preditor otimizado para o sinal observado, dando uma nova interpretação ao método root-MUSIC. Da mesma forma, os métodos baseados no critério de “distância” entre os subespaços também podem ser interpretados através de conceito de predição linear através da mesma reparametrização utilizada para os métodos baseados no critério ML, dada pela relação entre as matrizes de projeção $\mathbf{\Pi}_A^\perp = \mathbf{\Pi}_B$.

Os princípios da predição linear também foram encontrados em métodos de estimação de parâmetros baseados na máxima-verossimilhança (ML). Nesse caso, a ponte entre os dois princípios de estimação foi possível através da reparametrização do problema original e do conceito de subespaços.

Concluimos, portanto, que os diversos métodos de estimação de parâmetros encontrados na literatura possuem fortes relações com o conceito básico de predição linear, seja através da ordem do filtro utilizado, número de amostras ou da forma de utilização da distinção entre os subespaços que compõem o sinal observado. Tais variações para o mesmo conceito influenciam diretamente a solução para o problema e, conseqüentemente, as características de resolução, complexidade numérica e desempenho do estimador obtido.

Capítulo 7

Conclusão

7.1 Resumo das Conclusões

A necessidade de alta-resolução nas aplicações atuais de estimação do parâmetro DOA exige métodos com capacidade de estimar parâmetros próximos entre si mesmo em baixas relações sinal-ruído. A correlação eventualmente existente entre as fontes de sinais, aliada à pequena quantidade dos dados disponíveis, torna tal problema de estimação complexo e com ampla aplicação. O método da máxima-verossimilhança (ML), apesar de perfeitamente aplicável ao problema, apresenta esforço computacional proibitivo. O grande desafio para qualquer método alternativo é procurar ser computacionalmente viável, sem contudo comprometer seu desempenho quando comparado ao método ML.

Como resultado de uma ampla pesquisa, classificamos e propomos a descrição dos métodos alternativos ao método ML em três categorias. A primeira consiste dos métodos baseados no próprio critério ML, a segunda compreende os métodos baseados na predição linear e a terceira representa os métodos baseados em subespaços.

Apesar desses métodos possuírem origens e hipóteses distintas para as suas concepções, os mesmos foram aqui interpretados e descritos como a busca das estimativas que minimizam critérios semelhantes.

Com relação à primeira categoria de métodos, a diferença entre a implementação direta do critério ML e os métodos alternativos IQML e MODE foi apresentada como um procedimento de reparametrização visando a redução do esforço computacional. Particularmente, o método MODE foi interpretado como um procedimento de estimação direto derivado do algoritmo iterativo IQML, no qual as iterações foram eliminadas através do uso de estimativas auxiliares mais próximas dos seus valores ótimos. Por esse motivo o método MODE não sofre problemas de convergência como o método IQML, além de apresentar menor esforço computacional, embora ambos exijam a minimização de uma função custo de quarta ordem. Além dessas características o método MODE faz uso de uma versão da matriz de correlação composta apenas pela porção associada ao subespaço de sinal.

Os métodos baseados em predição linear foram apresentados através de sua formulação tradicional, utilizando o mesmo polinômio $b(z)$ adotado na reparametrização dos métodos IQML e MODE. Foi também apresentada a sua formulação alternativa baseada na distinção entre os subespaços de sinal e de ruído da matriz de dados associada ao problema. Dessa forma foi possível relacioná-los não apenas aos métodos baseados no ML, como também aos métodos

modernos baseados em subespaços.

Além disso, ficou evidente que a utilização de informações adicionais em sua formulação, como a distinção entre os subespaços utilizada pelo método FBLP Modificado ou a característica de filtro branqueador utilizada no método WTLS, permite a obtenção de melhores resultados sem um aumento considerável do esforço computacional.

A comparação com os métodos baseados no ML permitiu constatar que o processo de reparametrização, adotado pelos métodos IQML e MODE, pode ser visto como a utilização de um filtro preditor de ordem igual ao número de parâmetros desconhecidos. Essa relação revelou ainda a menor complexidade relativa dos métodos baseados na predição linear, bem como sua menor capacidade de resolução quando os sinais são correlacionados e possuem baixa relação sinal-ruído. Uma vantagem da predição linear em relação a esses métodos é a possibilidade do aumento da ordem do preditor para melhorar a capacidade de resolução. Este trabalho apresentou ainda a proposta de extensão do procedimento de predição linear sobre vários experimentos através da estimação de uma matriz de correlação adequada.

Quanto à terceira categoria de métodos, os métodos MUSIC e Norma-mínima, concebidos anteriormente aos métodos IQML e MODE, puderam ser descritos como alternativas do critério ML baseadas em subespaços. Os métodos Norma-mínima e Pisarenko foram interpretados como casos particulares do método MUSIC. Através dessa descrição geral, foi possível compreender que, apesar da redução do esforço computacional oferecida por esses métodos, através de uma função custo de segunda ordem, os mesmos não foram concebidos para sinais correlacionados, principalmente em baixa relação sinal-ruído.

Quando comparados com os métodos baseados na predição linear, foi possível uma nova interpretação para o método MUSIC que deu origem ao método root-MUSIC.

Foi ainda possível estabelecer uma relação entre os métodos MUSIC, IQML e MODE, na qual o método MUSIC foi identificado como o resultado do primeiro passo do processo de minimização comum aos métodos IQML e MODE, constatando-se, assim, a superioridade de desempenho desses últimos e o menor esforço computacional do método MUSIC.

A apresentação dos métodos baseados em subespaços que utilizam um critério de “distância” entre o subespaço de sinal e sua aproximação, como no caso do SSF e sua versão baseada no subespaço de ruído, NSF, permitiu constatar a equivalência desse último com o método MUSIC. Foi ainda possível classificar o método ESPRIT nessa categoria devido à semelhança em suas formulações. Uma generalização para esse critério foi obtida adicionando-se uma matriz de ponderações em sua formulação. O método denominado de WSF faz uso dessa ponderação e apresenta a mesma função custo do método MODE.

Concluimos que os diversos métodos de estimação de parâmetros encontrados na literatura possuem fortes relações com o conceito básico de predição linear ou com o critério ML, seja através da ordem do filtro utilizado, da forma de utilização dos dados disponíveis ou do uso da distinção entre os subespaços que compõem o sinal observado. Tais variações para um mesmo conceito influenciam diretamente na solução do problema original e, conseqüentemente, nas características de resolução, complexidade numérica e de desempenho do estimador obtido. Particularmente, a forma de manipular os subespaços se mostrou de grande importância na obtenção de melhores estimativas.

Esse trabalho evidenciou, portanto, a vasta possibilidade de relações entre os métodos existentes. A classificação dos métodos em três categorias e a utilização de uma formulação unificada para descrevê-los se apresentam como suas principais contribuições.

7.2 Sugestões para o Prosseguimento do Trabalho

Baseado nas idéias básicas presentes nos métodos IQML e MODE, podemos sugerir algumas novas alternativas que deixamos para pesquisas futuras.

Embora tenhamos desenvolvido o critério ML para o modelo estocástico do sinal, os métodos descritos fazem uso apenas do critério ML determinístico. Baseado nessa observação, podemos sugerir a implementação de um método baseado no critério ML estocástico. Acreditamos que tal critério, embora mais complexo, é mais adequado considerando a observação de um sinal prático. A forma de implementação dessa proposta que evita uma busca multidimensional está sob investigação.

Uma outra proposta é a utilização de um modelo linear com ordem elevada na reparametrização utilizada pelo IQML e MODE, fazendo $L > M$. Desta forma, acreditamos ser possível melhorar a resolução desses métodos, a exemplo do que ocorre nos métodos baseado na predição linear. A inserção ou modificação das restrições impostas nesses métodos, sobretudo através de considerações com respeito aos subespaços, também podem contribuir para o surgimento de novas alternativas de métodos baseados no critério ML.

Considerando os procedimentos de minimização apresentados, podemos sugerir a utilização dos procedimentos LS ou TLS através da SVD como alternativas para a solução do problema de minimização enfrentado pelos métodos MODE e IQML. Dessa forma, será possível a implementação da restrição de não-trivialidade implicitamente, além de se mostrar uma alternativa de solução eficiente e robusta computacionalmente. Tais soluções, que geralmente possuem norma-mínima, possibilitará ainda o contraste com a solução ML obtida pela implementação direta do critério ML.

Baseado nas idéias centrais presentes nos métodos de predição linear apresentados, sugerimos uma alternativa que consiste na adição de uma matriz de ponderação \mathbf{W} ao método WTLS-LP, a qual está presente nos procedimentos CTLS e STLS. Estudos, ainda em andamento, levantam suspeitas de que tal ponderação não está presente no procedimento WTLS por tal procedimento pressupor a existência de um ruído branco e não estruturado. No entanto, devido à ação do filtro preditor sobre as amostras de sinal ruidoso, teremos tais amostras estruturadas. Nesse caso, a inserção da matriz de ponderação se torna necessária e natural para satisfazer a condição de ruído branco e não estruturado. Tal ponderação também não é observada nos métodos IQML e MODE. Considerando a semelhança entre esses métodos, tal sugestão também pode ser estendida aos mesmos. No entanto, suspeitamos que tal ponderação esteja implicitamente presente na função custo derivada para o método MODE através da matriz de ponderação \mathbf{W}_{MODE} .

Baseado na equivalência da formulação entre os métodos WTLS-LP e IQML, podemos sugerir a utilização de seus algoritmos de solução para ambos os métodos indistintamente. A diferença na inicialização do algoritmo WTLS, quando comparado com os outros algoritmos, sugere a possibilidade de obtenção da solução ótima em dois passos apenas, como no algoritmo MODE.

Considerando a existência de uma versão do método WTLS-LP baseado em subespaço, podemos sugerir uma restrição no subespaço semelhante àquela presente no método FBLP Modificado, dando origem ao método WTLS-LP Modificado. Por se tratar de um método baseado em predição linear, podemos ainda considerar a utilização de um preditor progressivo-regressivo com o objetivo de melhorar a precisão das estimativas obtidas, dando origem ao

método WTLS-FBLP. Através da semelhança de formulação entre os métodos WTLS-LP e MODE, tais sugestões podem também ser aplicadas à reparametrização presente nos métodos IQML e MODE, dando origem a uma nova variação para esses métodos.

Quanto aos métodos baseados em subespaços, uma sugestão consiste na utilização de uma matriz de ponderação sobre o método NSF, dando origem ao método NSF Ponderado ou WNSF, a exemplo do método WSF.

Através da relação entre os problemas de otimização enfrentados pelos métodos apresentados e o procedimento TLS de minimização, surge a possibilidade de utilizar os procedimentos CTLS ou STLS como um aprimoramento na solução desses métodos.

Novas propostas ainda estão sob investigação e serão objeto de trabalhos futuros.

Conforme foi enfatizado no início desse trabalho, nos restringimos à comparações teóricas entre os métodos apresentados. Para comprovar tais constatações sugerimos, portanto, a implementação dos algoritmos e a obtenção de resultados numéricos via simulação computacional. Acreditamos que tais resultados não apenas confirmarão as expectativas de desempenho e esforço computacional como também comprovará a utilidade da descrição unificada apresentada.

Referências Bibliográficas

- [Abatzoglou91] Abatzoglou, T. J., Mendel, J. M. e Harada, G. A.: “The Constrained Total Least Squares Techniques and its Applications to Harmonic Superresolution.”, IEEE Trans. on Signal Processing, Vol. 39, Maio 1991.
- [Alves2000a] Alves, C. A., Lopes, A. e Colares, R. F.: “Constraints Implementation for IQML and MODE Direction-of-Arrival Estimators”, 43rd Midwest Symposium on Circuits and Systems - MWSCAS 2000, Michigan State University (MSU), Michigan, USA, Ago. 2000.
- [Alves2000b] Alves, C. A., Colares, R. F. e Lopes, A.: “Análise de Métodos para DOA em Função de Restrições de Otimização”, XVIII Simpósio Brasileiro de Telecomunicações, Hotel Serra Azul, Gramado, RS, Set. 2000.
- [Antunes92] Antunes, E. J. B.: “Detecção de Frequências Através de Predição Linear.”, Tese de Mestrado, UNICAMP, Campinas-SP, Maio 1992.
- [Bresler86] Bresler, Y., e Macovski, A.: “Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise,” IEEE Trans. on Acoustics, Speech, and S. P., Vol. 34, Out. 1986.
- [Box76] Box, G. E. P. e Jenkins, G. M.: “Time Series Analysis: Forecasting and Control.”, Prentice-Hall, 2a Ed., 1976.
- [Buckley90] Buckley, K. e Xu, X. L.: “Spatial-Spectrum Estimation in a Location Sector.”, IEEE Trans. on Acoustic, Speech and S. P., Vol. 38, Nov. 1990.
- [Cadzow90] Cadzow, J. A.: “Multiple Source Location-The Signal Subspace Approach.”, IEEE Trans. on Acoustic, Speech and S. P., Vol. 38, Jul. 1990.
- [Colares97] Colares, R. F.: “Aprimorando Método para Detecção de Frequências Baseado na Norma-mínima”, Tese de Mestrado, UNICAMP, Campinas-SP, Set. 1997.
- [Colares98] Colares, R. F. e Lopes, A.: “Improving The Threshold Of The Minimum-Norm Frequency Detection Method.”, anais do International Telecommunications Symposium - ITS'98, São Paulo, Brazil, Ago. 1998
- [Colares99a] Colares, R. F. e Lopes, A.: “Improving DOA Estimation Methods Using a Priori Knowledge About The Sources Location”, 42nd Midwest Symposium on Circuits and Systems, New Mexico State University, Novo México, USA, Ago. 1999.

- [Colares99b] Colares, R. F. e Lopes, A.: “Visão Unificada dos Métodos Modernos para DOA Através de Arranjos de Sensores Lineares e Uniformes ”, XVII Simpósio Brasileiro de Telecomunicações, Parthenon-Pasárgada Residence - Praia da Costa, Vila Velha, ES, Set. 1999.
- [Colares2000a] Colares, R. F., Alves, C. A. e Lopes, A.: “Proposta para Aplicação do Método WTLS em DOA ”, XVIII Simpósio Brasileiro de Telecomunicações, Hotel Serra Azul, Gramado, RS, Set. 2000.
- [Colares2000b] Colares, R. F., Alves, C. A. e Lopes, A.: “Descrição Unificada dos Métodos Modernos para Estimação D.O.A. em Arranjo de Sensores ”, Submetido a Revista da Sociedade Brasileira de Telecomunicações - SBrT em agosto de 2000.
- [DeMoor90] De Moor, B. e Vandewalle, J.: “A Unifying Theorem for Linear and Total Linear Least Squares.”, IEEE Trans. on Automatic Control, Vol. 35, Maio 1990.
- [DeMoor93a] De Moor, B.: “Generalizations of the Singular Value and QR Decompositions.”, SIAM J. on Matrix Analysis and Appl., 1993.
- [DeMoor93b] De Moor, B.: “Strutured Total Least Squares and L2 Approximation Problem.”, Linear Algebra and its Applications, Special Issue on Numerical Linear Algebra Methods in Control, Signal and Systems (Van Dooren, Ammar, Nichols, Mehrmann, Eds.), Vol. 188-189, 1993.
- [DeMoor94] De Moor, B.: “Total Least Squares for Affinely Structured Matrices and the Noisy Realization Problem.”, IEEE Trans. on Signal Processing, Vol. 42, Nov. 1994.
- [Dickinson79] Dickinson, B.: “Efficient Solution of Linear Equations with Banded Toeplitz Matrices.”, IEEE Trans. on Accoust., Speech, Signal Processing, Vol. 27, Ago. 1979.
- [Dowling96] Dowling, E. M. e DeGroat, R. D.: “The Equivalence of the Total Least Squares and Minimum Norm Methods.”, IEEE Trans. on Signal Processing, Vol. 39, Ago. 1996.
- [Ermolaev94] Ermolaev, V. T. e Flaksman, A. G.: “Fast Algorithm for Minimum-Norm Direction-of-Arrival Estimation”, IEEE Trans. on Signal Processing, Vol. 42, Set. 1994.
- [Golub80] Golub, G. H. e Van Loan, C. F.: “An Analysis of the Total Least Squares Problem.”, SIAM J. Numer. Anal., Vol. 17, No. 6, Dez. 1980.
- [Golub89] Golub, G. H. e VanLoan, C. F.: “Matrix Computation” , Johns Hopkins University Press, Baltimore, MD., 2a. ed., 1989.

- [Haykin85] Haykin S.: "Array Signal Processing", Prentice-Hall, Englewood Cliffs, New Jersey, 1985.
- [Haykin89] Haykin, S.: "Modern Filters.", MacMillan Ed., 1989.
- [Hua90] Hua, Y. e Sarkar, T. K.: "On the Total Least Squares Linear Prediction Method for Frequency Estimation.", IEEE Trans. on Acoust. Speech, and Signal Processing, Vol. 38, Dez 1990.
- [Jansson99] Jansson, M., Göransson, B. e Ottersten, B.: "A Subspace Method for Direction of Arrival Estimation of Uncorrelated Emitter Signals.", IEEE Trans. on Signal Processing, Vol. 47, Abril 1999.
- [Kaveh86] Kaveh, M e Barabell, A. J.: "The Statistical Performance of the MUSIC and the Minimum-Norm Algorithms in Resolving Plane Waves in Noise.", IEEE Trans. on Acoustic, Speech and S. P., Vol. 34, Abril 1986.
- [Kay88] Kay, S. M.: "Modern Spectral Estimation: Theory and Application.", Prentice-Hall, Englewood Cliffs, New Jersey, 1988.
- [Kay93] Kay, S. M.: "Fundamentals of Statistical Processing, Volume I: Estimation Theory.", Prentice-Hall, Englewood Cliffs, New Jersey, 1993.
- [Krim92] Krim, H., Forster, P. e Proakis, J. G.: "Operator Approach to Performance Analysis of Root-MUSIC and Root Min-Norm.", IEEE Trans. on Acoustic, Speech and S. P., Vol. 40, Jul. 1992.
- [Krim96] Krim, H e Viberg, M.: "Two Decades of Array Signal Processing Research", IEEE Signal Processing Magazine, Vol. 13, No. 4, Jul., 1996.
- [Kumaresa83a] Kumaresan, R. e Tufts, D. W.: "Estimating the Angles of Arrival of Multiple Plane Waves.", IEEE Trans on Aerospace and Electronic Systems, Vol. 19, Jan. 1983.
- [Kumaresa83b] Kumaresan, R.: "On the Zeros of the Linear Prediction-error Filter for Deterministic Signals.", IEEE Trans. on Acoustic, Speech e S. P., Vol. 71, Fev. 1983.
- [Kumaresa84] 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.", Proc. of IEEE, Vol. 72, Set. 1984.
- [Lemmerlin96] Lemmerling, P., De Moor, B. e Van Huffel, B.: "On the Equivalence of Constrained Total Least Squares and Structured Total Least Squares.", IEEE Trans. on Signal Processing, Vol. 44, Nov. 1996.
- [Lemos97] Lemos, R. P.: "Mínimos-quadrados Totais e Máxima-verossimilhança em Estimacão de Freqüências.", Tese de Doutorado, UNICAMP, Campinas-SP, Jun. 1997.

- [Li98] Li, J., Stoica, P. e Liu, Z.: “Comparative Study of IQML and MODE Direction-of-Arrival Estimators.”, IEEE Trans. on Acoust., Speech, Signal Processing, Vol. 46, Jan. 1998.
- [Lueberger73] Lueberger, D.: “An Introduction to Linear and Nonlinear Programming.”, Reading, MA, Addison-Wesley, 1973.
- [Makhoul75] Makhoul, J.: “Linear Prediction: A Tutorial Review.”, Proceedings of the IEEE, Vol.63, Abril 1975.
- [Ottersten92] Ottersten, B., Viberg, M. e Kailath, T.: “Analysis of Subspace Fitting and ML Techniques for Parameter Estimation from Sensor Array Data.”, IEEE Trans. on Signal Processing, Vol . 40, Mar. 1992.
- [Ottersten93] Ottersten, B., Viberg, M., Stoica, P. e Nehorai, A.: “Exact and Large Sample ML Techniques for Parameter Estimation and Detection in Array Processing ”, Haykin, Litva, e Shepherd, editors, Radar Array Processing, pg. 99-151. Springer-Verlag, Berlin, 1993.
- [Papoulis91] Papoulis, A.: “Probability, Random Variables, and Stochastic Processes”, 3a. ed., New York, NY, MacGraw-Hill, 1991.
- [Paulraj86] Paulraj, A., Roy, R. e Kailath, T.: “Subspace Rotational Approach to Signal Parameter Estimation.”, Proceedings of the IEEE, Jul. 1986.
- [Paulraj97] Paulraj, A. J. e Papadias, C. B.: “Space-Time Processing for Wireless Communications”, IEEE Signal Processing Magazine, Nov., 1997.
- [Pillai89] Pillai S. Unnikrishna: “Array Signal Processing”, Springer-Verlag, New York, 1989.
- [Pisarenko73] Pisarenko, V. F.: “The Retrieval of Harmonics from a Covariance Function.”, Geophys. J. Roy. Astron. Soc., Vol. 33, 1973.
- [Rahman87] Rahman, M. A. e Yu, K.B.: “Total Least Squares Approach for Frequency Estimation using Linear Prediction.”, IEEE Trans. on Acoust., Speech, Signal Processing, Vol. 35, Out. 1987.
- [Reddi79] Reddi, M. M.: “Multiple Source Location - A Digital Approach.”, IEEE Trans. on Aerospace, Eletronic and System, Vol. 15, Jan. 1979.
- [Roy89] Roy, R. e Kailath, T.: “ESPRIT - Estimation of Signal Parameters via Rotational Invariance Techniques.”, IEEE Trans, on Acoustic, Speech and S. P., Vol. 37, Jul. 1989.
- [Sacchi98] Sacchi M. D., Ulrych T. J. e Walker C. J.: “Interpolation and Extrapolation Using a High-Resolution Discrete Fourier Transform”, IEEE Transactions on Signal Processing, Vol. 46, No. 1, Jan. 1998.

- [Schmidt86] Schmidt, R. O.: "Multiple Emitter Location and Signal Parameter Estimation," IEEE Trans. on Ant. and Prop., vol. 34, 1986.
- [Shan85] Shan, T. J., Wax, M. e Kailath, T.: "On Spatial Smoothing for Direction of Arrival Estimation of Coherent Signal.", IEEE Trans. on Acoust. Speech, Signal Processing, Vol. 33, No. 4, Abril 1985.
- [Shamsunde94] Shamsunder, S. e Giannakis, G.: "Signal Selective Localization of NonGaussian Cyclostationary Sources ", IEEE Transaction on Signal Processing, Vol. 42, Out. 1994.
- [Sheinvald96] Sheinvald, J., Wax M.e Weiss, A. J.: "On Maximum-Likelihood Localization of Coherent Signals.", IEEE Trans. on Signal Processing, Vol. 44, Out. 1996.
- [Sorenson70] Sorenson, H. W.: "Least-squares estimation: from Gauss to Kalman.", IEEE Spectrum, Julho 1970.
- [Stoica89] Stoica, P. e Nehorai, A.: "MUSIC, Maximum Likelihood and Cramér-Rao Bound.", IEEE Trans. on Acoustic, Speech and S. P., Vol. 37, Maio 1989.
- [Stoica90a] Stoica, P. e Sharman, K.: "A Novel Eigenanalysis Method for Direction Estimation.", Proc. of IEE, Part. F, Fev. 1990.
- [Stoica90b] Stoica, P. e Sharman, K. C.: "Maximum Likelihood Methods for Direction-of-Arrival Estimation," IEEE Trans. on Acoustics, Speech, and S. P., Vol. 38, No. 7, Jul. 1990.
- [Stoica90c] Stoica, P. e Nehorai, A.: "Performance Study of Conditional and unconditional Direction-of-Arrival Estimation ", IEEE Trans. on Acoustic, Speech and Signal Processing, Vol.38, No. 10, Out. 1990.
- [Stoica90d] Stoica, P. e Nehorai, A.: "MUSIC, Maximum Likelihood and Cramér-Rao Bound: Futher Results and Comparisons", IEEE Trans. on Acoustic, Speech and S. P., Vol. 38, Dec. 1990.
- [Stoica91] Stoica, P. e Söderström, T.: "Statistical Analysis of MUSIC and Subspace Rotational Estimates of Sinusoidal Frequencies.", IEEE Trans. on Acoustic, Speech and S. P., Vol. 39, Ago. 1991.
- [Stoica97] Stoica, P., Li, J. e Söderström, T.: "On the Inconsistency of IQML", Signal Processing, Elsevier Science B. V., No. 56, 1997.
- [Tufts82] Tufts, D. W., e Kumaresan, R.: "Estimation of Frequencies of Multiple Sinusoids: Making Linear Prediction Perform Like Maximum Likelihood" Proc. of the IEEE, Vol. 70, No. 9, Set. 1982.
- [VanHuffel91] Van Huffel, S. e Vandewalle, J.: "The Total Least Squares Problem: Computational Aspects and Analysis.", SIAM Frontiers in Applied Mathematics, Vol. 9, Philadelphia, 1991.

- [VanHuffel93] Van Huffel, S., De Moor, B. e Chen.H.: “Relationships Between Structured and Constrained TLS, with Applications to Signal Enhancement.”, System and Networks: Mathematical Theory and Applications; Proceedings of the Int. Symposium MTNS-93, Regensburg, Alemanha, Ago. 1993.
- [Viberg91a] Viberg, M. e Ottersten, B.: “Sensor Array Processing Based on Subspace Fitting.” IEEE Trans. on Signal Processing, Vol. 39, Maio 1991.
- [Viberg91b] Viberg, M., Ottersten, B. e Kailath, T.: “Detection and Estimation in Sensor Array Using Weighted Subspace Fitting.” IEEE Trans. on Signal Processing, Vol.39, Nov. 1991.
- [Wax91] Wax, M.: “Detection and Localization of Multiple Sources Via the Stochastic Signals Model.”, IEEE Trans. on Signal Processing, Vol. 39, Nov. 1991.
- [Zatman98] Zatman, M.: “How Narrow is Narrowband.”, IEE Proc. on Radar, Sonar and Navig., Vol 145, No. 2, Abril 1998.