THOMAS EDSON FILGUEIRAS FILHO

SOBRE A VELOCIDADE DE CONVERGÊNCIA DA
FILTRAGEM ADAPTATIVA IIR

Tese apresentada à Escola Politécnica
da Universidade de São Paulo para
a obtenção do Título de Doutor em
Engenharia.

São Paulo

2008

THOMAS EDSON FILGUEIRAS FILHO

SOBRE A VELOCIDADE DE CONVERGÊNCIA DA
FILTRAGEM ADAPTATIVA IIR

Tese apresentada à Escola Politécnica
da Universidade de São Paulo para
a obtenção do Título de Doutor em
Engenharia.

Área de Concentração:
Sistemas Eletrônicos

Orientador:
Prof.Dr
Phillip Mark Seymour Burt

São Paulo

2008

Este exemplar foi revisado e alterado em relação à versão original, sob responsabilidade única do autor e com anuência de seu orientador.

São Paulo, 6 de Dezembro de 2008

Assinatura do autor

Assinatura do orientador

FICHA CATALOGRÁFICA

Filgueiras Filho, Thomas Edson

Sobre a Velocidade de Convergência da Filtragem Adaptativa IIR, São Paulo, São Paulo, 2008.
132p.

Tese (Doutorado) – Escola Politécnica da Universidade de São Paulo. Departamento de Telecomunicações e Controle.

1. Filtros Elétricos Adaptativos I. Universidade de São Paulo. Escola Politécnica. Departamento de Telecomunicações e Controle II.t.

Para Breno, pela companhia em todos os momentos, Eliane, pelo carinho e dedicação, e para meus pais pelo constante suporte e incentivo.

AGRADECIMENTOS

Agradeço ao Prof. Phillip M. S. Burt pela ajuda e orientação além da revisão minuciosa do texto com sugestões e críticas valiosas que foram fundamentais para a confecção deste trabalho.

Aos Profs. Maria D. Miranda e Vítor Heloiz Nascimento, pela leitura do texto e sugestões fornecidas no exame de qualificação.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ), pelo suporte financeiro durante parte deste trabalho.

A Escola Politécnica do Universidade de São Paulo por todas as oportunidades oferecidas ao longo de todos estes anos.

RESUMO

Filtros adaptativos com resposta ao impulso infinita (IIR) podem substituir com vantagens aqueles com respostas ao impulso finitas (FIR). Entre estas vantagens está o seu reduzido número de parâmetros que leva a uma menor complexidade computacional na obtenção de respostas similares. Porém, a utilização de filtros adaptativos IIR apresenta alguns problemas práticos, sendo o mais destacado sua convergência lenta. Este problema aparece principalmente quando algoritmos baseados no gradiente são utilizados para a adaptação dos coeficientes do filtro. A abordagem baseada na teoria de realização balanceada de sistemas, previamente utilizada para se analisar filtros com entrada branca, se mostrou uma ferramenta útil para entender o que faz com que um filtro convirja lentamente. Este método já foi aplicado com sucesso na análise de filtros adaptativos IIR com entrada branca nas configurações de identificação e de identificação inversa. Neste trabalho aplicaremos este mesmo método para o caso de entrada não branca. Será mostrado que a configuração de identificação inversa é um caso particular da configuração de identificação com entrada não-branca, podendo ambas serem tratadas conjuntamente. Também será mostrado que o sistema que controla as propriedades de convergência não é mais o sistema desconhecido que se está tentando identificar, e sim um sistema relacionado a este e a densidade espectral da entrada. No caso de entrada branca, esta análise levou ao algoritmo de aproximações sucessivas, o qual, utilizando um bloco auxiliar, tenta fazer com que o filtro adaptativo enxergue um sistema de rápida convergência. Será apresentada uma generalização deste algoritmo para o caso de entrada não-branca, inclusive serão apontadas limitações do mesmo quanto a valores dos passos de adaptação. Simulações numéricas serão usadas para ilustrar todos os resultados obtidos.

ABSTRACT

Adaptive filters with infinite impulse response (IIR) can replace with advantages the ones with finite impulse response (FIR). One of these advantages is the reduced number of parameters which leads to a smaller computational complexity giving similar responses. However, the use of adaptive IIR filters has some practical issue, being the most prominent its slow convergence. This issue is mainly seen when gradient descent algorithms are applied to adapt the filters coefficients. The approach based on the balanced realization of systems, previously used to analyze the convergence speed of adaptive IIR filter with white input, has seemed to be a useful tool in understanding what causes the slow convergence in a filter. This approach was successful applied to the analyses of the identification and inverse identification configurations. In this work we will apply this same approach to the non-white input case. It will be shown that the inverse identification configuration is a special case of the identification configuration with non-white input, so both can be addressed together. It will also be shown that the convergence properties are no more set by the characteristics of the unknown system, but by the characteristics of a system related to it and the input spectral density function. In the white input case, the results of this analysis were used to propose the successive approximations algorithm, which uses an auxiliary block trying to make the adaptive filter sees a system with faster convergence. A more general form of this algorithm that includes the non-white input case will be presented, and some drawbacks regarding the adaptation stepsize will be pointed out. Numerical simulations will be used to illustrate all the obtained results.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SÍMBOLOS

1	Introdução	1
1.1	Definição do problema de filtragem adaptativa IIR	1
1.2	Abordagens de filtragem adaptativa IIR	4
1.3	Contribuições originais	5
1.4	Organização do trabalho	6
2	Aspectos de aproximação racional $\mathcal{H}_2$ e filtragem adaptativa IIR	8
2.1	Visão geral do capítulo	8
2.2	Os espaços de funções e produto interno	9
2.2.1	Os espaços L_2 e $\mathcal{H}_2$	10
2.2.2	Propriedades do produto interno	11
2.3	Tópicos de sistemas lineares	13
2.3.1	Espaço de estados e função de transferência	13
2.3.2	Gramianos de controlabilidade e de observabilidade	14
2.3.3	Valores singulares de Hankel e realização balanceada	17
2.3.4	Sistemas com entrada estocástica	18
2.4	Teoria de Beurling-Lax	19
2.4.1	Subespaço ortogonal a vetor de controlabilidade	19
2.4.2	Base para $\mathcal{H}_2$	23
2.5	Aproximação racional $\mathcal{H}_2$	25
2.5.1	Decomposição do erro quadrático	26
2.5.2	Superfície de erro reduzida	27
2.5.3	Pontos estacionários da superfície de erro quadrático	29
2.5.4	Erro quadrático médio	32
2.5.5	Extensão para a norma L_2 ponderada	33
2.5.6	Identificação com entrada de espectro racional	33
2.5.7	Identificação inversa	35
2.6	Estrutura de filtros IIR	36
2.6.1	Forma direta	36
2.6.2	Forma em treliça normalizada	39

2.7	Algoritmos de filtragem adaptativa IIR baseados no gradiente descendente	40
2.7.1	Forma direta	40
2.7.2	Forma em treliça normalizada	43
2.7.3	Identificação inversa com filtro na forma direta	45
2.8	Análise da velocidade de convergência com entrada branca	46
2.8.1	Parâmetros BLP	49
2.8.2	Descrição local da adaptação	50
2.9	Algoritmo de aproximações sucessivas (SA)	52
2.9.1	Definição do algoritmo	53
2.9.2	Exemplo numérico	55
2.10	Velocidade de convergência de filtros adaptativos IIR na configuração inversa	55
2.11	Conclusão	58
3	Efeitos de uma entrada de espectro racional	60
3.1	Dependência da velocidade de convergência com relação ao espectro da entrada	60
3.2	Dependência com relação aos zeros de $H(z)$	65
3.3	Dependência com relação à realização de $\hat{H}(z)$	67
3.4	Identificação inversa com entrada não-branca	70
3.5	Conclusão e motivação para análise	70
4	Velocidade de convergência com entrada de espectro racional	72
4.1	Unificação das configurações de identificação inversa e identificação com entrada não-branca	73
4.2	Cálculo dos zeros ótimos	73
4.3	Base para a decomposição do erro quadrático	76
4.4	Velocidade de convergência determinada por $H_p(z)$	79
4.5	A superfície de erro restrita	80
4.5.1	Forma da superfície de erro restrita	80
4.5.2	Exemplos Numéricos	83
4.6	Propriedades do sistema projetado $H_p(z)$	85
4.6.1	Forma de Hankel ponderada	85
4.6.2	Norma de $G(z)H(z)$ e relação com os passos de adaptação	87
4.7	Condição para pequeno espalhamento dos autovalores de $\mathbf{J}\mathbf{J}^t\mathbf{\Sigma}^2$	91
4.7.1	Igualdade dos valores singulares	91
4.7.2	Exemplos numéricos	92
4.7.2.1	Teste do efeito de $H_p(z)$ passa-tudo	92
4.7.2.2	Variação de $C_s(z)$ com a mudança dos pólos de $H_s(z)$	102
4.8	Algoritmo de aproximações sucessivas com entrada não-branca	105
4.8.1	Descrição do algoritmo	105
4.8.2	Exemplo numérico	107
4.9	Conclusão	108

5	Conclusão e trabalhos futuros	110
5.1	Resultados obtidos	110
5.2	Trabalhos futuros	111
A	Algoritmo de filtragem adaptativa IIR na forma direta	112
B	Algoritmo de filtragem adaptativa IIR na forma em treliça normalizada	115
C	Algoritmo de filtragem adaptativa IIR na forma direta em configuração de identificação inversa	119
D	Método de identificação utilizando o erro de equação	122
E	Resultados da teoria de matrizes	126
E.1	Invariância de autovalores à permutação	126
E.2	Positividade de autovalores	126
E.3	Fatoração e raiz quadrada	127
E.4	Positividade de autovalores de produto de matrizes	127
E.5	SVD	127
E.6	Computação de $\mathbf{K}$ e $\mathbf{W}$	128
E.7	Computação da realização balanceada	128

LISTA DE FIGURAS

1.1	Filtro adaptativo em (A) configuração de identificação e (B) configuração de identificação inversa.	2
2.1	Filtro adaptativo em configuração de identificação.	32
2.2	Configuração de identificação com entrada de espectro racional dada por $u(n) = G(z)w(n)$	34
2.3	Filtro adaptativo na configuração de identificação inversa.	35
2.4	Filtro na forma direta obtido a partir da equação de diferenças.	37
2.5	Filtro mínimo obtido a partir da forma da Figura 2.4.	38
2.6	Forma em treliça normalizada.	39
2.7	Geração dos sinais regressores filtrados no algoritmo na forma direta.	42
2.8	Obtenção dos regressores em filtros adaptativos na forma em treliça normalizada.	44
2.9	Geração dos sinais regressores filtrados no algoritmo de identificação inversa na forma direta.	47
2.10	Exemplo de convergência lenta de filtros adaptativos IIR. Os parâmetros utilizados se encontram na Tabela 2.1 junto com o inverso dos pólos e zeros da função sendo identificada.	48
2.11	Adaptação dos parâmetros dos zeros e seus valores ótimos dados os pólos ($\circ$).	49
2.12	Identificação de um sistema $H(z)$ passa-tudo onde $H(z^{-1})$ tem pólos em $0,8\angle \pm 45^\circ$ e $0,8\angle \pm 110^\circ$ e zeros recíprocos a estes.	53
2.13	Diagrama em blocos do algoritmo de aproximações sucessivas para filtragem adaptativa IIR.	54
2.14	Erro obtido com o algoritmo de aproximações sucessivas (A) e com um filtro adaptativo na forma direta adaptado pelo gradiente estocástico (B) na identificação do sistema $H(z)$ conforme a Tabela 2.3.	56
2.15	Exemplo de convergência lenta de filtros adaptativos IIR na configuração de identificação inversa. O inverso dos pólos e zeros da função $H(z)$ sendo identificada se encontram na Tabela 2.4 junto com os parâmetros utilizados na simulação.	57
2.16	Pólos (" $\circ$ ") e zeros (" $\times$ ") de $H(z^{-1})$ (esquerda) e $\hat{H}_I(z^{-1})$ (direita).	57
3.1	Resposta em frequência de $H(z)$ (linha pontilhada) e de $G(z)$ (linha contínua).	62
3.2	Identificação de um sistema $H(z)$ passa-tudo com os pólos e zeros da Tabela 3.1 e entrada não-branca.	62

3.3	Identificação de um sistema passa-tudo utilizando (A) entrada branca e (B) entrada não-branca.	63
3.4	Resposta em frequência de $H(z)$ (linha pontilhada) e de $G(z)$ (linha contínua) para pólos e zeros da Tabela 3.2.	64
3.5	Resultado da simulação numérica cujas funções e parâmetros se encontram na Tabela 3.2 utilizando (A) entrada não-branca e (B) entrada branca.	64
3.6	Resposta em frequência de $H_1(z)$ (A), $H_2(z)$ (B) e de $G(z)$ (C) da Tabela 3.3.	66
3.7	Identificação dos sistemas $H_1(z)$ (A) e $H_2(z)$ (B), cujos pólos e zeros se encontram na Tabela 3.3 junto com os do filtro $G(z)$ usado para gerar a entrada $u(n)$	66
3.8	Identificação do sistema passa tudo $H(z)$ cujos pólos e zeros se encontram na Tabela 3.4. A curva (A) mostra o erro quadrático médio obtido com entrada não-branca e a (B) com entrada branca. Em ambos os casos o filtro adaptativo IIR utilizado se encontra na forma em treliça.	68
3.9	Identificação do sistema da tabela 3.5 utilizando filtro adaptativo em treliça. A curva (A) mostra o erro quadrático médio obtido com entrada não-branca e (B) o obtido com entrada branca. . . .	69
3.10	Erro obtido na identificação inversa de $H(z)$	71
4.1	(A) - Caso particular da configuração de identificação com entrada não-branca dada por $u(n) = H(z)w(n)$ com sistema sendo identificado igual a $H_I(z) = 1/H(z)$. (B) - Passo intermediário para a obtenção da configuração de identificação inversa. (C) - Configuração de identificação inversa.	74
4.2	Parâmetros $b_k(n)$ (linhas sólidas) e seus valores ótimos $b_{ok}(n)$ (quadrados e linhas pontilhadas), deslocados como indicado para melhor visualização.	76
4.3	Pólos (“o”) e zeros (“x”) de $H(z^{-1})$ (esquerda) e $H_p(z^{-1})$ (direita).	83
4.4	$E[e^2(n)]$ para o algoritmo de gradiente exato e $e^2(n)$ para o algoritmo de gradiente estocástico (traços superiores), parcela $\sum_{k=0}^{M+L} (\bar{p}_k - \bar{v}_k)^2$ de $E[e^2(n)]$ (traço inferior).	84
4.5	Evolução dos autovalores $\lambda_k(n)$ de $\mathbf{J}\mathbf{J}^t \mathbf{\Sigma}^2$ associada a $H_p(z)$	84
4.6	Evolução dos componentes $\beta_k(n)$ de $\boldsymbol{\alpha}(n)$ nas direções dos autovetores associados aos autovalores $\lambda_k(n)$	85
4.7	Pólos (“o”) e zeros (“x”) de $H(z^{-1})$ (esquerda) e $H_s(z^{-1})$ (direita).	93
4.8	Módulo das respostas em frequência de $G(z)$ (linha sólida) e de $H_s(z)$ (linha tracejada).	93
4.9	Identificação de $H_1(z)$ com entrada branca: $E[e^2(n)]$ para o algoritmo de gradiente exato e $e^2(n)$ para o algoritmo de gradiente estocástico (traços superiores), parcela $\sum_{k=0}^{M+L} (\bar{p}_k - \bar{v}_k)^2$ de $E[e^2(n)]$ (traço inferior).	94
4.10	Identificação de $H_1(z)$ com entrada branca: autovalores $\lambda_k(n)$ de $\mathbf{J}\mathbf{J}^t \mathbf{\Sigma}^2$ associado a $H_p(z)$	95

4.11	Identificação de $H_1(z)$ com entrada branca: componentes $\beta_k(n)$ de $\alpha(n)$ nas direções dos autovetores associados aos autovalores $\lambda_k(n)$.	95
4.12	Identificação de $H_s(z)$ com entrada colorida: $E[e^2(n)]$ para o algoritmo de gradiente exato e $e^2(n)$ para o algoritmo de gradiente estocástico (traços superiores), parcela $\sum_{k=0}^{M+L} (\bar{\rho}_k - \bar{\nu}_k)^2$ de $E[e^2(n)]$ (traço inferior).	96
4.13	Identificação de $H_s(z)$ com entrada colorida: autovalores $\lambda_k(n)$ de $\mathbf{J}\mathbf{J}^t\mathbf{\Sigma}^2$ associado a $H_p(z)$	96
4.14	Identificação de $H_s(z)$ com entrada colorida: componentes $\beta_k(n)$ de $\alpha(n)$ nas direções dos autovetores associados aos autovalores $\lambda_k(n)$.	97
4.15	Identificação de $H_1(z)$ com entrada branca, $\mu_A = 0,005$: $E[e^2(n)]$ para o algoritmo de gradiente exato e $e^2(n)$ para o algoritmo de gradiente estocástico (traços superiores), parcela $\sum_{k=0}^{M+L} (\bar{\rho}_k - \bar{\nu}_k)^2$ de $E[e^2(n)]$ (traço inferior).	98
4.16	Identificação de $H_1(z)$ com entrada branca, $\mu_A = 0,005$: autovalores $\lambda_k(n)$ de $\mathbf{J}\mathbf{J}^t\mathbf{\Sigma}^2$ associado a $H_p(z)$	99
4.17	Identificação de $H_1(z)$ com entrada branca, $\mu_A = 0,005$: componentes $\beta_k(n)$ de $\alpha(n)$ nas direções dos autovetores associados aos autovalores $\lambda_k(n)$.	99
4.18	Identificação de $H_s(z)$ com entrada colorida, $\mu_A = 0,005$: $E[e^2(n)]$ para o algoritmo de gradiente exato e $e^2(n)$ para o algoritmo de gradiente estocástico (traços superiores), parcela $\sum_{k=0}^{M+L} (\bar{\rho}_k - \bar{\nu}_k)^2$ de $E[e^2(n)]$ (traço inferior).	100
4.19	Identificação de $H_s(z)$ com entrada colorida, $\mu_A = 0,005$: autovalores $\lambda_k(n)$ de $\mathbf{J}\mathbf{J}^t\mathbf{\Sigma}^2$ associado a $H_p(z)$	100
4.20	Identificação de $H_s(z)$ com entrada colorida, $\mu_A = 0,005$: componentes $\beta_k(n)$ de $\alpha(n)$ nas direções dos autovetores associados aos autovalores $\lambda_k(n)$.	101
4.21	Identificação de $H_1(z)$ com entrada branca, $\mu_A = 0,02$: $E[e^2(n)]$ para o algoritmo de gradiente exato e $e^2(n)$ para o algoritmo de gradiente estocástico (traços superiores), parcela $\sum_{k=0}^{M+L} (\bar{\rho}_k - \bar{\nu}_k)^2$ de $E[e^2(n)]$ (traço inferior).	102
4.22	Identificação de $H_1(z)$ com entrada branca, $\mu_A = 0,02$: autovalores $\lambda_k(n)$ de $\mathbf{J}\mathbf{J}^t\mathbf{\Sigma}^2$ associado a $H_p(z)$	102
4.23	Identificação de $H_1(z)$ com entrada branca, $\mu_A = 0,02$: componentes $\beta_k(n)$ de $\alpha(n)$ nas direções dos autovetores associados aos autovalores $\lambda_k(n)$.	103
4.24	Módulo das respostas em frequência de $G(z)$ (linha sólida) e de $H_s(z)$ (linha tracejada) com um novo conjunto de pólos.	104
4.25	Identificação de $H_s(z)$ com entrada não-branca: $e^2(n)$ para o algoritmo de gradiente estocástico.	104
4.26	Diagrama em blocos do algoritmo de aproximações sucessivas para o caso de entrada não-branca.	106
4.27	Algoritmo de aproximações sucessivas (A) e algoritmo de gradiente na forma direta (B).	108

A.1	Geração dos sinais regressores filtrados no algoritmo na forma direta.	113
B.1	Obtenção dos regressores em filtros adaptativos na forma em treliça normalizada.	116
C.1	Geração dos sinais regressores filtrados no algoritmo na forma direta.	120

LISTA DE TABELAS

2.1	Pólos e zeros de $H(z^{-1})$ junto com os parâmetros utilizados na simulação da Figura 2.10.	47
2.2	Inverso dos pólos e zeros do sistema $H(z)$ sendo identificado junto com os parâmetros do filtro adaptativo IIR usado na simulação da Figura 2.12.	52
2.3	Pólos e zeros de $H(z^{-1})$ junto com os parâmetros utilizados nas simulações da Figura 2.14.	55
2.4	Inverso dos pólos do sistema $H(z)$ sendo identificado e parâmetros usados na simulação da Figura 2.15.	56
3.1	Pólos e zeros do sistema sendo identificado e do filtro utilizado para gerar a entrada não-branca junto com os parâmetros do filtro adaptativo IIR.	61
3.2	Pólos e zeros do sistema sendo identificado e do filtro utilizado para gerar a entrada não-branca, junto com os parâmetros do filtro adaptativo IIR no caso em que a entrada não-branca possibilita uma convergência mais rápida.	63
3.3	Pólos e zeros do sistemas sendo identificados e do filtro $G(z)$ utilizado para gerar a entrada não-branca. Parâmetros dos filtros adaptativos IIR.	65
3.4	Pólos e zeros do sistema sendo identificado e do filtro utilizado para gerar a entrada não-branca junto com os parâmetros dos filtros adaptativos IIR. Neste caso os valores de $B_0(z)$ e $A_0(z)$ foram convertidos para os parâmetros do filtro em treliça.	67
3.5	Pólos e zeros do sistema sendo identificado e do filtro utilizado para gerar a entrada não-branca junto com os parâmetros dos filtros adaptativos IIR.	69
3.6	Inverso dos pólos e zeros do sistema $H(z)$ sendo identificado e do filtro $G(z)$ utilizado para gerar a entrada não-branca junto com os parâmetros do filtro adaptativo IIR usado para a identificação de $H(z)$	70
4.1	Pólos e zeros das funções utilizadas no exemplo da Figura 3.2 junto com suas respectivas normas.	90
4.2	Pólos e zeros das funções utilizadas no exemplo da Figura 3.5 junto com suas respectivas normas.	90

4.3	Inverso dos pólos e zeros do sistema $H(z)$ sendo identificado e do filtro $G(z)$ utilizado para gerar a entrada não-branca junto com os parâmetros do filtro adaptativo IIR na forma direta e com o algoritmo SA usado na simulação numérica da Figura 4.27.	107
-----	---	-----

LISTA DE SÍMBOLOS

n	Índice de tempo
$x(n)$	Função do tempo n
z	Variável complexa. Corresponde ao atraso unitário ($zu(n) = u(n - 1)$)
L_2	Espaço de funções estáveis
$\mathcal{H}_2$	Espaço de funções estáveis e causais
$f(z)$	Função da variável complexa z
$\langle f(z), w(z) \rangle$	Produto interno entre as funções $f(z)$ e $w(z)$
$\langle f(z), w(z) \rangle_{S_u}$	Produto interno entre as funções $f(z)$ e $w(z)$ ponderado pela função $S_u(z)$
$\ f(z)\ $	Norma de $f(z)$
$\ f(z)\ _{S_u}$	Norma de $f(z)$ ponderada por $S_u(z)$
$H(z)$	Sistema sendo identificado
$\hat{H}(z)$	Filtro adaptativo
$\deg[f(z)]$	Ordem de $f(z)$. Corresponde ao número mínimo de retardos necessários para implementar $f(z)$
$\mathcal{C}$	Matriz de controlabilidade de horizonte infinito
$\mathbf{K}$	Gramiano de controlabilidade
$\mathcal{C}(z)$	vetor de controlabilidade de $(\mathbf{A}, \mathbf{b})$
$\mathcal{O}$	Matriz de observabilidade de horizonte infinito
$\mathbf{W}$	Gramiano de observabilidade
$E[\cdot]$	Operador esperança
$E[e^2(n)]$	Erro quadrático médio
$\mathbf{x}$	Vetor coluna (letra minúscula em negrito)
$\mathbf{X}$	Matriz (letra maiúscula em negrito)
Γ_H	Forma de Hankel associada ao sistema $H(z)$
σ_k	Autovalores da forma de Hankel

Σ	Matriz diagonal cujo elementos são os autovalores σ_k
α	Vetor com os parâmetros da forma balanceada de um sistema
$\mathbf{J}$	Matriz de sensibilidade
b_k	Coefficientes do polinômio do denominador da forma direta
a_k	Coefficientes do polinômio do numerador da forma direta
θ_k	Ângulos de rotação da forma em treliça
ν_k	Pesos de cada estágio da forma em treliça
μ	Passo de adaptação
μ_A	Passo de adaptação dos parâmetros dos pólos da forma direta
μ_B	Passo de adaptação dos parâmetros dos zeros da forma direta
μ_ν	Passo de adaptação dos parâmetros ν_k da forma em treliça
μ_θ	Passo de adaptação dos parâmetros θ_k da forma em treliça
n_a	Número de amostras utilizadas na fase de adaptação do algoritmo de aproximações sucessivas
n_x	Número de amostras utilizadas na fase de aproximação do algoritmo de aproximações sucessivas
n_t	Número de amostras utilizadas na fase de transição do algoritmo de aproximações sucessivas
n_w	Número de amostras utilizadas na fase de espera do algoritmo de aproximações sucessivas
$H_p(z)$	Sistema projetado
$H_s(z)$	Sistema com zeros calculados para que $H_p(z)$ seja passa-tudo
s, c	Parâmetros do sistema $H_p(z)$ passa-tudo usados no cálculo de $H_s(z)$
$\mathcal{C}$	Matriz de controlabilidade de horizonte infinito
$\mathbf{K}$	Gramiano de controlabilidade
$\mathcal{C}(z)$	vetor de controlabilidade de $(\mathbf{A}, \mathbf{b})$
$\mathcal{O}$	Matriz de observabilidade de horizonte infinito
$\mathbf{W}$	Gramiano de observabilidade

CAPÍTULO 1 INTRODUÇÃO

Filtros cuja função de transferência é racional são capazes de fornecer longas respostas impulsivas com um número pequeno de parâmetros. Estes filtros são conhecidos como filtros IIR (*Infinite Impulse Response*) pois, pelo menos em teoria, esta resposta é infinita. Esta característica permite que estes filtros tenham respostas bem mais seletivas que um filtro FIR (filtro de resposta ao impulso finita ou filtro só de zeros) com o mesmo número de parâmetros, o que os torna uma opção vantajosa sobre estes.

Do mesmo modo, filtros adaptativos IIR podem substituir filtros adaptativos FIR com vantagens sobre seu desempenho [1]-[5], e, devido ao pequeno número de parâmetros a serem adaptados, com menor complexidade computacional. Porém, certos problemas, que não existem no caso de filtros adaptativos FIR, desencorajam em princípio a sua utilização: o filtro pode se tornar instável, a superfície de erro pode apresentar mínimos locais e, mesmo para entradas bem condicionadas, a adaptação pode atingir regiões onde a convergência é lenta. A seguir definiremos o problema de filtragem adaptativa e depois discutiremos como estes problemas vêm sendo abordados ao longo do tempo por vários pesquisadores (por exemplo [6]-[20]).

1.1 Definição do problema de filtragem adaptativa IIR

Consideremos a saída $y(n) = H(z)u(n) + \eta(n)$ de um sistema $H(z)$ submetido a uma entrada $u(n)$, sendo o ruído $\eta(n)$ independente de $u(n)$. Nesta notação mista, z é o atraso unitário de modo que $zu(n) = u(n - 1)$. Esta notação foi escolhida

Figura 1.1: Filtro adaptativo em (A) configuração de identificação e (B) configuração de identificação inversa.

por ser compacta e visualmente agradável. Além disto, há a conveniência dos expoentes de z serem iguais aos correspondentes índices da resposta impulsiva.

Conhecendo apenas $u(n)$ e $y(n)$ podemos tentar identificar o sistema $H(z)$ adaptando um filtro $\hat{H}(z)$ de modo a minimizar o erro quadrático médio $E[e^2(n)]$, onde $e(n) = y(n) - \hat{y}(n)$ e $\hat{y}(n) = \hat{H}(z)u(n)$ é a saída do filtro adaptativo. Este caso corresponde à configuração de identificação mostrada na Figura 1.1(A). Os parâmetros do filtro $\hat{H}(z)$ são atualizados em função de $y(n)$ e do erro $e(n)$ através de um algoritmo recursivo. Quando a entrada $u(n)$ é branca o problema de filtragem adaptativa corresponde à minimização do quadrado da norma L_2 $\|H(z) - \hat{H}(z)\|^2$.

Podemos também adaptar um filtro $\hat{H}_I(z)$ para identificar o sistema $H(z)$ minimizando o erro quadrático médio $E[(u(n) - \hat{u}(n))^2]$, onde $\hat{u}(n) = \hat{H}_I(z)y(n) = \hat{H}_I(z)H(z)u(n)$. O subscrito I indica que o sistema inverso $H_I(z) = 1/H(z)$ está sendo identificado. Esta configuração de identificação inversa é mostrada na Figura 1.1(B).

Se no caso de identificação a função $\hat{H}(z)$ for um polinômio de grau M ,

$$\hat{H}(z) = h_0 + h_1z + h_2z^2 + \dots + h_Mz^M,$$

teremos um filtro adaptativo FIR. Caso $\hat{H}(z)$ seja racional de ordem M ,

$$\hat{H}(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1z + \dots + b_Mz^M}{1 + a_1z + \dots + a_Mz^M}$$

teremos um filtro adaptativo IIR. O mesmo vale para $\hat{H}_I(z)$ no caso de identificação inversa. Quando o filtro é implementado na forma direta, os próprios coeficientes b_k e a_k são adaptados. Em outras estruturas, como a em treliça, são adaptados parâmetros relacionados com estes coeficientes [5]. Podemos notar que em $A(z)$ o primeiro coeficiente é $a_0 = 1$, ou seja, $A(z)$ é mônico. No restante deste trabalho sempre consideraremos que o denominador das funções racionais envolvidas é mônico.

Quando o sistema sendo identificado possui ordem menor do que a de $\hat{H}(z)$ (ou $\hat{H}_I(z)$), temos o caso de *sobremodelamento*, onde há mais parâmetros adaptados do que no sistema real. Caso ambas as ordens sejam iguais, temos o caso de *modelamento suficiente*, sendo que nestas duas situações é possível a identificação exata do sistema. Já quando a ordem do sistema sendo identificado for maior do que a do filtro adaptativo, temos um *submodelamento*, onde não podemos mais identificar exatamente o sistema, apenas obter uma aproximação deste.

Quando estivermos falando de entrada não-branca, consideraremos que esta é da forma $u(n) = G(z)w(n)$, onde $G(z)$ é um filtro qualquer e $w(n)$ é um sinal branco de variância unitária. Quando tivermos a função $G(z) = E(z)/F(z)$, ou seja, uma função racional, denominaremos a entrada $u(n)$ como *entrada de espectro racional*. Esta denominação está ligada à função densidade espectral da entrada que será abordada mais adiante. De uma maneira geral, não podemos, a partir da adaptação de $\hat{H}(z)$ com entrada branca, inferir o comportamento da adaptação com entrada não-branca. Em algumas vezes a primeira terá melhores propriedades, em outras, a segunda.

1.2 Abordagens de filtragem adaptativa IIR

Como mencionado anteriormente, filtros adaptativos IIR podem apresentar um desempenho melhor que filtros adaptativos FIR devido à realimentação da saída [5, 3]. Isto é notado ao se tentar identificar sistemas que possuam pólos e zeros, como muitos sistemas físicos. Um filtro adaptativo FIR pode apenas aproximar este sistema enquanto um filtro adaptativo IIR pode identificá-lo perfeitamente com muito menos parâmetros que no caso FIR. Este reduzido conjunto de parâmetros leva a outra vantagem dos filtros adaptativos IIR, sua menor complexidade computacional.

Porém, filtros adaptativos IIR podem apresentar problemas que não existem no caso de filtros adaptativos FIR. O mais aparente destes problemas é a possibilidade do filtro se tornar instável durante a adaptação. Este problema pode ser eliminado com o uso de estruturas cujo controle de estabilidade é computacionalmente simples [5, 21].

Quanto aos mínimos locais que possam existir na superfície de erro, estes podem ser evitados utilizando algoritmos que não se baseiam no gradiente do erro. Outra maneira de se evitar estes mínimos locais é iniciar a adaptação em um ponto próximo ao ponto de mínimo como em [5, 10, 11], ou ainda utilizando algoritmos diferentes em cada etapa da adaptação [9, 12, 13]. Em [22] é mostrado que, se a ordem do sistema sendo identificado for conhecida, podemos escolher a ordem do filtro adaptativo de modo que não existam mínimos locais na superfície de erro.

Há ainda o problema da convergência lenta dos filtros adaptativos IIR. Pode-se aumentar a velocidade de convergência em certos casos utilizando estruturas não-diretas de filtragem [5] ou com o uso da estrutura polifásica proposta em [14]-[17]. A influência da posição dos pólos do sistema sendo identificado na velocidade de convergência na vizinhança de pontos de mínimo é analisada em [8]. No caso mais geral (vizinhança de pontos de mínimo ou não) esta dependência é analisada em [14]-[16].

Outro aspecto que merece atenção é a rápida degradação da velocidade de

convergência de algoritmos simples com o aumento da ordem do filtro. É facilmente verificado por meio de simulações que, mesmo na ausência de mínimos locais, filtros com ordem superior a 2 podem ficar praticamente estacionados em certas regiões do espaço de parâmetros onde o erro quadrático está longe de ser satisfatório. Algoritmos que buscam usar a inversa da matriz Hessiana [4], genericamente chamados de algoritmos de Newton, são menos suscetíveis a este efeito, mas apresentam uma maior complexidade computacional, eliminando uma das principais motivações para a utilização de filtros adaptativos IIR.

Em [23] é proposta uma parametrização dos pólos baseada na teoria de realização balanceada e é mostrada a relação entre a velocidade de convergência de algoritmos baseados no gradiente com o espalhamento dos valores singulares de Hankel do sistema sendo identificado e com a própria parametrização do pólos. Com estes resultados, em [24] é proposto o *algoritmo de aproximações sucessivas*, que, através da inclusão de um bloco auxiliar, combinado com o sistema que se deseja identificar, proporciona uma maior velocidade de convergência. Em [19] esta análise é estendida para a configuração de identificação inversa e em [20] uma versão modificada do algoritmo de aproximações sucessivas para esta configuração é apresentada.

1.3 Contribuições originais

A principal contribuição deste trabalho é a análise efetuada no Capítulo 4. Neste capítulo os resultados de [23] para a configuração de identificação e de [19] para a configuração de identificação inversa, ambas com entrada branca, são estendidos para quando a entrada é não-branca. Também é mostrado que a identificação inversa é um caso particular da identificação com entrada não-branca, sendo possível tratar ambas as configurações conjuntamente.

Nesta análise vemos que as propriedades de convergência de filtros adaptativos IIR com entrada não-branca são controladas pelas características de um sistema $H_p(z)$ relacionado com o sistema original $H(z)$ sendo identificado e com a densidade espectral da entrada. A partir disto foram obtidas condições para

que estes filtros convirjam mais rapidamente. Finalmente, algumas limitações do caso de entrada não-branca são abordadas.

Com os resultados obtidos o algoritmo de aproximações sucessivas foi modificado para esta nova situação. Este algoritmo, como no caso de entrada branca, propicia uma maior velocidade de convergência.

Os resultados das Seções 4.1 até 4.5 e parte dos resultados da Seções 4.7 e 4.8 fazem parte de [25], apresentado no *XVI European Signal Processing Conference* (EUSIPCO 2008) e que se encontra no final deste trabalho.

1.4 Organização do trabalho

A maior parte dos resultados referentes à velocidade de convergência de filtros adaptativos IIR citados anteriormente foi obtida para o caso de entrada branca. Neste trabalho vamos aplicar as mesmas técnicas para o caso de entrada não-branca, mais especificamente entradas de espectro racional, de modo a obtermos condições que possam fazer com que filtros submetidos a este tipo de entrada convirjam mais rapidamente.

A primeira parte deste trabalho consistirá das bases teóricas para o desenvolvimento que se seguirá. Desta forma, no Capítulo 2 são abordados os espaços de funções L_2 e $\mathcal{H}_2$ que utilizaremos, bem como as propriedades de seu produto interno. Também serão apresentados tópicos relativos a sistemas lineares, usados a seguir para mostrar alguns resultados úteis que nos permitirão analisar o erro quadrático obtido na aproximação racional de uma função qualquer. É mostrada a relação entre o problema da filtragem adaptativa IIR e o de aproximação racional, o que nos permite utilizar os resultados anteriores na análise do primeiro problema. Abordaremos a estrutura de filtros IIR e os algoritmos baseados no gradiente utilizados para a adaptação de seus coeficientes. Terminando o capítulo temos alguns resultados da análise de convergência de filtros adaptativos IIR efetuada em [23], notadamente a parametrização BLP dos pólos. Em seguida abordamos o algoritmo de aproximações sucessivas de [24].

No capítulo 3 são apresentados resultados de simulações numéricas mostrando

o que ocorre com a velocidade de convergência de filtros adaptativos IIR quando a entrada não é branca. O resultado destas simulações servirá de motivação para a análise que se seguirá no Capítulo 4. Neste capítulo mostraremos como a configuração de identificação inversa pode ser considerada um caso particular do caso de identificação com entrada não-branca, o que nos permitirá tratar ambos de maneira conjunta. Uma base para a decomposição do erro quadrático para esta situação será desenvolvida, permitindo estender alguns resultados do Capítulo 2 para cobrir estas entradas. Uma condição para que haja uma maior velocidade de convergência com entradas não-brancas é então obtida, permitindo alterar o algoritmo de aproximações sucessivas para este caso.

O capítulo 5 conclui o trabalho e apresenta alguns pontos em aberto para investigação futura. Alguns algoritmos úteis são apresentados nos apêndices de forma a fornecer informações sobre como foram feitos os cálculos das simulações realizadas.

CAPÍTULO 2 ASPECTOS DE APROXIMAÇÃO RACIONAL $\mathcal{H}_2$ E FILTRAGEM ADAPTATIVA IIR

A seguir revisaremos alguns resultados relativos à aproximação de uma função estável e causal $H(z)$ por uma função racional $\hat{H}(z)$ (mais abreviadamente, aproximação racional $\mathcal{H}_2$), assim como resultados da análise da velocidade de convergência de filtros adaptativos IIR. De modo a prover maior auto-suficiência ao texto, optamos por apresentar demonstrações de alguns destes resultados e também rever conceitos matemáticos importantes no contexto dos mesmos. Isto levou o capítulo a ser relativamente longo e diversificado, fazendo sua leitura ser possivelmente trabalhosa. Como forma de auxiliar esta leitura, então, apresentamos inicialmente um resumo mais detalhado do capítulo.

2.1 Visão geral do capítulo

Começamos na Seção 2.2 discutindo os espaços de funções L_2 (funções estáveis) e seu subespaço $\mathcal{H}_2$ (funções estáveis e causais). Introduzimos em seguida uma operação fundamental envolvendo as funções pertencentes a estes espaços —o produto interno— e apresentamos suas propriedades de maior interesse.

Funções racionais em $\mathcal{H}_2$ podem ser vistas como funções de transferência de sistemas dinâmicos lineares de dimensão finita, que é o tema da Seção 2.3. Neste âmbito, os conceitos de controlabilidade e observabilidade levam à teoria da realização balanceada e ao conceito de valores singulares de Hankel do sistema, que terão um papel fundamental na análise da velocidade de convergência de

filtros adaptativos IIR.

O conceito de controlabilidade, por sua vez, também é útil para uma formulação mais concreta da teoria de Beurling-Lax sobre espaços invariantes, discutida na Seção 2.4. Esta teoria leva a uma determinada base ortonormal completa para o espaço $\mathcal{H}_2$, que é outro elemento fundamental da análise da velocidade de convergência de filtros adaptativos IIR.

Uma primeira conexão dos conceitos discutidos com o problema de filtragem adaptativa IIR é realizada na Seção 2.5, que trata do problema da aproximação racional $\mathcal{H}_2$. A teoria de Beurling-Lax e os resultados derivados dela, vistos na Seção 2.4, permitem caracterizar a forma que devem ter as soluções deste problema. A obtenção iterativa destas soluções leva ao problema de filtragem adaptativa IIR propriamente dito. Para tanto, consideramos inicialmente, na Seção 2.6, diferentes estruturas de filtragem IIR e seus respectivos conjuntos de parâmetros a serem adaptados. Na Seção 2.7 são então discutidos algoritmos para adaptar estes parâmetros.

A partir da Seção 2.8, a apresentação, antes baseada em [5], passa a ser baseada em [24][23] e [21]. Nesta seção, o erro quadrático no contexto da aproximação racional $\mathcal{H}_2$ é reescrito usando conceitos da teoria da realização balanceada. Isto leva a uma nova parametrização dos pólos do filtro adaptativo, mais útil para a análise da velocidade de convergência de filtros adaptativos IIR (análise, até agora, limitada ao caso de entrada branca). Conforme discutido na Seção 2.9, estes resultados levam também a um novo algoritmo adaptativo IIR, que visa uma velocidade de convergência maior do que a dos algoritmos convencionais.

2.2 Os espaços de funções e produto interno

Nesta seção abordaremos os espaços de funções L_2 e $\mathcal{H}_2$ aos quais pertencem as funções que serão utilizadas ao longo deste trabalho. Após a apresentação destes espaços definiremos seu produto interno e apresentaremos algumas propriedades deste.

2.2.1 Os espaços L_2 e $\mathcal{H}_2$

O espaço L_2 tem como elementos funções $f(z)$ de uma variável complexa que têm norma L_2 finita, ou seja

$$\|f(z)\| \triangleq \left[\frac{1}{2\pi j} \oint_{|z|=1} |f(z)|^2 \frac{dz}{z} \right]^{1/2} < \infty,$$

sendo esta norma induzida pelo produto interno

$$\langle f(z), x(z) \rangle \triangleq \frac{1}{2\pi j} \oint_{|z|=1} f(z) x(z^{-1}) \frac{dz}{z},$$

onde $f(z), x(z) \in L_2$ e subentende-se que a integral de linha é feita no sentido anti-horário.

Uma função $f(z) \in L_2$ pode ser expandida em uma série de Fourier $\sum_{k=-\infty}^{\infty} f_k z^k$, $|z| = 1$, de modo que

$$\lim_{N \rightarrow \infty} \left\| f(z) - \sum_{k=-N}^N f_k z^k \right\| = 0, \quad |z| = 1.$$

Esta convergência em norma L_2 é sub-entendida ao escrevermos simplesmente

$$f(z) = \sum_{k=-\infty}^{\infty} f_k z^k, \quad |z| = 1.$$

Em certos momentos será conveniente separarmos uma função $f(z) \in L_2$ em suas projeções estritamente causal

$$[f(z)]_+ \triangleq \sum_{k=1}^{\infty} f_k z^k$$

e anti-causal

$$[f(z)]_{\ominus} \triangleq \sum_{k=-\infty}^0 f_k z^k,$$

o que nos permite escrever

$$f(z) = [f(z)]_+ + [f(z)]_{\ominus}.$$

Empregaremos ainda a notação $[f(z)]_{\oplus} \triangleq \sum_{k=0}^{\infty} f_k z^k$ para a projeção causal de $f(z)$ e $[f(z)]_- \triangleq \sum_{k=-\infty}^{-1} f_k z^k$ para sua projeção estritamente anti-causal.

Um subespaço importante de L_2 é o espaço de Hardy $\mathcal{H}_2$, definido como o conjunto das funções $f(z) \in L_2$ cuja expansão em série apresenta somente

componentes causais, ou seja

$$\mathcal{H}_2 = \left\{ f(z) \mid f(z) \in L_2, f(z) = \sum_{k=0}^{\infty} f_k z^k, |z| = 1 \right\}. \quad (2.1)$$

Em outras palavras, é o conjunto de funções que representam sistemas de tempo discreto lineares, invariantes no tempo, estáveis e causais.

Se $f(z)$ for uma função racional

$$f(z) = \frac{q_0 + q_1 z + \dots + q_M z^M}{1 + p_1 z + \dots + p_M z^M}$$

então $f(z) \in \mathcal{H}_2$ se as raízes de seu denominador (ou seja, seus pólos) se encontrarem fora do círculo unitário. Denominaremos $\deg[f(z)]$ o número mínimo de retardos necessários para implementar a função $f(z)$, e neste caso $\deg[f(z)] = M$.

2.2.2 Propriedades do produto interno

Como mencionado acima, o produto interno entre duas funções pertencentes a L_2 pode ser escrito como

$$\langle f(z), x(z) \rangle \triangleq \frac{1}{2\pi j} \oint_{|z|=1} f(z) x(z^{-1}) \frac{dz}{z} \quad (2.2)$$

considerando que a integral é feita no sentido anti-horário. A seguir apontaremos algumas propriedades deste produto interno que serão úteis no decorrer do trabalho.

Propriedade 2.1 *Sejam as funções $f(z) = z^k$, $k \in \mathbb{Z}$ e $x(z) = z^l$, $l \in \mathbb{Z}$, que pertencem a L_2 . O produto interno entre elas será*

$$\langle f(z), x(z) \rangle = \begin{cases} 1, & k = l \\ 0, & k \neq l \end{cases}.$$

A demonstração é direta utilizando a equação (2.2) e a troca de variáveis $z = e^{j\omega}$:

$$\begin{aligned}
\langle f(z), x(z) \rangle &= \frac{1}{2\pi j} \oint_{|z|=1} z^k z^{-l} \frac{dz}{z} \\
&= \frac{1}{2\pi j} \oint_{|z|=1} z^{k-l} \frac{dz}{z} \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j(k-l)\omega} d\omega, \quad z = e^{j\omega} \\
&= \begin{cases} 1, & k = l \\ 0, & k \neq l \end{cases}.
\end{aligned}$$

Propriedade 2.2 Para quaisquer funções $f(z), x(z), w(z) \in \mathcal{H}_2$, vale a seguinte propriedade:

$$\langle f(z), w(z)x(z) \rangle = \langle f(z)w(z^{-1}), x(z) \rangle. \quad (2.3)$$

Para provar esta propriedade é só utilizar a equação (2.2):

$$\begin{aligned}
\langle f(z), w(z)x(z) \rangle &= \frac{1}{2\pi j} \oint_{|z|=1} f(z)w(z^{-1})x(z^{-1}) \frac{dz}{z} \\
&= \langle f(z)w(z^{-1}), x(z) \rangle.
\end{aligned}$$

Propriedade 2.3 Sejam as funções $f(z) = \sum_{k=-\infty}^{\infty} f_k z^k$ e $x(z) = \sum_{k=-\infty}^{\infty} x_k z^k$ pertencentes a L_2 . Podemos escrever o produto interno entre elas como

$$\langle f(z), x(z) \rangle = \sum_{k=-\infty}^{\infty} f_k x_k.$$

Esta propriedade pode ser provada usando a equação (2.2):

$$\begin{aligned}
\langle f(z), x(z) \rangle &= \frac{1}{2\pi j} \oint_{|z|=1} f(z)x(z^{-1}) \frac{dz}{z} \\
&= \frac{1}{2\pi j} \oint_{|z|=1} \left(\sum_{k=-\infty}^{\infty} f_k z^k \right) \left(\sum_{l=-\infty}^{\infty} x_l z^{-l} \right) \frac{dz}{z} \\
&= \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \left(\frac{1}{2\pi j} \oint_{|z|=1} f_k x_l z^{k-l} \frac{dz}{z} \right) \\
&= \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} f_k x_l \left(\frac{1}{2\pi j} \oint_{|z|=1} z^{k-l} \frac{dz}{z} \right) \\
&= \sum_{k=-\infty}^{\infty} f_k x_k.
\end{aligned}$$

Sendo que a última passagem decorre da Propriedade 2.1.

Propriedade 2.4 *Seja $f(z) = \sum_{k=0}^{\infty} f_k z^k$ uma função causal e $x(z) = \sum_{k=-\infty}^0 x_k z^k$ uma função anti-causal. O produto interno entre as duas funções será*

$$\langle f(z), x(z) \rangle = f_0 x_0.$$

A Propriedade 2.3 junto com o fato de $f_k = 0, k < 0$ e $x_k = 0, k > 0$ levam a este resultado. Em particular, se $f_0 = 0$ ou $x_0 = 0$, $\langle f(z), x(z) \rangle = 0$.

2.3 Tópicos de sistemas lineares

Considerando que a função

$$\hat{H}(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1 z + \cdots + b_M z^M}{1 + a_1 z + \cdots + a_M z^M}$$

pertence ao espaço $\mathcal{H}_2$, podemos tratá-la como sendo a função de transferência de um sistema linear de tempo discreto estável, causal e de ordem $\deg[\hat{H}(z)] \leq M$. Isto nos permitirá nas seções seguintes formular e deduzir propriedades do problema de aproximação racional em um contexto mais concreto e mais próximo do problema principal que iremos abordar, o da filtragem adaptativa IIR.

Começaremos apresentando a descrição deste sistema em espaço de estados, depois abordaremos seus gramianos de controlabilidade e de observabilidade, o que nos permitirá definir o que é um sistema balanceado e forma de Hankel de sistemas lineares. No final desta seção apresentaremos a relação entre sistemas lineares e o produto interno definido anteriormente, o que nos será útil no resto do trabalho.

2.3.1 Espaço de estados e função de transferência

O sistema $\hat{H}(z)$ pode ser descrito no espaço de estados como

$$\mathbf{x}(n+1) = \mathbf{A}\mathbf{x}(n) + \mathbf{b}u(n) \quad (2.4)$$

$$\hat{y}(n) = \mathbf{c}^t \mathbf{x}(n) + du(n) \quad (2.5)$$

onde $u(n)$ é a entrada do sistema, $\hat{y}(n) = \hat{H}(z)u(n)$ é sua saída e $\mathbf{x}(n)$ é o vetor de estados com M elementos. Reescrevendo a equação (2.4) acima como

$$\begin{aligned} \mathbf{x}(n+1) &= z\mathbf{A}\mathbf{x}(n) + \mathbf{b}u(n) \\ \mathbf{x}(n+1)(\mathbf{I} - z\mathbf{A}) &= \mathbf{b}u(n) \\ \mathbf{x}(n+1) &= (\mathbf{I} - z\mathbf{A})^{-1} \mathbf{b}u(n) \end{aligned} \quad (2.6)$$

e a equação (2.5) como

$$\hat{y}(n) = z\mathbf{c}^t \mathbf{x}(n+1) + du(n),$$

que, substituindo o valor de (2.6), se transforma em

$$\begin{aligned} \hat{y}(n) &= z\mathbf{c}^t (\mathbf{I} - z\mathbf{A})^{-1} \mathbf{b}u(n) + du(n) \\ \hat{H}(z)u(n) &= (d + z\mathbf{c}^t (\mathbf{I} - z\mathbf{A})^{-1} \mathbf{b}) u(n), \end{aligned}$$

concluimos que a função de transferência de $\hat{H}(z)$ será

$$\hat{H}(z) = d + z\mathbf{c}^t (\mathbf{I} - z\mathbf{A})^{-1} \mathbf{b} = \sum_{k=0}^{\infty} \hat{h}_k z^k. \quad (2.7)$$

Este desenvolvimento mostra a facilidade de se trabalhar com a notação mista adotada.

2.3.2 Gramianos de controlabilidade e de observabilidade

Define-se a matriz de controlabilidade de horizonte infinito como

$$\mathcal{C} \triangleq \begin{bmatrix} \mathbf{b} & \mathbf{A}\mathbf{b} & \mathbf{A}^2\mathbf{b} & \dots \end{bmatrix} \quad (2.8)$$

e o par $(\mathbf{A}, \mathbf{b})$ é dito completamente controlável se $\mathcal{C}$ tiver posto M . O gramiano $\mathbf{K}$ desta matriz é dado por

$$\mathbf{K} = \mathcal{C}\mathcal{C}^t = \sum_{k=0}^{\infty} \mathbf{A}^k \mathbf{b} (\mathbf{A}^k \mathbf{b})^t, \quad (2.9)$$

o qual satisfaz a equação de Lyapunov [5]

$$\begin{aligned}
\mathbf{K} &= \sum_{k=0}^{\infty} \mathbf{A}^k \mathbf{b} (\mathbf{A}^k \mathbf{b})^t \\
&= \sum_{k=0}^{\infty} \mathbf{A}^k \mathbf{b} \mathbf{b}^t (\mathbf{A}^t)^k \\
&= \sum_{k=1}^{\infty} \mathbf{A}^k \mathbf{b} \mathbf{b}^t (\mathbf{A}^t)^k + \mathbf{b} \mathbf{b}^t \\
&= \mathbf{A} \left(\sum_{k=0}^{\infty} \mathbf{A}^k \mathbf{b} \mathbf{b}^t (\mathbf{A}^t)^k \right) \mathbf{A}^t + \mathbf{b} \mathbf{b}^t \\
&= \mathbf{A} \mathbf{K} \mathbf{A}^t + \mathbf{b} \mathbf{b}^t.
\end{aligned}$$

A matriz $\mathbf{K}$ é conhecida como gramiano de controlabilidade.

Por construção, $\mathbf{K} = \mathcal{C}\mathcal{C}^t$ é simétrica e positiva semi-definida. Além disso, se o par $(\mathbf{A}, \mathbf{b})$ for completamente controlável, $\mathbf{K}$ é simétrica e positiva definida e, portanto, seus autovalores são estritamente positivos (ver o Apêndice E, onde também é discutida a computação de $\mathbf{K}$).

Cada linha da matriz (2.8) contém os termos da expansão em série de potências da função de transferência da entrada $u(n)$ para um dos elementos do estado $\mathbf{x}(n+1)$

$$\mathbf{x}(n+1) = \mathcal{C}(z)u(n), \quad (2.10)$$

com $\mathcal{C}(z) = [\mathcal{C}_1(z) \ \mathcal{C}_2(z) \ \dots \ \mathcal{C}_M(z)]^t$ dado por

$$\mathcal{C}(z) \triangleq \sum_{k=0}^{\infty} \mathbf{A}^k \mathbf{b} z^k = (\mathbf{I} - z\mathbf{A})^{-1} \mathbf{b}.$$

Este vetor é denominado *vetor de controlabilidade de $(\mathbf{A}, \mathbf{b})$* e seus elementos $\mathcal{C}_{k+1}(z)$ são da forma

$$\mathcal{C}_{k+1}(z) = P_k(z)/A(z), \ k = 0, \dots, M-1, \quad (2.11)$$

onde as funções $P_k(z)$ têm $\deg[P_k(z)] < M$.

De maneira análoga à matriz (2.8), define-se a matriz de observabilidade de

horizonte infinito como

$$\mathcal{O} \triangleq \begin{bmatrix} \mathbf{c}^t \\ \mathbf{c}^t \mathbf{A} \\ \mathbf{c}^t \mathbf{A}^2 \\ \vdots \end{bmatrix} \quad (2.12)$$

cujo gramiano é

$$\mathbf{W} = \mathcal{O}^t \mathcal{O} = \sum_{k=0}^{\infty} (\mathbf{c}^t \mathbf{A}^k)^t \mathbf{c}^t \mathbf{A}^k \quad (2.13)$$

e que satisfaz a equação de Lyapunov

$$\mathbf{W} = \mathbf{A}^t \mathbf{W} \mathbf{A} + \mathbf{c} \mathbf{c}^t.$$

O par $(\mathbf{A}, \mathbf{c})$ é dito completamente observável se $\mathcal{O}$ tiver posto M . A matriz $\mathbf{W}$ é conhecida como o gramiano de observabilidade. Valem propriedades de positividade análogas às do gramiano $\mathbf{K}$.

Os termos $(\mathbf{A}, \mathbf{b}, \mathbf{c}, d)$, bem como os gramianos $\mathbf{K}$ e $\mathbf{W}$, dependem da realização específica de $\hat{H}(z)$. Quando esta realização é completamente controlável e observável, dizemos que ela é mínima. Todas as realizações mínimas de $\hat{H}(z)$ estão relacionadas através de uma matriz inversível $\mathbf{T}$ de dimensões $M \times M$. De fato, seja o vetor de estados $\mathbf{z}(n) = \mathbf{T} \mathbf{x}(n)$ para qualquer n , o que nos leva a descrição em espaço de estados

$$\begin{aligned} \mathbf{z}(n) &= [\mathbf{T} \mathbf{A} \mathbf{T}^{-1}] \mathbf{z}(n) + [\mathbf{T} \mathbf{b}] u(n) \\ \hat{y}(n) &= [\mathbf{c}^t \mathbf{T}^{-1}] \mathbf{z}(n) + d u(n). \end{aligned}$$

A resposta impulsiva $\hat{h}_k$, $k \geq 1$, deste sistema será

$$\hat{h}_k = [\mathbf{c}^t \mathbf{T}^{-1}] [\mathbf{T} \mathbf{A} \mathbf{T}^{-1}]^{k-1} [\mathbf{T} \mathbf{b}] = \mathbf{c}^t \mathbf{A}^{k-1} \mathbf{b},$$

a qual é a mesma que a da equação (2.7). Portanto $(\mathbf{A}' = \mathbf{T} \mathbf{A} \mathbf{T}^{-1}, \mathbf{b}' = \mathbf{T} \mathbf{b}, \mathbf{c}'^t = \mathbf{c}^t \mathbf{T}^{-1}, d)$ é uma outra realização de $\hat{H}(z)$. Nesta nova realização os gramianos de controlabilidade e de observabilidade são dados, respectivamente, por

$$\mathbf{K}' = \mathbf{T} \mathbf{K} \mathbf{T}^t \quad (2.14)$$

$$\mathbf{W}' = \mathbf{T}^{-t} \mathbf{W} \mathbf{T}^{-1}. \quad (2.15)$$

2.3.3 Valores singulares de Hankel e realização balanceada

A forma de Hankel associada a $\hat{H}(z) = \hat{h}_0 + \hat{h}_1 z + \hat{h}_2 z^2 + \dots$ é dada por

$$\Gamma_{\hat{H}} = \begin{bmatrix} \hat{h}_1 & \hat{h}_2 & \hat{h}_3 & \cdots \\ \hat{h}_2 & \hat{h}_3 & \hat{h}_4 & \cdots \\ \hat{h}_3 & \hat{h}_4 & \hat{h}_5 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \quad (2.16)$$

onde notamos a exclusão de $\hat{h}_0$. Utilizando as equações (2.7), (2.9) e (2.13), podemos escrever

$$\begin{aligned} \Gamma_{\hat{H}} &= \begin{bmatrix} \mathbf{c}^t \\ \mathbf{c}^t \mathbf{A} \\ \mathbf{c}^t \mathbf{A}^2 \\ \vdots \end{bmatrix} \begin{bmatrix} \mathbf{b} & \mathbf{A}\mathbf{b} & \mathbf{A}^2\mathbf{b} & \cdots \end{bmatrix} \\ &= \mathcal{O}\mathcal{C}. \end{aligned} \quad (2.17)$$

Temos então

$$\Gamma_{\hat{H}}^t \Gamma_{\hat{H}} = \mathcal{C}^t \mathcal{O}^t \mathcal{O} \mathcal{C},$$

e permutando $\mathcal{C}^t \mathcal{O}^t \mathcal{O}$ com $\mathcal{C}$ temos

$$\mathcal{C} \mathcal{C}^t \mathcal{O}^t \mathcal{O} = \mathbf{K} \mathbf{W},$$

sendo que a permutação mantém a igualdade entre os autovalores não-nulos de $\Gamma_{\hat{H}}^t \Gamma_{\hat{H}}$ e $\mathbf{K} \mathbf{W}$ (ver o Apêndice E).

Consideremos uma realização mínima de $\hat{H}(z)$, ou seja, uma realização para a qual os gramianos $\mathbf{K}$ e $\mathbf{W}$ são matrizes positivas definidas. Notamos que neste caso os M autovalores do produto $\mathbf{K} \mathbf{W}$ são estritamente positivos (ver o Apêndice E). Definem-se os *valores singulares de Hankel* σ_k de $\hat{H}(z)$ como as raízes quadradas destes autovalores e convencionam-se ordená-los como

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_M > 0.$$

Notamos que os valores singulares de Hankel não dependem da realização mínima particular sendo considerada: de (2.14) e (2.15) temos que

$$\mathbf{K}' \mathbf{W}' = \mathbf{T} \mathbf{K} \mathbf{T}^t \mathbf{T}^{-t} \mathbf{W} \mathbf{T}^{-1} = \mathbf{T} \mathbf{K} \mathbf{W} \mathbf{T}^{-1},$$

ou seja, o produto $\mathbf{KW}$ sofre uma transformação de similaridade, o que não altera seus autovalores. Conforme discutido no Apêndice E, existe uma realização mínima denominada de *realização balanceada*, para a qual os gramianos de controlabilidade e observabilidade são diagonais e compostos pelos valores singulares de Hankel de $\hat{H}(z)$:

$$\mathbf{K}_b = \mathbf{W}_b = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_M \end{bmatrix} \triangleq \mathbf{\Sigma}. \quad (2.18)$$

Os valores singulares de Hankel do sistema sendo identificado $H(z)$ foram utilizados em [23] para analisar a velocidade de convergência de filtros adaptativos IIR, conforme discutido adiante na Seção 2.8.

2.3.4 Sistemas com entrada estocástica

Seja um sinal $u(n)$ cuja função densidade espectral é dada por

$$S_u(z) = \sum_{k=-\infty}^{\infty} E[u(n)u(n+k)] z^k. \quad (2.19)$$

Se aplicarmos este sinal na entrada dos sistemas lineares $X(z) = \sum_{k=0}^{\infty} x_k z^k$ e $Y(z) = \sum_{k=0}^{\infty} y_k z^k$ obteremos os sinais $\tilde{x}(n) = X(z)u(n)$ e $\tilde{y}(n) = Y(z)u(n)$. A correlação entre estes sinais será

$$\begin{aligned} E[\tilde{x}(n)\tilde{y}(n)] &= E[X(z)u(n)Y(z)u(n)] \\ &= E\left[\sum_{k=0}^{\infty} x_k u(n-k) \sum_{l=0}^{\infty} y_l u(n-l)\right] \\ &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} x_k y_l E[u(n-k)u(n-l)] \\ &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} x_k y_l E[u(m)u(m+k-l)], \quad m = n - k \\ &= \sum_{k=0}^{\infty} x_k \sum_{p=k}^{\infty} y_{p-k} E[u(m)u(m+p)], \quad p = k - l \end{aligned}$$

a segunda somatória contém o k -ésimo termo da expansão em série de $S_u(z)Y(z)$, nos levando, com a Propriedade 2.3, a

$$\begin{aligned} E[\tilde{x}(n)\tilde{y}(n)] &= \langle X(z), S_u(z)Y(z) \rangle \\ &= \langle X(z)S_u(z), Y(z) \rangle. \end{aligned} \quad (2.20)$$

Isto significa que a correlação entre as respostas de dois sistemas excitados pela mesma entrada $u(n)$ será igual ao produto interno de suas funções de transferência $X(z)$ e $Y(z)$ ponderado pela função densidade espectral $S_u(z)$ da entrada. Em particular, se a entrada for branca teremos

$$E[\tilde{x}(n)\tilde{y}(n)] = \langle X(z), Y(z) \rangle. \quad (2.21)$$

2.4 Teoria de Beurling-Lax

A teoria de Beurling-Lax trata de subespaços de $\mathcal{H}_2$ invariantes a deslocamentos e suas relações com funções passa-tudo. Um subespaço $\mathcal{S}_r$ de $\mathcal{H}_2$ invariante ao deslocamento para a direita é aquele em que dada uma função $f(z) \in \mathcal{S}_r$ temos que $zf(z) \in \mathcal{S}_r$. Analogamente, um subespaço $\mathcal{S}_l$ de $\mathcal{H}_2$ invariante ao deslocamento para a esquerda é aquele em que para qualquer $f(z) \in \mathcal{S}_l$, temos $[z^{-1}f(z)]_{\oplus} \in \mathcal{S}_l$. Mais detalhes quanto a estes espaços podem ser encontrados em [5, cap. 3].

Iniciaremos apresentando o chamado Teorema de Beurling-Lax em uma forma ligada à teoria de sistemas lineares abordada anteriormente. Em seguida, este resultado será utilizado para obter uma base para o espaço $\mathcal{H}_2$, a qual será utilizada em análises posteriores.

2.4.1 Subespaço ortogonal a vetor de controlabilidade

A seguir demonstraremos o resultado que nos ajudará no restante deste trabalho, um teorema que relaciona um vetor de controlabilidade $\mathcal{C}(z)$ com um conjunto de funções formadas a partir do deslocamento de uma função passa-tudo. No

que segue consideraremos que $\mathcal{C}(z)$ é associado a um par $(\mathbf{A}, \mathbf{b})$ completamente controlável. O seguinte lema será útil na demonstração:

Lema 2.5 *Seja $V(z)$ a função passa-tudo cujos pólos são os mesmos das funções $\mathcal{C}_k(z) = P_k(z)/A(z)$ que compõem o vetor de controlabilidade $\mathcal{C}(z)$. Então*

$$\langle \mathcal{C}(z), z^k V(z) \rangle = \mathbf{0}, \quad k \geq 0. \quad (2.22)$$

Nesta notação o produto interno entre o vetor $\mathcal{C}(z)$ e a função $z^k V(z)$ é um vetor que tem como elementos o produto interno entre cada componente de $\mathcal{C}(z)$ e $z^k V(z)$, ou seja

$$\begin{aligned} \langle \mathcal{C}(z), z^k V(z) \rangle &= \left\langle \begin{bmatrix} \mathcal{C}_1(z) \\ \mathcal{C}_2(z) \\ \vdots \\ \mathcal{C}_M(z) \end{bmatrix}, z^k V(z) \right\rangle \\ &= \begin{bmatrix} \langle \mathcal{C}_1(z), z^k V(z) \rangle \\ \langle \mathcal{C}_2(z), z^k V(z) \rangle \\ \vdots \\ \langle \mathcal{C}_M(z), z^k V(z) \rangle \end{bmatrix}. \end{aligned}$$

Podemos ver que (2.22) é verdadeira reescrevendo-a como

$$\begin{aligned} \langle \mathcal{C}(z), z^k V(z) \rangle &= \langle \mathcal{C}(z) V(z^{-1}), z^k \rangle \\ &= \left\langle \begin{bmatrix} \mathcal{C}_1(z) \\ \vdots \\ \mathcal{C}_M(z) \end{bmatrix}, V(z^{-1}), z^k \right\rangle \\ &= \left\langle \begin{bmatrix} P_0(z) \\ \vdots \\ P_{M-1}(z) \end{bmatrix}, \frac{1}{A(z)} \frac{z^{-M} A(z)}{A(z^{-1})}, z^k \right\rangle \\ &= \left\langle \begin{bmatrix} P_0(z) \\ \vdots \\ P_{M-1}(z) \end{bmatrix}, \frac{z^{-M}}{A(z^{-1})}, z^k \right\rangle \\ &= \mathbf{0} \end{aligned}$$

sendo que a última igualdade decorre do fato de z^k ser causal para $k \geq 0$ e do fato de que os polinômios $z^{-M}P_l(z)$ são estritamente anticausais (pois $\deg[P_l(z)] < M$).

Teorema 2.6 (Teorema de Beurling-Lax) *Se $\mathcal{C}(z)$ é um vetor de controlabilidade e $f(z) \in \mathcal{H}_2$, então*

$$\langle \mathcal{C}(z), f(z) \rangle = 0 \quad (2.23)$$

se e somente se

$$f(z) = g(z)V(z), \quad g(z) \in \mathcal{H}_2 \quad (2.24)$$

onde

$$V(z) = \frac{z^M A(z^{-1})}{A(z)} = \frac{\overline{A}(z)}{A(z)} \quad (2.25)$$

é uma função passa-tudo chamada de complemento passa-tudo de $\mathcal{C}(z)$.

Para mostrar que a condição (2.24) é suficiente, suponha que

$$f(z) = V(z) \sum_{k=0}^{\infty} g_k z^k, \quad \sum_{k=0}^{\infty} g_k^2 < \infty.$$

Então

$$\begin{aligned} \langle \mathcal{C}(z), f(z) \rangle &= \left\langle \mathcal{C}(z), V(z) \sum_{k=0}^{\infty} g_k z^k \right\rangle \\ &= \sum_{k=0}^{\infty} g_k \langle \mathcal{C}(z), z^k V(z) \rangle \end{aligned}$$

onde, pelo Lema 2.5, esta somatória é igual a zero.

Para mostrar a necessidade da condição (2.24), considere agora que a equação (2.23) é válida e que $f(z) \in \mathcal{H}_2$. Interpretando $f(z)$ como a função de transferência de um sistema estável e causal cuja entrada é a sequência $\{u(\cdot)\}$, temos:

$$y(n) = f(z)u(n) = \sum_{k=0}^{\infty} f_k u(n-k).$$

Caso $\{u(\cdot)\}$ seja uma sequência branca de variância unitária,

$$f_k = E[y(n)u(n-k)] \quad (2.26)$$

e se esta mesma sequência for aplicada à entrada de um sistema $Q(z)$ que tenha como vetor de controlabilidade $\mathcal{C}(z)$

$$\begin{bmatrix} \mathbf{x}(n+1) \\ q(n) \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{c}^t & d \end{bmatrix}}_{\mathbf{Q}} \begin{bmatrix} \mathbf{x}(n) \\ u(n) \end{bmatrix}, \quad (2.27)$$

podemos escrever, usando (2.10) e (2.20),

$$\begin{aligned} E[\mathbf{x}(n+1)y(n)] &= E[\mathcal{C}(z)u(n)f(z)u(n)] \\ &= \langle \mathcal{C}(z), f(z) \rangle = \mathbf{0}. \end{aligned}$$

Supondo que o sistema em (2.27) representa um filtro ortogonal, isto é, $\mathbf{Q}^t \mathbf{Q} = \mathbf{Q} \mathbf{Q}^t = \mathbf{I}$, podemos calcular o sistema adjunto

$$\begin{bmatrix} \mathbf{x}(n) \\ u(n) \end{bmatrix} = \begin{bmatrix} \mathbf{A}^t & \mathbf{c} \\ \mathbf{b}^t & d \end{bmatrix} \begin{bmatrix} \mathbf{x}(n+1) \\ q(n) \end{bmatrix} \quad (2.28)$$

que agora regride no tempo. As entradas passadas podem então ser reconstruídas iterando a recursão acima para $n-1, n-2, \dots$, obtendo-se

$$\begin{bmatrix} u(n) \\ u(n-1) \\ \vdots \\ u(n-k) \end{bmatrix} = \begin{bmatrix} \mathbf{b}^t \\ \mathbf{b}^t \mathbf{A}^t \\ \vdots \\ \mathbf{b}^t (\mathbf{A}^k)^t \end{bmatrix} \mathbf{x}(n+1) + \begin{bmatrix} q_0 & 0 & \cdots & 0 \\ q_1 & q_0 & \ddots & \vdots \\ \vdots & \ddots & q_0 & 0 \\ q_k & \cdots & q_1 & q_0 \end{bmatrix} \begin{bmatrix} q(n) \\ q(n-1) \\ \vdots \\ q(n-k) \end{bmatrix} \quad (2.29)$$

onde

$$q_l = \begin{cases} d, & l = 0 \\ (\mathbf{c}^t \mathbf{A}^{l-1} \mathbf{b})^t, & l = 1, 2, \dots \end{cases}$$

é a resposta impulsiva de $Q(z)$.

Substituindo (2.29) em (2.26) e usando o fato de que $E[\mathbf{x}(n+1)y(n)] = \mathbf{0}$, temos

$$\begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_k \end{bmatrix} = \begin{bmatrix} E[y(n)u(n)] \\ E[y(n)u(n-1)] \\ \vdots \\ E[y(n)u(n-k)] \end{bmatrix} = \begin{bmatrix} q_0 & 0 & \cdots & 0 \\ q_1 & q_0 & \ddots & \vdots \\ \vdots & \ddots & q_0 & 0 \\ q_k & \cdots & q_1 & q_0 \end{bmatrix} \begin{bmatrix} E[y(n)q(n)] \\ E[y(n)q(n-1)] \\ \vdots \\ E[y(n)q(n-k)] \end{bmatrix} \quad (2.30)$$

o que mostra que a sequência $\{f_k\}$ é obtida a partir de uma operação de convolução. Desta maneira

$$\begin{aligned} f(z) &= \sum_{k=0}^{\infty} f_k z^k \\ &= \left(\sum_{k=0}^{\infty} v_k z^k \right) \left(\sum_{k=0}^{\infty} \underbrace{E[y(n)q(n-k)]}_{g_k} z^k \right) \\ &= Q(z)g(z). \end{aligned}$$

Como a equação (2.28) só é válida se a matriz $\mathbf{Q}$ em (2.27) for ortogonal, o que só é possível se $Q(z)$ for passa-tudo [5, seção 2.4.3], então chegamos a

$$Q(z) = \frac{z^M A(z^{-1})}{A(z)} = V(z)$$

e

$$f(z) = V(z)g(z).$$

Para completar a prova falta provar que $g(z) \in \mathcal{H}_2$. Por definição

$$\begin{aligned} g_k &= E[y(n)q(n-k)] \\ &= \langle f(z), z^k V(z) \rangle \\ &= \langle V(z^{-1})f(z), z^k \rangle. \end{aligned}$$

Usando a desigualdade de Schwarz e a igualdade de Parseval, temos

$$\sum_{k=0}^{\infty} g_k^2 \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |V(e^{j\omega})|^2 |f(z)|^2 d\omega = \|f(z)\|^2 < \infty.$$

Então $g(z)$ possui uma norma L_2 finita e é causal, logo $g(z) \in \mathcal{H}_2$ e o teorema está provado.

2.4.2 Base para $\mathcal{H}_2$

A equação (2.11) nos mostra que todos os componentes de $\mathcal{C}(z)$ possuem o mesmo denominador $A(z)$, e portanto os mesmos pólos, e como os polinômios $P_k(z)$ têm grau menor do que M (ordem do polinômio $A(z)$), resulta que podemos escrever qualquer função do tipo

$$Q(z) = \frac{q_0 + q_1 z + \cdots + q_{M-1} z^{M-1}}{A(z)}$$

combinando os elementos do vetor de controlabilidade $\mathcal{C}(z)$. Reunindo todas as funções $Q(z)$ possíveis de serem escritas desta maneira em um subespaço $\mathcal{Q}$ de $\mathcal{H}_2$, vemos que este é gerado pelos elementos $P_k(z)/A(z)$ de $\mathcal{C}(z)$. Pelo Teorema 2.6 temos que o espaço ortogonal a $\mathcal{Q}$ pode ser escrito como $\mathcal{Q}^\perp = V(z)\mathcal{H}_2$, com $V(z)$ dado por (2.25). Invertendo-se este raciocínio, temos que $\mathcal{Q} = \mathcal{H}_2 \ominus V(z)\mathcal{H}_2$, o que nos permite enunciar:

Propriedade 2.7 *Dado um par $(\mathbf{A}, \mathbf{b})$ completamente controlável, as M funções de seu vetor de controlabilidade $\mathcal{C}(z)$ e o conjunto $\{z^k V(z)\}_{k=0}^\infty$, onde $V(z)$ é o complemento passa-tudo de $\mathcal{C}(z)$, formam uma base completa para o espaço $\mathcal{H}_2$. Além disto, se tivermos uma realização onde*

$$\langle \mathcal{C}(z), \mathcal{C}(z) \rangle = \mathbf{I},$$

esta base será também ortonormal.

Vale ressaltar que esta base é determinada pelo denominador $A(z)$ das funções que compõem $\mathcal{C}(z)$, pois tanto os polinômios $P_k(z)$ de (2.11) quanto a função $V(z)$ dependem exclusivamente dele.

Consideremos agora $F(z) = f_0 + \dots + f_M z^M \in \mathcal{H}_2$ e o produto interno

$$\left\langle \frac{F(z)}{A(z)}, \mathcal{C}(z) \right\rangle = \begin{bmatrix} \psi_0 \\ \psi_1 \\ \vdots \\ \psi_{M-1} \end{bmatrix}.$$

Usando o Lema 2.5 vemos que os coeficientes ψ_k só serão todos nulos se $F(z)/A(z) = V(z)$ e que para qualquer outro caso pelo menos um deles será diferente de zero.

Observemos por outro lado o que acontece com o produto interno entre $F(z)/A(z)$ e as funções $\{z^k V(z)\}_{k=0}^\infty$:

$$\begin{aligned} \left\langle \frac{F(z)}{A(z)}, z^k \frac{\overline{A}(z)}{A(z)} \right\rangle &= \left\langle \frac{F(z)}{A(z)}, z^k \frac{z^M A(z^{-1})}{A(z)} \right\rangle \\ &= \left\langle \frac{F(z) z^{-M} A(z)}{A(z)}, \frac{z^k}{A(z)} \right\rangle \\ &= \left\langle F(z) z^{-M}, \frac{z^k}{A(z)} \right\rangle \\ &= \begin{cases} f_M, & k = 0 \\ 0, & k > 0. \end{cases} \end{aligned}$$

Isto nos mostra que somente o produto interno $\psi_M = \langle F(z)/A(z), V(z) \rangle = f_M$ é diferente de zero, o que sugere separar $V(z)$ das demais funções do conjunto $\{z^k V(z)\}_{k=0}^\infty$.

Desta maneira, utilizando a equação (2.11), podemos escrever a base da Propriedade 2.7 como

$$\left\{ \frac{P_k(z)}{A(z)} \right\}_{k=0}^M \cup \{z^k V(z)\}_{k=1}^\infty, \quad (2.31)$$

com $P_M(z) = z^M A(z^{-1}) = \overline{A}(z)$. Notamos que qualquer função com grau menor ou igual a M que tenha como pólos os zeros de $A(z)$ pode ser escrita utilizando apenas o primeiro conjunto de funções,

$$\frac{F(z)}{A(z)} = \sum_{k=0}^M \psi_k \frac{P_k(z)}{A(z)}.$$

Funções com outros pólos precisam, em princípio, de todas as funções da base.

2.5 Aproximação racional $\mathcal{H}_2$

Como mencionado anteriormente, o problema de aproximação racional tem uma forte ligação com o problema de filtragem adaptativa IIR. Nesta seção serão abordados alguns dos principais aspectos da aproximação racional de funções que pertencem ao espaço $\mathcal{H}_2$, sendo que estes resultados nos auxiliarão nas análises efetuadas nos capítulos seguintes.

Seja

$$H(z) = \sum_{k=0}^{\infty} h_k z^k \in \mathcal{H}_2$$

uma função causal e estável desconhecida. Na aproximação racional $\mathcal{H}_2$ queremos encontrar uma função racional

$$\hat{H}(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1 z + \cdots + b_M z^M}{1 + a_1 z + \cdots + a_M z^M}, \quad (2.32)$$

com $\deg[\hat{H}(z)] \leq M$, de modo a minimizar o erro quadrático

$$\|H(z) - \hat{H}(z)\|^2 = \langle H(z) - \hat{H}(z), H(z) - \hat{H}(z) \rangle. \quad (2.33)$$

A escolha desta função custo é interessante pois é composta por um conjunto de funções diferenciáveis pertencentes a um espaço com produto interno. Isto nos permite usar o formalismo dos espaços de Hilbert para caracterizar soluções através de restrições de ortogonalidade. Apesar disto, este problema não possui sempre uma solução analítica fechada e métodos numéricos geralmente são a única maneira efetiva de se encontrar uma solução. Alguns destes métodos numéricos, como os que serão abordados neste trabalho, utilizam o gradiente da função custo para adaptar os parâmetros da função $\hat{H}(z)$.

2.5.1 Decomposição do erro quadrático

Utilizaremos agora a base (2.31) para obter uma decomposição do erro quadrático que nos permita analisar melhor o que ocorre quando variamos os zeros e os pólos de $\hat{H}(z)$.

Consideremos a base formada a partir dos pólos de $\hat{H}(z)$. Podemos escrever então:

$$\hat{H}(z) = \sum_{k=0}^M \nu_k \frac{P_k(z)}{A(z)} \quad (2.34)$$

$$H(z) = \sum_{k=0}^M \rho_k \frac{P_k(z)}{A(z)} + \sum_{k=1}^{\infty} \tau_k z^k V(z), \quad (2.35)$$

onde

$$\nu_k = \left\langle \hat{H}(z), \frac{P_k(z)}{A(z)} \right\rangle \quad (2.36)$$

$$\rho_k = \left\langle H(z), \frac{P_k(z)}{A(z)} \right\rangle \quad (2.37)$$

$$\tau_k = \langle H(z), z^k V(z) \rangle. \quad (2.38)$$

Admitindo que a base da Propriedade 2.7 é ortonormal, e utilizando (2.34) e (2.35), podemos escrever o erro quadrático médio da aproximação racional $\mathcal{H}_2$ como [5]

$$\begin{aligned} \|H(z) - \hat{H}(z)\|^2 &= \langle H(z) - \hat{H}(z), H(z) - \hat{H}(z) \rangle \\ \|H(z) - \hat{H}(z)\|^2 &= \sum_{k=0}^M (\rho_k - \nu_k)^2 + \sum_{k=1}^{\infty} \tau_k^2. \end{aligned} \quad (2.39)$$

Esta forma de decompor o erro será útil para analisarmos os pontos estacionários da superfície de erro e definirmos uma superfície de erro reduzida que depende somente dos pólos de $\hat{H}(z)$.

2.5.2 Superfície de erro reduzida

Vemos nas equações (2.36) a (2.38) que os coeficientes ν_k dependem tanto dos pólos quanto dos zeros de $\hat{H}(z)$, e que ρ_k e τ_k dependem apenas dos pólos. Isto significa que a representação de $H(z)$ depende somente dos pólos de $\hat{H}(z)$. Mantendo fixos os pólos de $\hat{H}(z)$ e variando apenas seus zeros, podemos minimizar o erro quadrático $\|H(z) - \hat{H}(z)\|^2$. Nestas condições, o erro mínimo será atingido quando $\hat{H}(z)$ for igual à projeção de $H(z)$ no espaço gerado pelo primeiro conjunto de funções da base em (2.31). Isto é equivalente a termos

$$\rho_k = \nu_k, \quad k = 0, 1, \dots, M \quad (2.40)$$

nas equações (2.34) e (2.35). Com isto os zeros de $\hat{H}(z)$ estão otimizados dado seus pólos e podemos escrever

$$H(z) - \hat{H}(z) = \sum_{k=1}^{\infty} \tau_k z^k V(z).$$

Definindo $g(z) = \sum_{k=1}^{\infty} \tau_k z^k$, temos a seguinte propriedade:

Propriedade 2.8 Quando os pólos de $\hat{H}(z)$ são fixados e os seus zeros são escolhidos de modo a minimizar $\|H(z) - \hat{H}(z)\|^2$, o erro $H(z) - \hat{H}(z)$ tem a forma:

$$H(z) - \hat{H}(z) = g(z)V(z), \quad z^{-1}g(z) \in \mathcal{H}_2 \quad (2.41)$$

para algum $g(z)$ e com $V(z)$ dado por (2.25).

Notar que $z^{-1}g(z) \in \mathcal{H}_2$ significa que $g(z)$ é estritamente causal.

Quando a Propriedade 2.8 é válida, dizemos que o erro se encontra na *superfície de erro reduzida*, e a equação (2.39) se reduz a

$$\|H(z) - \hat{H}(z)\|^2 = \|g(z)\|^2 = \sum_{k=1}^{\infty} \tau_k^2, \quad (2.42)$$

onde a primeira igualdade vem do fato de $V(z)$ ser passa-tudo de norma unitária. Neste caso o erro quadrático pode ser escrito em termos da forma de Hankel de $H(z)$, como mostrado a seguir.

Seguindo a Propriedade 2.3, os coeficientes τ_k podem ser escritos como

$$\tau_k = \langle H(z), z^k V(z) \rangle = \sum_{l=k}^{\infty} h_l v_{l-k}, \quad k \geq 1$$

onde h_l são os coeficientes da expansão em série de $H(z) = \sum_{l=0}^{\infty} h_l z^l$ e v_l , os da expansão de $V(z) = \sum_{l=0}^{\infty} v_l z^l$. Em forma matricial temos [5]

$$\mathbf{g} = \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \vdots \end{bmatrix} = \underbrace{\begin{bmatrix} h_1 & h_2 & h_3 & \cdots \\ h_2 & h_3 & h_4 & \cdots \\ h_3 & h_4 & h_5 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}}_{\Gamma_H} \begin{bmatrix} v_0 \\ v_1 \\ v_2 \\ \vdots \end{bmatrix} = \Gamma_H \mathbf{v}, \quad (2.43)$$

com Γ_H sendo a forma de Hankel de $H(z)$ e os elementos do vetor $\mathbf{v}$ sendo os coeficientes v_l . Com isto, e levando em conta que Γ_H é simétrica, a equação (2.42) pode ser reescrita como

$$\|H(z) - \hat{H}(z)\|^2 = \mathbf{g}^t \mathbf{g} = \mathbf{v}^t \Gamma_H^2 \mathbf{v}. \quad (2.44)$$

Esta equação será útil para definir uma parametrização dos pólos baseada na teoria de realização balanceada, conforme será visto na Seção 2.8.1.

2.5.3 Pontos estacionários da superfície de erro quadrático

Através da equação (2.32) vemos que a função $\hat{H}(z)$ possui $2M + 1$ parâmetros a serem ajustados, $M + 1$ no numerador e M no denominador. Este número de parâmetros independe da realização de $\hat{H}(z)$, e os parâmetros serão agrupados em um vetor $\mathbf{w} = [w_0 \ w_1 \ \dots \ w_{2M}]^t$ que os representa de maneira genérica. Define-se um ponto estacionário da superfície de erro quadrático como:

Propriedade 2.9 $\hat{H}(z)$ é um ponto estacionário de $\|H(z) - \hat{H}(z)\|^2$ se para qualquer w_j do conjunto de parâmetros de $\hat{H}(z)$ valer

$$\frac{\partial}{\partial w_j} \|H(z) - \hat{H}(z)\|^2 = 0. \quad (2.45)$$

Tomando como exemplo a realização na forma direta, temos

$$\mathbf{w} = \begin{bmatrix} b_o \\ \vdots \\ b_M \\ a_1 \\ \vdots \\ a_M \end{bmatrix},$$

com as derivadas da Propriedade 2.9 iguais a

$$\frac{\partial}{\partial \mathbf{b}} \|H(z) - \hat{H}(z)\|^2 = -2 \left\langle \begin{bmatrix} 1/A(z) \\ z/A(z) \\ \vdots \\ z^M/A(z) \end{bmatrix}, H(z) - \hat{H}(z) \right\rangle \quad (2.46)$$

$$\frac{\partial}{\partial \mathbf{a}} \|H(z) - \hat{H}(z)\|^2 = -2 \left\langle \begin{bmatrix} z/A(z) \\ z^2/A(z) \\ \vdots \\ z^M/A(z) \end{bmatrix}, \hat{H}(z), H(z) - \hat{H}(z) \right\rangle. \quad (2.47)$$

Considerando que na equação (2.46) o vetor

$$\mathcal{C}_a = \begin{bmatrix} 1/A(z) \\ z/A(z) \\ \vdots \\ z^M/A(z) \end{bmatrix}$$

equivale a um vetor de controlabilidade aumentado, temos que seu complemento passa-tudo será igual a $zV(z)$, com $V(z) = z^M A(z^{-1})/A(z)$. Igualando (2.46) a zero e usando o Teorema 2.6,

$$\frac{\partial}{\partial \mathbf{b}} \left\| H(z) - \hat{H}(z) \right\|^2 = \mathbf{0} \Leftrightarrow H(z) - \hat{H}(z) = g(z)V(z), \quad z^{-1}g(z) \in \mathcal{H}_2,$$

que já era esperado dada a Propriedade 2.8. Isto mostra que os pontos estacionários da superfície de erro quadrático pertencem à superfície de erro reduzida.

Agora, a equação (2.47) só será igual a zero se

$$\left\langle \begin{bmatrix} z/A(z) \\ z^2/A(z) \\ \vdots \\ z^M/A(z) \end{bmatrix} \hat{H}(z), H(z) - \hat{H}(z) \right\rangle = \mathbf{0}.$$

Usando o resultado acima e alguma manipulação algébrica, isto equivale a

$$\left\langle \begin{bmatrix} 1/A(z) \\ z/A(z) \\ \vdots \\ z^{M-1}/A(z) \end{bmatrix}, z^{-1}g(z)V(z)\hat{H}(z^{-1}) \right\rangle = \mathbf{0},$$

onde vemos claramente os elementos $\mathcal{C}_k(z)$ do vetor de controlabilidade definidos em (2.11). Usando novamente o Teorema 2.6, temos que em um ponto estacionário da superfície de erro

$$\begin{aligned} z^{-1}g(z)V(z)\hat{H}(z^{-1}) &= q(z)V(z), \quad q(z) \in \mathcal{H}(z) \\ z^{-1}g(z)\frac{z^M A(z^{-1})}{A(z)}\frac{B(z^{-1})}{A(z^{-1})} &= q(z)V(z) \\ z^{-1}g(z)\frac{z^M B(z^{-1})}{A(z)} &= q(z)V(z). \end{aligned}$$

Agora, se $z^M B(z^{-1})/A(z)$ tiver zeros em comum com $V(z)$, teremos cancelamento de pólos e zeros em $\hat{H}(z)$. Conforme discutido em [5], se este cancelamento ocorrer, $\hat{H}(z)$ será um ponto de sela. Agora, se $z^M B(z^{-1})/A(z)$ não tiver zeros em comum com $V(z)$, os zeros de $V(z)$ deverão aparecer em $z^{-1}g(z)$. Desta maneira,

$$\begin{aligned} z^{-1}g(z) &= Q(z)V(z) \\ g(z) &= zQ(z)V(z), \quad Q(z) \in \mathcal{H}_2, \end{aligned}$$

para algum $Q(z)$, nos levando ao seguinte teorema:

Teorema 2.10 (Teorema de Walsh) *Se $\deg[\hat{H}(z)] = M$, então a função racional $\hat{H}(z) = B(z)/A(z)$ será um ponto estacionário de $\|H(z) - \hat{H}(z)\|^2$ se e somente se*

$$H(z) - \hat{H}(z) = zQ(z)V^2(z)$$

para algum $Q(z) \in \mathcal{H}_2$ e $V(z) = z^M A(z^{-1})/A(z)$.

Esta propriedade caracteriza todos os pontos estacionários, incluindo mínimos locais e pontos de sela, para os quais $\deg[\hat{H}(z)] = M$. Máximos locais não existem, pois qualquer ponto estacionário deve pertencer à superfície de erro reduzida e deste modo devem ser um mínimo com relação aos parâmetros dos zeros de $\hat{H}(z)$.

Supomos agora que $\deg[H(z)] = M$, onde M é a ordem da função $\hat{H}(z)$ procurada. O Teorema 2.10 nos diz que em qualquer ponto estacionário onde $\deg[\hat{H}(z)] = M$ temos

$$H(z) - \hat{H}(z) = zQ(z)V^2(z).$$

O lado esquerdo desta equação possui ordem não superior a $2M$, mas o lado direito possui $2M + 1$ zeros, sendo $2M$ zeros de $V^2(z)$ mais um zero em $z = 0$. Desta maneira, $\hat{H}(z)$ interpola $H(z)$ em $2M + 1$ pontos, o que implica em $\hat{H}(z) = H(z)$. Portanto, se $\deg[\hat{H}(z)] = \deg[H(z)]$, a função custo $\|H(z) - \hat{H}(z)\|^2$ tem somente um ponto de mínimo dado por $\hat{H}(z) = H(z)$. Qualquer outro ponto estacionário deve ser um ponto de sela, pois haverá cancelamentos de pólos e zeros em $\hat{H}(z)$.

Figura 2.1: Filtro adaptativo em configuração de identificação.

Em [22] é mostrado que para o caso de sobremodelamento ($\deg[H(z)] < M$), temos $B(z) = B_1(z)L(z)$ e $A(z) = A_1(z)L(z)$ com $\deg[L(z)] \leq M - N$. Variando $L(z)$ teremos infinitos pontos de mínimo globais onde o Teorema 2.10 é válido. Neste caso $H(z) = B_1(z)/A_1(z)$.

No caso de submodelamento ($M < N$), o Teorema 2.10 não pode ser usado para mostrar a inexistência de mínimos locais. Portanto, estes podem eventualmente existir.

2.5.4 Erro quadrático médio

Observemos a configuração de identificação apresentada no capítulo anterior e reproduzida na Figura 2.1. Vemos que tanto o sistema desconhecido $H(z)$ quanto o sistema adaptado $\hat{H}(z)$ são alimentados pela mesma entrada $u(n)$ e que o erro é dado por $e(n) = y(n) - \hat{y}(n)$.

O erro quadrático médio é a variância do erro da saída $e(n)$ dada por

$$\begin{aligned} E[e^2(n)] &= E[(y(n) - \hat{y}(n))^2] \\ &= E\left[\left((H(z) - \hat{H}(z))u(n)\right)^2\right], \end{aligned}$$

que pela equação (2.20) equivale a

$$E[e^2(n)] = \left\langle H(z) - \hat{H}(z), S_u(z) (H(z) - \hat{H}(z)) \right\rangle. \quad (2.48)$$

Se $u(n)$ for branco de variância unitária e média nula, então

$$E[e^2(n)] = \left\langle H(z) - \hat{H}(z), H(z) - \hat{H}(z) \right\rangle \quad (2.49)$$

que é igual a (2.33). Isto justifica a afirmação de que o problema de filtragem adaptativa e o de aproximação racional $\mathcal{H}_2$ são equivalentes.

2.5.5 Extensão para a norma L_2 ponderada

As propriedades discutidas nas seções anteriores dizem respeito à norma L_2 sem ponderação. A seguir será apresentada a extensão destas propriedades quando a norma utilizada é da forma

$$\left\| H(z) - \hat{H}(z) \right\|_{S_u}^2 \triangleq \left\langle H(z) - \hat{H}(z), S_u(z) \left[H(z) - \hat{H}(z) \right] \right\rangle,$$

onde $S_u(z)$ é dado por (2.19). Como mostra a equação (2.48), este tipo de norma aparece no caso de filtragem adaptativa com entrada não-branca ($S_u(z) \neq \sigma_u^2$).

Pode-se mostrar que a Propriedade 2.8 e o Teorema 2.10 podem ser reescritos como [5]:

Propriedade 2.11 *Quando os pólos de $\hat{H}(z)$ são fixados e os seus zeros são escolhidos de modo a minimizar $\|H(z) - \hat{H}(z)\|_{S_u}^2$, o erro $H(z) - \hat{H}(z)$ é tal que:*

$$\left[S_u(z) \left(H(z) - \hat{H}(z) \right) \right]_+ = g(z)V(z), \quad z^{-1}g(z) \in \mathcal{H}_2 \quad (2.50)$$

para algum $g(z)$ e com $V(z)$ dado por (2.25).

Teorema 2.12 *Se $\deg[\hat{H}(z)] = M$, então $\hat{H}(z) = B(z)/A(z)$ será um ponto estacionário de $\|H(z) - \hat{H}(z)\|_{S_u}^2$ se e somente se*

$$\left[S_u(z) \left(H(z) - \hat{H}(z) \right) \right]_+ = zQ(z)V^2(z)$$

para algum $Q(z) \in \mathcal{H}_2$ e $V(z) = z^M A(z^{-1})/A(z)$.

2.5.6 Identificação com entrada de espectro racional

Em boa parte deste trabalho estaremos interessados em filtros adaptativos cuja entrada é da forma $u(n) = G(z)w(n)$, onde a função racional $G(z) = E(z)/F(z)$ tem $\deg[G(z)] = L$ e $w(n)$ é branco com variância unitária. A Figura 2.2 mostra

Figura 2.2: Configuração de identificação com entrada de espectro racional dada por $u(n) = G(z)w(n)$.

a configuração resultante da escolha deste tipo de entrada. A função densidade espectral do processo $\{u(\cdot)\}$ da entrada será dada neste caso por

$$S_u(z) = G(z^{-1})G(z). \quad (2.51)$$

Com os sinais

$$\begin{aligned} y(n) &= H(z)G(z)w(n) \\ \hat{y}(n) &= \hat{H}(z)G(z)w(n) \\ e(n) &= G(z) \left(H(z) - \hat{H}(z) \right) w(n), \end{aligned}$$

resulta

$$\begin{aligned} E[e^2(n)] &= \left\langle G(z) \left(H(z) - \hat{H}(z) \right), G(z) \left(H(z) - \hat{H}(z) \right) \right\rangle \\ &= \left\langle H(z) - \hat{H}(z), H(z) - \hat{H}(z) \right\rangle_{S_u} \end{aligned}$$

que é equivalente a (2.20). Ou seja, o problema de adaptar $\hat{H}(z)$ para minimizar $E[e^2(n)]$ no caso de entrada de espectro racional é equivalente ao problema de se minimizar a norma L_2 ponderada $\|H(z) - \hat{H}(z)\|_{S_u}^2$.

Cabe ressaltar que a fatoração em (2.51) não é única, pois, para $G(z)$ de mínima fase, qualquer outra fatoração pode ser escrita como $F(z) = G(z)P(z)$, onde $P(z)$ é passa-tudo de norma unitária, o que levaria a $u(n) = G(z)P(z)w(n)$, com $P(z)w(n)$ sendo ainda branco de variância unitária. A existência desta função $P(z)$ não altera o erro médio quadrático nem suas derivadas, e portanto nos restringiremos a $G(z)$ de mínima fase no restante deste trabalho.

Figura 2.3: Filtro adaptativo na configuração de identificação inversa.

Além disto, é possível mostrar [22] que, para garantir que não haja mínimos locais, precisamos ter $\deg[\hat{H}(z)] \geq \deg[H(z)] + \deg[G(z)]$. Ou seja, para o caso de entrada de espectro racional, temos que ter sobremodelamento para garantir a existência de um único ponto de mínimo global, e para o caso de entrada branca ($L = 0$), basta modelamento suficiente.

2.5.7 Identificação inversa

A Figura 2.3 mostra um filtro adaptativo na configuração de identificação inversa. Nesta configuração o erro $e(n)$ não é dado pela diferença entre a saída do sistema desconhecido e a do filtro adaptativo. Ao invés disto o erro é a diferença entre a entrada $u(n)$ e uma aproximação desta entrada dada pela filtragem do sinal de saída $y(n)$ do sistema sendo identificado, ou seja

$$\begin{aligned}
 e(n) &= u(n) - \hat{u}(n) \\
 &= u(n) - \hat{H}_I(z)y(n) \\
 &= u(n) - H(z)\hat{H}_I(z)u(n) \\
 &= \left(1 - H(z)\hat{H}_I(z)\right)u(n).
 \end{aligned}$$

O filtro adaptativo $\hat{H}_I(z)$ é então adaptado de forma a minimizar o erro médio quadrático

$$E[e^2(n)] = \left\langle 1 - H(z)\hat{H}_I(z), S_u(z) \left(1 - H(z)\hat{H}_I(z)\right) \right\rangle,$$

onde $S_u(z)$ é a densidade espectral da entrada $u(n)$, sendo que esta igualdade é

obtida de forma similar à usada para chegar a (2.20). Caso a entrada seja branca,

$$\begin{aligned} E[e^2(n)] &= \left\langle 1 - H(z)\hat{H}_I(z), 1 - H(z)\hat{H}_I(z) \right\rangle \\ &= \left\| 1 - H(z)\hat{H}_I(z) \right\|^2. \end{aligned}$$

Vemos que o sistema $H(z)$ não é identificado diretamente, mas sim o seu inverso $H_I(z) = 1/H(z)$. Isto significa que se $H(z)$ for da forma

$$H(z) = \frac{C(z)}{D(z)}$$

com $\deg[H(z)] = N$, e $\hat{H}_I(z)$ for racional de ordem $\deg[\hat{H}(z)] = M = N$, então o mínimo global da superfície de erro quadrático será

$$\hat{H}(z) = \frac{D(z)}{C(z)}.$$

Nos capítulos seguintes veremos que a configuração de identificação inversa pode ser tratada como um caso particular da identificação com entrada de espectro racional. Isto nos permitirá analisar os dois casos de forma conjunta.

2.6 Estrutura de filtros IIR

Como mencionado anteriormente, todas as realizações mínimas de um filtro recursivo estão ligadas através de uma matriz de mudança de coordenadas $\mathbf{T}$. No restante deste trabalho iremos considerar duas estruturas para realização de filtros recursivos: a forma direta e a forma em treliça normalizada.

2.6.1 Forma direta

Seja $\hat{H}(z) = B(z)/A(z)$ como em (2.32). A saída $\hat{y}(n) = \hat{H}(z)u(n)$ pode ser escrita como $A(z)\hat{y}(n) = B(z)u(n)$, levando à equação de diferenças

$$\hat{y}(n) + a_1\hat{y}(n-1) + \dots + a_M\hat{y}(n-M) = b_0x(n) + b_1x(n-1) + \dots + b_Mx(n-M).$$

A Figura 2.4 mostra um filtro na forma direta obtido a partir desta equação de diferenças. Alterando a ordem dos subfiltros $B(z)$ e $A(z)$ chegamos à realização

Figura 2.4: Filtro na forma direta obtido a partir da equação de diferenças.

mínima da Figura 2.5. Uma descrição no espaço de estados deste filtro será

$$\begin{aligned}
 \mathbf{x}(n+1) &= \underbrace{\begin{bmatrix} -a_1 & -a_2 & \cdots & -a_M \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}}_{\mathbf{A}} \mathbf{x}(n) + \underbrace{\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{b}} u(n) \\
 \hat{y}(n) &= \underbrace{\begin{bmatrix} b_1 - b_0 a_1 & b_2 - b_0 a_2 & \cdots & b_M - b_0 a_M \end{bmatrix}}_{\mathbf{c}^t} \mathbf{x}(n) + \underbrace{b_0}_d u(n)
 \end{aligned}$$

de onde calculamos seu vetor de controlabilidade como

$$\mathcal{C}(z) = \begin{bmatrix} \frac{1}{A(z)} \\ \frac{z}{A(z)} \\ \vdots \\ \frac{z^{M-1}}{A(z)} \end{bmatrix}. \quad (2.52)$$

O filtro da Figura 2.5 admite também a seguinte descrição em espaço de

Figura 2.5: Filtro mínimo obtido a partir da forma da Figura 2.4.

estados aumentada:

$$\begin{bmatrix} \mathbf{x}(n+1) \\ w(n) \end{bmatrix} = \underbrace{\begin{bmatrix} -a_1 & -a_2 & \cdots & -a_M & 1 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}}_{\mathbf{Q}} \begin{bmatrix} \mathbf{x}(n) \\ u(n) \end{bmatrix} \quad (2.53)$$

$$\hat{y}(n) = \begin{bmatrix} b_0 & b_1 & b_2 & \cdots & b_M \end{bmatrix} \begin{bmatrix} \mathbf{x}(n+1) \\ w(n) \end{bmatrix},$$

sendo esta mais apropriada para a confecção e análise de algoritmos.

Figura 2.6: Forma em treliça normalizada.

2.6.2 Forma em treliça normalizada

Outra realização possível para um filtro recursivo é a forma em treliça normalizada da Figura 2.6. Esta realização pode ser descrita em espaço de estados como [5]

$$\begin{bmatrix} \mathbf{x}(n+1) \\ w(n) \end{bmatrix} = \underbrace{\begin{bmatrix} -\text{sen}\theta_1 & \times & \cdots & \times & \times \\ \cos\theta_1 & \times & \cdots & \times & \times \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & \cos\theta_{M-1} & \times & \times \\ 0 & \cdots & 0 & \cos\theta_M & \text{sen}\theta_M \end{bmatrix}}_{\mathbf{Q}} \begin{bmatrix} \mathbf{x}(n) \\ u(n) \end{bmatrix}$$

$$\hat{y}(n) = \begin{bmatrix} \nu_0 & \nu_1 & \nu_2 & \cdots & \nu_M \end{bmatrix} \begin{bmatrix} \mathbf{x}(n+1) \\ w(n) \end{bmatrix},$$

que lembra a forma da equação (2.53). Nesta realização os parâmetros que representam o filtro são os ângulos de rotação θ_k , $k = 1, \dots, M$ mais os coeficientes ν_l , $l = 0, \dots, M$.

Enquanto a forma direta é mais apropriada para a análise de certos aspectos e mais intuitiva para o desenvolvimento de algoritmos, a forma em treliça normalizada é útil para outros aspectos analíticos e possui melhores características numéricas, o que a torna preferível em algumas aplicações.

2.7 Algoritmos de filtragem adaptativa IIR baseados no gradiente descendente

Entre os algoritmos para adaptação dos parâmetros de um filtro adaptativo, aqueles baseados no gradiente descendente estão entre os mais usados. Portanto, não é surpresa que estes algoritmos também sejam usados no caso de filtros adaptativos IIR.

Estes algoritmos procuram os parâmetros de uma função racional $\hat{H}(z)$ que minimiza o valor quadrático médio do sinal de erro

$$e(n) = y(n) - \hat{y}(n) = [H(z) - \hat{H}(z)] u(n)$$

que é dado pela sua variância

$$E[e^2(n)] = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_u(e^{j\omega}) |H(e^{j\omega}) - \hat{H}(e^{j\omega})|^2 d\omega. \quad (2.54)$$

Esta procura é feita ajustando os parâmetros do filtro $\hat{H}(z)$ na direção negativa do gradiente desta função custo. A seguir serão descritos os algoritmos para a forma direta e a forma em treliça normalizada que serão utilizados no restante deste trabalho.

2.7.1 Forma direta

Na forma direta são adaptados os parâmetros $b_0, b_1, \dots, b_M$ e $a_1, \dots, a_M$ da função racional

$$\hat{H}(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1 z + \dots + b_M z^M}{1 + a_1 z + \dots + a_M z^M}$$

de forma a minimizar a função custo (2.54). Isto é feito atualizando os parâmetros na direção contrária ao gradiente da função custo,

$$\begin{bmatrix} b_0(n+1) \\ \vdots \\ b_M(n+1) \\ a_1(n+1) \\ \vdots \\ a_M(n+1) \end{bmatrix} = \begin{bmatrix} b_0(n) \\ \vdots \\ b_M(n) \\ a_1(n) \\ \vdots \\ a_M(n) \end{bmatrix} - \frac{\mu}{2} \begin{bmatrix} \partial E[e^2(n)] / \partial b_0(n) \\ \vdots \\ \partial E[e^2(n)] / \partial b_M(n) \\ \partial E[e^2(n)] / \partial a_1(n) \\ \vdots \\ \partial E[e^2(n)] / \partial a_M(n) \end{bmatrix}, \quad \mu > 0.$$

O passo de adaptação μ controla a taxa de convergência.

As derivadas $\partial E[e^2(n)]/\partial b_k(n)$ e $\partial E[e^2(n)]/\partial a_k(n)$ são obtidas examinando como a função custo $E[e^2(n)]$ varia com os parâmetros de $\hat{H}(z)$. Desta maneira, são calculadas no ponto do espaço de parâmetros ocupado por $\{b_k(n)\}$ e $\{a_k(n)\}$ no instante n , o que justifica a inclusão do índice de tempo nestes parâmetros.

De acordo com (2.48), a função custo (2.54) pode ser escrita como

$$E[e^2(n)] = \left\langle H(z) - \hat{H}(z), S_u(z) (H(z) - \hat{H}(z)) \right\rangle,$$

e usando as fórmulas

$$\frac{\partial E[e^2(n)]}{\partial b_k} = -2 \left\langle \frac{\partial \hat{H}(z)}{\partial b_k}, S_u(z) (H(z) - \hat{H}(z)) \right\rangle \quad (2.55)$$

$$\frac{\partial E[e^2(n)]}{\partial a_k} = -2 \left\langle \frac{\partial \hat{H}(z)}{\partial a_k}, S_u(z) (H(z) - \hat{H}(z)) \right\rangle \quad (2.56)$$

chegamos a

$$\begin{bmatrix} b_0(n+1) \\ \vdots \\ b_M(n+1) \\ a_1(n+1) \\ \vdots \\ a_M(n+1) \end{bmatrix} = \begin{bmatrix} b_0(n) \\ \vdots \\ b_M(n) \\ a_1(n) \\ \vdots \\ a_M(n) \end{bmatrix} - \mu \left\langle \begin{bmatrix} 1/A(z) \\ \vdots \\ z^M/A(z) \\ -z\hat{H}(z)/A(z) \\ \vdots \\ -z^M\hat{H}(z)/A(z) \end{bmatrix}, S_u(z) (H(z) - \hat{H}(z)) \right\rangle. \quad (2.57)$$

Esta equação será utilizada para adaptar os parâmetros de $\hat{H}(z)$ nas simulações que utilizam o gradiente exato, sendo que os produtos internos podem ser calculados através do cálculo de resíduos [26].

Por conveniência, definimos os sinais regressores filtrados

$$\begin{aligned} \nabla_{b_k}(n) &= \frac{z^k}{A(z)} u(n), & k = 0, \dots, M \\ \nabla_{a_k}(n) &= -\frac{z^k}{A(z)} \hat{H}(z) u(n), & k = 1, \dots, M \end{aligned}$$

Figura 2.7: Geração dos sinais regressores filtrados no algoritmo na forma direta.

que em conjunto com (2.48) nos permite escrever (2.57) como

$$\begin{bmatrix} b_0(n+1) \\ \vdots \\ b_M(n+1) \\ a_1(n+1) \\ \vdots \\ a_M(n+1) \end{bmatrix} = \begin{bmatrix} b_0(n) \\ \vdots \\ b_M(n) \\ a_1(n) \\ \vdots \\ a_M(n) \end{bmatrix} - \mu E \left\{ e(n) \begin{bmatrix} \nabla_{b_0}(n) \\ \vdots \\ \nabla_{b_M}(n) \\ \nabla_{a_1}(n) \\ \vdots \\ \nabla_{a_M}(n) \end{bmatrix} \right\}. \quad (2.58)$$

Na Figura 2.7 temos uma realização onde todos os sinais necessários para a atualização dos parâmetros estão disponíveis.

Apesar de termos todos os sinais disponíveis, não temos a esperança dos produtos necessários em (2.58). Na prática utilizamos uma aproximação estocástica

abandonando a operação de esperança, o que nos leva a

$$\begin{bmatrix} b_0(n+1) \\ \vdots \\ b_M(n+1) \\ a_1(n+1) \\ \vdots \\ a_M(n+1) \end{bmatrix} = \begin{bmatrix} b_0(n) \\ \vdots \\ b_M(n) \\ a_1(n) \\ \vdots \\ a_M(n) \end{bmatrix} - \mu e(n) \begin{bmatrix} \nabla_{b_0}(n) \\ \vdots \\ \nabla_{b_M}(n) \\ \nabla_{a_1}(n) \\ \vdots \\ \nabla_{a_M}(n) \end{bmatrix}. \quad (2.59)$$

Quando a adaptação for feita por esta fórmula, dizemos que ela utiliza o gradiente estocástico. No Apêndice A temos uma descrição do algoritmo implementado para este caso. Nas simulações numéricas que serão feitas adiante poderemos utilizar valores diferentes para a adaptação dos parâmetros dos pólos (μ_A) e dos zeros (μ_B).

2.7.2 Forma em treliça normalizada

O objetivo do filtro em treliça normalizada é o mesmo que o de filtros na forma direta, ambos procuram pontos de mínimo da função custo $E[e^2(n)]$, porém, em diferentes espaços de parâmetros. A principal vantagem da forma em treliça é a facilidade de verificação da estabilidade do filtro obtido a cada iteração, basta que se tenha $|\theta_k| < \pi/2$. Como será visto adiante, ela também tem vantagens sobre a forma direta relativas à velocidade de convergência.

O desenvolvimento do algoritmo adaptativo para a forma em treliça normalizada segue os mesmos passos da forma direta: o erro de saída $e(n)$ é diferenciado com relação aos parâmetros ν_k e θ_k do filtro para se obter um gradiente negativo. Ocorre que a obtenção de uma fórmula nos moldes de (2.59) é mais extensa que no caso da forma direta, e portanto não a reproduziremos aqui. Esta derivação pode ser encontrada em [5, sec. 7.4].

Além disto, ao se utilizar o gradiente com filtros em treliça, temos que a complexidade do algoritmo é da ordem de M^2 , com $\deg[\hat{H}(z)] = M$. Algoritmos com complexidade computacional da ordem de M podem ser obtidos através da utilização de um gradiente simplificado [5, sec. 7.5 a 7.7], como o que é mostrado

Figura 2.8: Obtenção dos regressores em filtros adaptativos na forma em treliça normalizada.

no Apêndice B. Neste caso, a adaptação dos parâmetros é feita através de

$$\nu_j(n+1) = \nu_j(n) + \mu_\nu \nabla_{\nu_j}(n) e(n)$$

$$\theta_j(n+1) = \theta_j(n) + \mu_\theta \nabla_{\theta_j}(n) e(n)$$

$$\text{Se } |\theta_j(n+1)| \geq \frac{\pi}{2}, \text{ então } \theta_j(n+1) = \theta_j(n),$$

onde μ_ν e μ_θ são os passos de adaptação utilizados para os parâmetros ν_j e θ_j , respectivamente.

A Figura 2.8, onde $\gamma_M = 1$ e

$$\gamma_j \triangleq \prod_{i=j+1}^M \cos \theta_i, \quad j = 1, \dots, M-1,$$

mostra como os regressores são obtidos. É interessante observar a semelhança com o modo de obtenção dos regressores da forma direta (Figura 2.7). Neste caso,

como foram feitas simplificações ao longo da dedução do gradiente, os regressores $\nabla_{\theta_j}(n)$ não correspondem exatamente às derivadas do erro quadrático em relação a θ_j . Mesmo assim, pode-se mostrar [5] que os pontos onde $E[\nabla_{\nu_j}(n)] = 0$ e $E[\nabla_{\theta_j}(n)] = 0$ são pontos estacionários do erro quadrático $\|H(z) - \hat{H}(z)\|_{S_u}^2$. Isto, conforme observado em simulações numéricas, sugere que pontos de convergência do algoritmo sejam pontos de mínimo do erro quadrático.

2.7.3 Identificação inversa com filtro na forma direta

Novamente, os passos para se chegar ao algoritmo adaptativo para a identificação inversa são os mesmos que para a configuração de identificação, e serão resumidos a seguir. Nos restringiremos a filtros adaptativos IIR da forma

$$\hat{H}(z) = \frac{B(z)}{A(z)} = \frac{b_o + b_1 z + \dots + b_M z^M}{1 + a_1 z + \dots + a_M z^M}$$

na forma direta. Neste caso, o erro tem a forma

$$\begin{aligned} e(n) &= u(n) - \hat{u}(n) \\ &= \left(1 - \hat{H}_I(z)H(z)\right) u(n), \end{aligned}$$

o erro quadrático médio é dado por

$$E[e^2(n)] = \left\langle 1 - H(z)\hat{H}_I(z), S_u(z) \left(1 - H(z)\hat{H}_I(z)\right) \right\rangle$$

e os elementos do vetor gradiente serão

$$\frac{\partial}{\partial b_k} E[e^2(n)] = -2 \left\langle \frac{z^k}{A(z)} H(z), S_u(z) \left(1 - H(z)\hat{H}_I(z)\right) \right\rangle \quad (2.60)$$

$$\frac{\partial}{\partial a_k} E[e^2(n)] = 2 \left\langle \frac{z^k}{A(z)} \hat{H}_I(z) H(z), S_u(z) \left(1 - H(z)\hat{H}_I(z)\right) \right\rangle. \quad (2.61)$$

Estas fórmulas podem ser usadas para se implementar um algoritmo de gradiente exato. Para obtermos um algoritmo que usa o gradiente estocástico, definimos os regressores filtrados

$$\begin{aligned} \nabla_{b_k}(n) &= \frac{z^k}{A(z)} y(n) = \frac{z^k}{A(z)} H(z) u(n), \quad k = 0, \dots, M \\ \nabla_{a_k}(n) &= -\frac{z^k}{A(z)} \hat{u}(n) = -\frac{z^k}{A(z)} \hat{H}_I(z) H(z) u(n), \quad k = 1, \dots, M. \end{aligned}$$

Com estes regressores e com o erro $e(n)$ temos, usando (2.20),

$$\begin{aligned}\frac{\partial}{\partial b_k} E[e^2(n)] &= -2E[e(n)\nabla_{b_k}(n)] \\ \frac{\partial}{\partial a_k} E[e^2(n)] &= -2E[e(n)\nabla_{a_k}(n)].\end{aligned}$$

Substituindo a esperança pelo próprio produto do erro com os regressores, chegamos a

$$\begin{bmatrix} b_0(n+1) \\ \vdots \\ b_M(n+1) \\ a_1(n+1) \\ \vdots \\ a_M(n+1) \end{bmatrix} = \begin{bmatrix} b_0(n) \\ \vdots \\ b_M(n) \\ a_1(n) \\ \vdots \\ a_M(n) \end{bmatrix} - \mu e(n) \begin{bmatrix} \nabla_{b_0}(n) \\ \vdots \\ \nabla_{b_M}(n) \\ \nabla_{a_1}(n) \\ \vdots \\ \nabla_{a_M}(n) \end{bmatrix}, \quad (2.62)$$

que tem a mesma forma que no caso de identificação com filtro na forma direta. A Figura 2.9 mostra uma realização de um filtro adaptativo IIR na configuração de identificação inversa junto com a geração dos sinais regressores. O algoritmo utilizado é descrito no Apêndice C.

2.8 Análise da velocidade de convergência com entrada branca

Como mencionado anteriormente, um dos problemas dos filtros adaptativos IIR é que estes podem alcançar regiões do espaço de parâmetros onde a convergência é lenta. A Figura 2.10 mostra um exemplo desta convergência lenta. O gradiente estocástico foi utilizado para identificar o sistema $H(z)$ cujos pólos e zeros se encontram na Tabela 2.1 junto com os parâmetros usados nesta simulação. Os passos utilizados são aproximadamente metade dos valores que levam a adaptação a divergir levando em conta o conjunto de realizações dos sinais. Em todas as simulações do restante deste trabalho serão utilizados passos com cerca da metade do valor máximo possível observado em simulações. Os polinômios $B_0(z)$ e $A_0(z)$ correspondem aos valores iniciais do filtro adaptativo. Nesta simulação vemos que para atingir um erro de -60dB são necessárias mais de 1×10^5 iterações.

Figura 2.9: Geração dos sinais regressores filtrados no algoritmo de identificação inversa na forma direta.

Função	Pólos	Zeros
$H(z^{-1})$	$0,8\angle \pm 45^\circ$ e $0,8\angle \pm 110^\circ$	$0,8\angle \pm 30^\circ$ e $0,5\angle \pm 60^\circ$
Parâmetros da simulação numérica		
Ordem de $\hat{H}(z)$	$M = 4$	
Passo μ_A	$1,5 \times 10^{-2}$	
Passo μ_B	2×10^{-2}	
$B_0(z), A_0(z)$	0,4 vezes os valores ótimos	

Tabela 2.1: Pólos e zeros de $H(z^{-1})$ junto com os parâmetros utilizados na simulação da Figura 2.10.

Figura 2.10: Exemplo de convergência lenta de filtros adaptativos IIR. Os parâmetros utilizados se encontram na Tabela 2.1 junto com o inverso dos pólos e zeros da função sendo identificada.

Como mostrado em [23] quando a convergência é lenta, esta tende a ocorrer próxima a superfície de erro reduzida. A Figura 2.11 ilustra este fato. Vemos que os valores dos coeficientes $b_k(n)$ seguem de perto os valores ótimos $b_{o,k}(n)$ dados os coeficientes $a_k(n)$.

O comportamento do erro na superfície de erro reduzida pode então ajudar na análise da convergência dos filtros adaptativos IIR. Ocorre que a equação (2.44) não é muito conveniente para estes fins, pois o vetor $\mathbf{v}$ tem infinitos termos e a relação dos parâmetros a serem otimizados com $\|H(z) - \hat{H}(z)\|^2$ é complicada.

Em [23] e [24] uma parametrização dos pólos baseada na teoria de realização balanceada foi proposta, tendo se mostrado eficaz na análise da velocidade de convergência de filtros adaptativos IIR. A seguir seus principais resultados serão revistos.

Figura 2.11: Adaptação dos parâmetros dos zeros e seus valores ótimos dados os pólos ($\circ$).

2.8.1 Parâmetros BLP

Ao tentar usar a equação (2.44) para analisar o comportamento de filtros adaptativos IIR, um problema é a dimensão infinita da forma de Hankel $\mathbf{\Gamma}_H$ e do vetor $\mathbf{v}$. Uma forma mais conveniente para a superfície de erro reduzida é obtida em [24], usando uma fatoração da forma de Hankel de $H(z)$ baseada na sua realização balanceada. Este procedimento é resumido a seguir.

Consideramos que $\deg[H(z)] = N$ e $\deg[\hat{H}(z)] = M$. Como visto na Seção (2.3.3), em uma realização balanceada podemos escrever os gramianos de controlabilidade e observabilidade como

$$\mathbf{K}_b = \mathcal{C}_b \mathcal{C}_b^t = \mathbf{\Sigma}$$

e

$$\mathbf{W}_b = \mathcal{O}_b^t \mathcal{O}_b = \mathbf{\Sigma},$$

respectivamente, onde $\mathcal{C}_b$ e $\mathcal{O}_b$ são as matrizes de controlabilidade e observabilidade da realização balanceada, respectivamente. As versões normalizadas destas

matrizes são definidas como

$$\overline{\mathcal{C}} \triangleq \Sigma^{-1/2} \mathcal{C}_b$$

e

$$\overline{\mathcal{O}} \triangleq \mathcal{O}_b \Sigma^{-1/2}.$$

Notamos que $\overline{\mathcal{C}}\overline{\mathcal{C}}^t = \overline{\mathcal{O}}^t\overline{\mathcal{O}} = \mathbf{I}$. Com (2.17) podemos escrever, então,

$$\Gamma_H = \mathcal{O}\mathcal{C} = \overline{\mathcal{O}}\Sigma\overline{\mathcal{C}},$$

e podemos, portanto, reescrever (2.44) como

$$\begin{aligned} \left\| H(z) - \hat{H}(z) \right\|^2 &= \mathbf{v}^t \Gamma_H^2 \mathbf{v} \\ &= \mathbf{v}^t \overline{\mathcal{C}}^t \Sigma \underbrace{\overline{\mathcal{O}}^t \overline{\mathcal{O}}}_{=\mathbf{I}} \Sigma \overline{\mathcal{C}} \mathbf{v} = \mathbf{v}^t \overline{\mathcal{C}}^t \Sigma^2 \overline{\mathcal{C}} \mathbf{v} \\ &= \boldsymbol{\alpha}^t \Sigma^2 \boldsymbol{\alpha}, \end{aligned} \quad (2.63)$$

onde $\boldsymbol{\alpha} \triangleq [\alpha_1 \ \alpha_2 \ \dots \ \alpha_N]^t = \overline{\mathcal{C}}\mathbf{v}$, com

$$\alpha_k = \langle \zeta_k(z), V(z) \rangle, \quad (2.64)$$

sendo $\zeta_k(z) = \zeta_{k,0} + \zeta_{k,1}z^1 + \dots$ a função de controlabilidade normalizada associada ao k -ésimo dos N estados da realização balanceada de $H(z)$. Os coeficientes $\zeta_{k,i}$ correspondem aos elementos da k -ésima das linhas de $\overline{\mathcal{C}}$. Como $\zeta_k(z)$ e $V(z)$ possuem norma unitária, resulta que $|\alpha_k| \leq 1$.

Em [23] é mostrado que, se $N \geq M$, então os coeficientes α_k constituem uma parametrização dos pólos de $\hat{H}(z)$, parametrização esta ligada ao sistema $H(z)$, de onde vem a denominação BLP (*Balanced form Linked Parameterization*).

2.8.2 Descrição local da adaptação

Admitamos que os zeros e a constante multiplicativa de $\hat{H}(z)$ são sempre ótimos dados os seus pólos, ou seja, a adaptação ocorre na superfície de erro reduzida. Os parâmetros dos pólos de $\hat{H}(z)$ são reunidos no vetor $\mathbf{w}$ e adaptados segundo o gradiente exato do erro quadrático médio. Temos então:

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \frac{\mu_A}{2} \frac{\partial}{\partial \mathbf{w}} \left\| H(z) - \hat{H}(z) \right\|^2 \Big|_{\mathbf{w}(n)}, \quad (2.65)$$

onde $\partial f / \partial \mathbf{w} \triangleq \begin{bmatrix} \partial f / \partial w_1 & \partial f / \partial w_2 & \dots \end{bmatrix}^t$.

Passemos agora à obtenção de uma descrição local da adaptação em termos dos parâmetros BLP, isto é, de uma expressão para $\boldsymbol{\alpha}(n+1)$ em função de $\boldsymbol{\alpha}(n)$. Uma aproximação de primeira ordem é

$$\boldsymbol{\alpha}(n+1) \approx \boldsymbol{\alpha}(n) + \underbrace{\frac{\partial \boldsymbol{\alpha}}{\partial \mathbf{w}} \bigg|_{\mathbf{w}(n)}}_{\triangleq \mathbf{J}} [\mathbf{w}(n+1) - \mathbf{w}(n)], \quad (2.66)$$

onde $\mathbf{J}$ é uma matriz Jacobiana de sensibilidade

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \alpha_1}{\partial w_1} & \frac{\partial \alpha_1}{\partial w_2} & \dots & \frac{\partial \alpha_1}{\partial w_M} \\ \frac{\partial \alpha_2}{\partial w_1} & \frac{\partial \alpha_2}{\partial w_2} & \dots & \frac{\partial \alpha_2}{\partial w_M} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial \alpha_N}{\partial w_1} & \frac{\partial \alpha_N}{\partial w_2} & \dots & \frac{\partial \alpha_N}{\partial w_M} \end{bmatrix}.$$

Agora, de (2.65) e usando a regra da cadeia da derivação $\partial f / \partial \mathbf{x} = [\partial \mathbf{y} / \partial \mathbf{x}]^t \partial f / \partial \mathbf{y} = \mathbf{J}^t \partial f / \partial \mathbf{y}$, temos

$$\mathbf{w}(n+1) - \mathbf{w}(n) = -\frac{\mu_A}{2} \mathbf{J}^t \frac{\partial}{\partial \boldsymbol{\alpha}} \left\| H(z) - \hat{H}(z) \right\|^2 \bigg|_{\mathbf{w}(n)}.$$

Por sua vez, da expressão (2.63) resulta que

$$\frac{\partial}{\partial \boldsymbol{\alpha}} \left\| H(z) - \hat{H}(z) \right\|^2 = 2 \boldsymbol{\Sigma}^2 \boldsymbol{\alpha},$$

e, portanto,

$$\mathbf{w}(n+1) - \mathbf{w}(n) = -\mu_A \mathbf{J}^t \boldsymbol{\Sigma}^2 \boldsymbol{\alpha}.$$

Com (2.66) chegamos finalmente a:

$$\boldsymbol{\alpha}(n+1) = [\mathbf{I} - \mu_A \mathbf{J} \mathbf{J}^t \boldsymbol{\Sigma}^2] \boldsymbol{\alpha}(n). \quad (2.67)$$

A velocidade de convergência de $\hat{H}(z)$ está ligada aos autovalores λ_k de $\mathbf{J} \mathbf{J}^t \boldsymbol{\Sigma}^2$, mais especificamente ao seu espalhamento $\chi(\mathbf{J} \mathbf{J}^t \boldsymbol{\Sigma}^2) = \lambda_{\max} / \lambda_{\min}$. Além disto, esta parametrização possui a propriedade de separar parcialmente o efeito do sistema sendo identificado, que afeta $\boldsymbol{\Sigma}$ e $\mathbf{J}$, e o efeito da parametrização, que afeta apenas $\mathbf{J}$.

Função	Pólos	Zeros
$H(z^{-1})$	$0,8\angle \pm 30^\circ$ e $0,9\angle \pm 100^\circ$	$1/0,8\angle \pm 30^\circ$ e $1/0,9\angle \pm 100^\circ$
Parâmetros da simulação numérica		
Ordem de $\hat{H}(z)$	$M = 4$	
Passo μ_A	2×10^{-2}	
Passo μ_B	3×10^{-2}	
$B_0(z), A_0(z)$	0,4 vezes os valores ótimos	

Tabela 2.2: Inverso dos pólos e zeros do sistema $H(z)$ sendo identificado junto com os parâmetros do filtro adaptativo IIR usado na simulação da Figura 2.12.

Em [23, 18] várias propriedades desta parametrização são discutidas. Em particular é mostrado que quando a forma de Hankel de $H(z)$ possui um pequeno espalhamento de seus valores singulares, contribuindo para um menor espalhamento dos autovalores de $\mathbf{J}\mathbf{J}^t\mathbf{\Sigma}^2$, a velocidade de convergência do filtro adaptativo é maior. Portanto, é de se esperar que, se o sistema $H(z)$ sendo identificado for um sistema passa-tudo (os quais têm todos os valores singulares iguais a 1), a convergência tenderá a ser mais rápida. Mais genericamente isto também vale se

$$H(z) = s \frac{\overline{D}(z)}{D(z)} + c,$$

que tem todos os autovalores de $\mathbf{\Gamma}_H$ iguais a s . A constante c aparece pois a forma de Hankel não leva em conta o termo direto h_0 .

Para ilustrar este fato, temos na Figura 2.12 o resultado da identificação de um sistema passa-tudo com os mesmos pólos do sistema da Seção 2.8. Os parâmetros usados nesta simulação se encontram na Tabela 2.2. Podemos notar que para atingir um erro de -60dB foram necessárias cerca de 1500 iterações.

2.9 Algoritmo de aproximações sucessivas (SA)

Utilizando os resultados apontados na seção anterior, em [24] foi proposto um algoritmo de filtragem adaptativa IIR, denominado algoritmo de aproximações sucessivas. Este algoritmo, através de um bloco auxiliar, busca tornar o sistema a ser identificado mais parecido com um sistema passa-tudo.

Figura 2.12: Identificação de um sistema $H(z)$ passa-tudo onde $H(z^{-1})$ tem pólos em $0,8 \angle \pm 45^\circ$ e $0,8 \angle \pm 110^\circ$ e zeros recíprocos a estes.

2.9.1 Definição do algoritmo

O algoritmo de aproximações sucessivas é [24]:

1. A partir de uma estimativa $\hat{H}(z) = B(z)/A(z)$ do sistema $H(z)$, calcula-se a função de transferência de um bloco auxiliar

$$H_a(z) = \frac{B(z) - \overline{A}(z)}{A(z)}$$

2. Um filtro adaptativo IIR $\hat{H}_b(z)$ faz a identificação de $H_b(z) = H(z) - H_a(z)$ durante um número estabelecido n_a de amostras. Para esta identificação pode ser usada em princípio qualquer realização de um filtro adaptativo IIR.

3. A partir do valor resultante de $\hat{H}_b(z)$, determina-se $H'(z) = \hat{H}_b(z) + H_a(z)$ (notar que todas as funções envolvidas são conhecidas) e uma nova estimativa $\hat{H}'(z) = B'(z)/A'(z)$ de $H(z)$ é obtida através da solução numérica da minimização do erro quadrático de equação

$$\|A'(z)H'(z) - B'(z)\|^2.$$

Figura 2.13: Diagrama em blocos do algoritmo de aproximações sucessivas para filtragem adaptativa IIR.

Um número estabelecido n_x de amostras é utilizado para este procedimento computacional, durante o qual a adaptação é congelada. O Apêndice D mostra como calcular $B'(z)$ e $A'(z)$ neste caso.

4. De forma a reduzir transitórios intensos, os parâmetros do bloco auxiliar e do filtro adaptativo sofrem uma transição suave, durante n_t amostras, para os novos valores decorrentes da nova estimativa $\hat{H}'(z)$. A adaptação continua congelada.
5. A adaptação continua congelada por mais n_w amostras, de modo a não ser influenciada por transitórios residuais.

A Figura 2.13 mostra um diagrama em blocos do algoritmo. Notamos que no terceiro passo não foi utilizado um filtro adaptativo para a obtenção de uma nova aproximação de $H(z)$. Isto porque, conforme discutido em [23] e [24], não podemos esperar uma convergência rápida neste caso. Porém, como todas as funções são conhecidas, uma solução numérica, mesmo que sub-ótima, mostrou-se mais adequada. No caso do segundo passo, é de se esperar que quando $\hat{H}(z) \approx H(z)$ o filtro $\hat{H}_b(z)$ convirja rapidamente, pois enxerga um sistema próximo a um sistema passa-tudo.

Função	Pólos	Zeros
$H(z^{-1})$	$0,8\angle \pm 45^\circ$ e $0,8\angle \pm 110^\circ$	$0,8\angle \pm 30^\circ$ e $0,5\angle \pm 60^\circ$
Parâmetros da simulação numérica (filtro na forma direta)		
Ordem de $\hat{H}(z)$	$M = 4$	
Passo μ_A	$1,5 \times 10^{-2}$	
Passo μ_B	2×10^{-2}	
$B_0(z), A_0(z)$	0,4 vezes os valores ótimos	
Parâmetros da simulação numérica (algoritmo SA)		
Ordem de $\hat{H}(z)$	$M = 4$	
Passo μ_A	$1,5 \times 10^{-2}$	
Passo μ_B	2×10^{-2}	
n_a	300	
n_x	100	
n_t	50	
n_w	50	
$B_0(z), A_0(z)$	0,4 vezes os valores ótimos	

Tabela 2.3: Pólos e zeros de $H(z^{-1})$ junto com os parâmetros utilizados nas simulações da Figura 2.14.

2.9.2 Exemplo numérico

O algoritmo SA foi utilizado para a identificação do mesmo sistema $H(z)$ da Seção 2.8, cujos inversos dos pólos e zeros se encontram na Tabela 2.3, junto com os parâmetros utilizados em ambas as simulações. A Figura 2.14 mostra o erro obtido para este filtro e para um filtro na forma direta adaptado pelo gradiente estocástico. Vemos que o algoritmo SA apresenta um grande aumento na velocidade de convergência se comparado com o filtro na forma direta.

2.10 Velocidade de convergência de filtros adaptativos IIR na configuração inversa

Em [19] é analisado o caso da convergência lenta de filtros adaptativos IIR na configuração de identificação inversa. Nesta seção apresentaremos alguns dos principais resultados desta análise.

Começamos com um exemplo de convergência lenta em configuração de identificação inversa. Na Tabela 2.4 temos o inverso dos pólos e zeros da função $H(z)$

Figura 2.14: Erro obtido com o algoritmo de aproximações sucessivas (A) e com um filtro adaptativo na forma direta adaptado pelo gradiente estocástico (B) na identificação do sistema $H(z)$ conforme a Tabela 2.3.

sendo identificada junto com os parâmetros utilizados na simulação numérica. O erro obtido é mostrado na Figura 2.15. Vemos que neste caso são necessárias cerca de 5×10^4 iterações para que um erro de -60dB seja atingido, mostrando que na configuração de identificação inversa podemos ter a mesma convergência lenta do caso de identificação. Na Figura 2.16 vemos que os pólos e zeros de $\hat{H}_I(z^{-1})$ obtido após 10^5 iterações correspondem aproximadamente aos pólos e zeros de $1/H(z)$, como havia sido dito na Seção 2.7.3.

Em [19] e [20] é mostrado que quando esta adaptação é lenta não são os valores

Função	Pólos	Zeros
$H(z^{-1})$	$0,8 \angle \pm 45^\circ$ e $0,8 \angle \pm 110^\circ$	$0,8 \angle \pm 115^\circ$ e $0,9 \angle \pm 70^\circ$
Parâmetros da simulação numérica		
Ordem de $\hat{H}(z)$	$M = 4$	
Passo μ_A	2×10^{-2}	
Passo μ_B	4×10^{-2}	
$B_0(z), A_0(z)$	0,1 vezes os valores ótimos	

Tabela 2.4: Inverso dos pólos do sistema $H(z)$ sendo identificado e parâmetros usados na simulação da Figura 2.15.

Figura 2.15: Exemplo de convergência lenta de filtros adaptativos IIR na configuração de identificação inversa. O inverso dos pólos e zeros da função $H(z)$ sendo identificada se encontram na Tabela 2.4 junto com os parâmetros utilizados na simulação.

Figura 2.16: Pólos (“o”) e zeros (“x”) de $H(z^{-1})$ (esquerda) e $\hat{H}_I(z^{-1})$ (direita).

singulares de $H(z)$ que são importantes para a velocidade de convergência do filtro adaptativo, e sim os do sistema "projetado"

$$H_p(z) = [H(z^{-1})H(z)]_{\oplus}. \quad (2.68)$$

Será mostrado no Capítulo 4 que a identificação inversa pode ser tratada como um caso particular da identificação com entrada não-branca. Uma demonstração de (2.68) será então desenvolvida neste contexto.

Se os valores singulares de Hankel $\sigma_{p,k}$ de $H_p(z)$ forem todos iguais, podemos esperar uma convergência mais rápida do filtro adaptativo. Esta condição é obtida se tivermos

$$H(z) = c \frac{D(z) - s\overline{D}(z)}{D(z)}, \quad (2.69)$$

que resultará em $\sigma_{p,k} = s$, $k = 1, \dots, N$. Com esta condição, em [20] o algoritmo de aproximações sucessivas foi estendido para o caso de identificação inversa. Os resultados obtidos para este caso são similares aos do caso de identificação e, portanto, não serão abordados aqui.

2.11 Conclusão

Neste capítulo apresentamos resultados referentes à aproximação racional $\mathcal{H}_2$. Estes resultados servirão nas análises do restante deste trabalho, principalmente no Capítulo 4. Também foram apresentadas realizações de filtros adaptativos IIR junto com algoritmos adaptativos para elas. Durante o restante do trabalho estas serão utilizadas em simulações numéricas.

No final do capítulo abordamos alguns resultados já existentes da análise da velocidade de convergência de filtros adaptativos IIR com entrada branca. Rapidamente mostramos o algoritmo de aproximações sucessivas decorrente destes resultados e vimos que ele possibilita uma maior velocidade de convergência se comparado com outros algoritmos. A seguir, alguns pontos da extensão dos resultados anteriores para a configuração de identificação inversa foram citados.

Nos próximos capítulos apresentaremos uma análise do que ocorre quando a entrada não é branca. Os resultados deste capítulo serão utilizados para che-

gamos a um melhor entendimento dos fatores que influenciam a velocidade de convergência de filtros adaptativos IIR com este tipo de entrada.

CAPÍTULO 3 EFEITOS DE UMA ENTRADA DE ESPECTRO RACIONAL

Neste capítulo abordaremos, através de simulações numéricas, os efeitos na convergência de filtros adaptativos IIR de uma entrada com espectro racional não necessariamente branco. Simulações deste tipo são comuns na fase exploratória de projetos envolvendo filtragem adaptativa, sendo que, no caso de filtragem adaptativa IIR, os resultados podem ser obscuros, em princípio. Desta forma, os resultados destas simulações numéricas servirão de motivação para a análise que será feita no capítulo seguinte.

Iniciaremos vendo o que ocorre quando o sinal de entrada passa de branco para não-branco. A seguir, mantendo a entrada não-branca, verificaremos o efeito dos zeros da função $H(z)$ em busca de indícios da existência de uma condição em que a convergência é mais rápida. Terminaremos com simulações que mostram o resultado de se utilizar uma realização diferente para o filtro adaptativo $\hat{H}(z)$.

Consideraremos que a função de entrada $u(n)$ possui espectro racional como na Seção 2.5.6, ou seja, $u(n) = G(z)w(n)$, com $w(n)$ branco de variância unitária.

3.1 Dependência da velocidade de convergência com relação ao espectro da entrada

Voltemos à função passa-tudo da Seção 2.8.1. Como visto anteriormente, com entrada branca tínhamos uma convergência rápida para a sua identificação. Veremos a seguir o que ocorre quando a entrada possui um espectro racional não-

Funções utilizadas		
Função	Pólos	Zeros
$H(z^{-1})$	$0,8\angle \pm 30^\circ$ e $0,9\angle \pm 100^\circ$	$1/0,8\angle \pm 30^\circ$ e $1/0,9\angle \pm 100^\circ$
$G(z^{-1})$	$0,7\angle \pm 110^\circ$	$0,75\angle \pm 30^\circ$
Parâmetros da simulação numérica		
Ordem de $\hat{H}(z)$	$M = 4$	
Passos μ_A, μ_B	2×10^{-2}	
$B_0(z), A_0(z)$	0, 1 vezes os valores ótimos	

Tabela 3.1: Pólos e zeros do sistema sendo identificado e do filtro utilizado para gerar a entrada não-branca junto com os parâmetros do filtro adaptativo IIR.

branco. Na Tabela 3.1 temos os pólos e zeros da função $H(z)$ sendo identificada e da função $G(z)$ usada para gerar a entrada $u(n)$ e na Figura 3.1, a resposta em frequência de ambas as funções.

A Figura 3.2 mostra o resultado de uma simulação numérica da identificação de $H(z)$ utilizando um filtro na forma direta adaptado pelo gradiente estocástico. Os parâmetros utilizados para esta simulação estão na Tabela 3.1. Vemos que agora, após cerca de 1×10^5 iterações o erro ainda é -30dB , um desempenho muito pior que para o caso de entrada branca mostrado na Figura 2.12 e reproduzido junto com o resultado desta simulação na Figura 3.3.

Este resultado levanta a hipótese de que a densidade espectral $\mathcal{S}_u(z)$ da entrada influencia as propriedades de convergência da adaptação de $\hat{H}(z)$. Desta maneira pode haver casos em que a convergência de filtros adaptativos seja mais rápidas com uma entrada não-branca. A simulação a seguir ilustra uma situação onde isto ocorre. As funções envolvidas bem como os parâmetros utilizados se encontram na Tabela 3.2. Na Figura 3.4 temos a resposta em frequência de $H(z)$ e $G(z)$.

Na Figura 3.5 temos o resultado das simulações. Vemos que para a entrada branca são necessárias 1×10^5 iterações para se alcançar um erro de -60dB , enquanto que no caso de entrada não-branca são necessárias apenas 1×10^4 iterações. Isto sugere que existem sistemas que são mais fáceis de serem identificados caso a entrada tenha um espectro mais adequado que o branco. Para estes casos então, uma eventual abordagem envolvendo o branqueamento da entrada não seria eficaz.

Figura 3.1: Resposta em frequência de $H(z)$ (linha pontilhada) e de $G(z)$ (linha contínua).

Figura 3.2: Identificação de um sistema $H(z)$ passa-tudo com os pólos e zeros da Tabela 3.1 e entrada não-branca.

Figura 3.3: Identificação de um sistema passa-tudo utilizando (A) entrada branca e (B) entrada não-branca.

Funções utilizadas		
Função	Pólos	Zeros
$H(z^{-1})$	$0,8\angle \pm 30^\circ$ e $0,9\angle \pm 100^\circ$	$1,6\angle \pm 55^\circ$ e $0,9\angle \pm 115^\circ$
$G(z^{-1})$	$0,7\angle \pm 110^\circ$	$0,75\angle \pm 30^\circ$
Parâmetros da simulação numérica (entrada branca)		
Ordem de $\hat{H}(z)$	$M = 4$	
Passo μ_A	2×10^{-2}	
Passo μ_B	5×10^{-2}	
$B_0(z), A_0(z)$	0,1 vezes os valores ótimos	
Parâmetros da simulação numérica (entrada não-branca)		
Ordem de $\hat{H}(z)$	$M = 4$	
Passo μ_A	$1,5 \times 10^{-1}$	
Passo μ_B	4×10^{-2}	
$B_0(z), A_0(z)$	0,1 vezes os valores ótimos	

Tabela 3.2: Pólos e zeros do sistema sendo identificado e do filtro utilizado para gerar a entrada não-branca, junto com os parâmetros do filtro adaptativo IIR no caso em que a entrada não-branca possibilita uma convergência mais rápida.

Figura 3.4: Resposta em frequência de $H(z)$ (linha pontilhada) e de $G(z)$ (linha contínua) para pólos e zeros da Tabela 3.2.

Figura 3.5: Resultado da simulação numérica cujas funções e parâmetros se encontram na Tabela 3.2 utilizando (A) entrada não-branca e (B) entrada branca.

Funções utilizadas		
Função	Pólos	Zeros
$H_1(z^{-1})$	$0,8\angle \pm 45^\circ$ e $0,8\angle \pm 110^\circ$	$1,6\angle \pm 55^\circ$ e $0,9\angle \pm 115^\circ$
$H_2(z^{-1})$	$0,8\angle \pm 45^\circ$ e $0,8\angle \pm 110^\circ$	$0,9\angle \pm 140^\circ$ e $0,8\angle \pm 115^\circ$
$G(z^{-1})$	$0,7\angle \pm 110^\circ$	$0,75\angle \pm 30^\circ$
Parâmetros da simulação numérica (identificação de $H_1(z^{-1})$)		
Ordem de $\hat{H}(z)$	$M = 4$	
Passo μ_A	2×10^{-2}	
Passo μ_B	4×10^{-2}	
$B_0(z), A_0(z)$	0,1 vezes os valores ótimos	
Parâmetros da simulação numérica (identificação de $H_2(z^{-1})$)		
Ordem de $\hat{H}(z)$	$M = 4$	
Passos μ_A, μ_B	5×10^{-2}	
$B_0(z), A_0(z)$	0,1 vezes os valores ótimos	

Tabela 3.3: Pólos e zeros dos sistemas sendo identificados e do filtro $G(z)$ utilizado para gerar a entrada não-branca. Parâmetros dos filtros adaptativos IIR.

3.2 Dependência com relação aos zeros de $H(z)$

Conforme visto no capítulo anterior, em [23] foi demonstrado que as posições dos zeros de $H(z)$ são importantes para a velocidade de convergência de filtros adaptativos IIR usados em sua identificação. Na próxima simulação compararemos a velocidade de convergência da identificação de dois sistemas com zeros diferentes mas com pólos iguais, submetidos a uma entrada não-branca.

Os pólos e zeros dos sistemas considerados estão na Tabela 3.3, junto com os pólos e zeros da função $G(z)$ usada para gerar a entrada $u(n)$ e os parâmetros dos filtros adaptativos IIR. As respostas em frequência destes sistemas se encontram na Figura 3.6 e na Figura 3.7 temos o resultado de ambas as simulações.

Vemos que para o sistema $H_1(z)$ a convergência é mais rápida que para o sistema $H_2(z)$. Este resultado levanta a hipótese de haver, como no caso de entrada branca, uma condição nos zeros do sistema sendo identificado que aumente a velocidade de convergência. Caso esta condição exista, abre-se a possibilidade de se modificar o algoritmo de aproximações sucessivas para incluir os casos de entrada não-branca.

Figura 3.6: Resposta em frequência de $H_1(z)$ (A), $H_2(z)$ (B) e de $G(z)$ (C) da Tabela 3.3.

Figura 3.7: Identificação dos sistemas $H_1(z)$ (A) e $H_2(z)$ (B), cujos pólos e zeros se encontram na Tabela 3.3 junto com os do filtro $G(z)$ usado para gerar a entrada $u(n)$.

Funções utilizadas		
Função	Pólos	Zeros
$H(z^{-1})$	$0,8\angle \pm 30^\circ$ e $0,9\angle \pm 100^\circ$	$1/0,8\angle \pm 30^\circ$ e $1/0,9\angle \pm 100^\circ$
$G(z^{-1})$	$0,7\angle \pm 110^\circ$	$0,75\angle \pm 30^\circ$
Parâmetros da simulação numérica (entrada branca)		
Ordem de $\hat{H}(z)$	$M = 4$	
Passo μ_θ	2×10^{-2}	
Passo μ_ν	1×10^{-1}	
$B_0(z), A_0(z)$	0,1 vezes os valores ótimos	
Parâmetros da simulação numérica (entrada não-branca)		
Ordem de $\hat{H}(z)$	$M = 4$	
Passo μ_θ	2×10^{-2}	
Passo μ_ν	4×10^{-2}	
$B_0(z), A_0(z)$	0,1 vezes os valores ótimos	

Tabela 3.4: Pólos e zeros do sistema sendo identificado e do filtro utilizado para gerar a entrada não-branca junto com os parâmetros dos filtros adaptativos IIR. Neste caso os valores de $B_0(z)$ e $A_0(z)$ foram convertidos para os parâmetros do filtro em treliça.

3.3 Dependência com relação à realização de $\hat{H}(z)$

Os filtros adaptativos considerados nas seções anteriores utilizavam a forma direta de realização. A seguir mostraremos algumas simulações onde os filtros adaptativos estão na forma em treliça e compararemos os resultados obtidos com os anteriores.

Consideremos novamente o filtro passa-tudo e a função $G(z)$ cujos pólos e zeros se encontram na Tabela 3.1 e são reproduzidos na Tabela 3.4. Na Figura 3.8 temos o resultado da identificação deste sistema usando a entrada não-branca gerada com o filtro $G(z)$ e usando entrada branca. Os parâmetros utilizados nestas simulações se encontram na Tabela 3.4.

Os resultados obtidos com os filtros adaptativo na forma em treliça são parecidos com aqueles obtidos para a forma direta mostradas na Figura 3.3: enquanto que no caso da entrada branca a convergência é rápida, o mesmo não ocorre para o caso de entrada não-branca. Cabe ressaltar que no caso de entrada branca a velocidade de convergência é maior para o filtro em treliça que para o filtro na forma direta, sendo que este fato já fora verificado em algumas simulações de [23]

Figura 3.8: Identificação do sistema passa tudo $H(z)$ cujos pólos e zeros se encontram na Tabela 3.4. A curva (A) mostra o erro quadrático médio obtido com entrada não-branca e a (B) com entrada branca. Em ambos os casos o filtro adaptativo IIR utilizado se encontra na forma em treliça.

onde os pólos de $H(z)$ se encontravam concentrados.

Consideremos agora o caso da Seção 3.2 que para a forma direta apresentou uma maior velocidade de convergência com entrada não-branca. Os resultados obtidos com filtros adaptativos em treliça se encontram na Figura 3.9 e os parâmetros utilizados, na Tabela 3.5.

Vemos nas Figuras 3.8 e 3.9 que os resultados são parecidos com os obtidos com filtros adaptativos IIR na forma direta. Na Figura 3.9 vemos que a utilização da forma em treliça já proporcionou um ganho na velocidade de convergência com entrada branca sobre a identificação utilizando a forma direta. Isto foi observado em [23] quando havia uma concentração dos pólos do sistema sendo observado.

Funções utilizadas		
Função	Pólos	Zeros
$H(z^{-1})$	$0,8\angle \pm 45^\circ$ e $0,8\angle \pm 110^\circ$	$1,6\angle \pm 55^\circ$ e $0,9\angle \pm 115^\circ$
$G(z^{-1})$	$0,7\angle \pm 110^\circ$	$0,75\angle \pm 30^\circ$
Parâmetros da simulação numérica (entrada branca)		
Ordem de $\hat{H}(z)$	$M = 4$	
Passos μ_θ, μ_ν	4×10^{-2}	
$B_0(z), A_0(z)$	0,1 vezes os valores ótimos	
Parâmetros da simulação numérica (entrada não-branca)		
Ordem de $\hat{H}(z)$	$M = 4$	
Passo μ_θ	2×10^{-1}	
Passo μ_ν	8×10^{-2}	
$B_0(z), A_0(z)$	0,1 vezes os valores ótimos	

Tabela 3.5: Pólos e zeros do sistema sendo identificado e do filtro utilizado para gerar a entrada não-branca junto com os parâmetros dos filtros adaptativos IIR.

Figura 3.9: Identificação do sistema da tabela 3.5 utilizando filtro adaptativo em treliça. A curva (A) mostra o erro quadrático médio obtido com entrada não-branca e (B) o obtido com entrada branca.

Funções utilizadas		
Função	Pólos	Zeros
$H(z^{-1})$	$0,8\angle \pm 30^\circ$ e $0,9\angle \pm 100^\circ$	$1/0,8\angle \pm 30^\circ$ e $1/0,9\angle \pm 100^\circ$
$G(z^{-1})$	$0,7\angle \pm 110^\circ$	$0,75\angle \pm 30^\circ$
Parâmetros da simulação numérica		
Ordem de $\hat{H}(z)$	$M = 4$	
Passo μ_A	3×10^{-2}	
Passo μ_B	2×10^{-2}	
$B_0(z), A_0(z)$	0,1 vezes os valores ótimos	

Tabela 3.6: Inverso dos pólos e zeros do sistema $H(z)$ sendo identificado e do filtro $G(z)$ utilizado para gerar a entrada não-branca junto com os parâmetros do filtro adaptativo IIR usado para a identificação de $H(z)$.

3.4 Identificação inversa com entrada não-branca

Consideremos agora um filtro adaptativo na configuração de identificação inversa da Figura 2.3. A Figura 3.10 mostra o erro obtido na identificação inversa do sistema $H(z)$ usado na Seção 2.8. Os parâmetros usados na adaptação estão na Tabela 3.6.

Vemos que neste caso, em 10^5 iterações o erro obtido é da ordem de -40dB , ou seja, um resultado um pouco melhor que no caso de identificação. Porém, após atingir este erro a velocidade de convergência ainda é muito baixa.

3.5 Conclusão e motivação para análise

Neste capítulo apresentamos algumas simulações que mostram que as propriedades de convergência de filtros adaptativos IIR com entrada branca não se mantêm no caso de entrada não-branca. Sistemas que são facilmente identificados com entrada branca nem sempre o são com entrada não-branca. Além disto, existem sistemas em que o inverso também é verdadeiro, eles são mais facilmente identificados com uma entrada não-branca adequada do que com entrada branca. Com isto, podemos afirmar que, apesar do branqueamento da entrada levar a um caso com propriedades conhecidas, isto não significa que haverá uma melhora na velocidade de convergência do filtro adaptativo.

Figura 3.10: Erro obtido na identificação inversa de $H(z)$.

Estes fatos justificam uma análise das propriedades de filtros submetidos a entradas de diferentes espectros, a qual será feita no capítulo seguinte. Com esta análise pretendemos estender os resultados obtidos para o caso de entrada branca e proporcionar um melhor entendimento do que ocorre no caso de filtros adaptativos com entrada não-branca. O caso de identificação inversa também será abordado.

CAPÍTULO 4 VELOCIDADE DE CONVERGÊNCIA COM ENTRADA DE ESPECTRO RACIONAL

Nos capítulos anteriores foram abordados alguns aspectos teóricos da aproximação racional $\mathcal{H}_2$ ligados ao problema de filtragem adaptativa IIR. Em particular, foram revistos resultados teóricos apresentados em [23], resultantes da análise de filtros adaptativos IIR com entrada branca. Simulações numéricas que mostram a dependência da velocidade de convergência de filtros adaptativos IIR com relação à densidade espectral da entrada também foram apresentadas.

Neste capítulo apresentaremos o principal resultado deste trabalho, que é uma extensão da análise para entrada branca efetuada anteriormente de modo a incluir também o caso de entrada não-branca, em particular uma entrada de espectro racional.

Iniciaremos mostrando que a configuração de identificação inversa pode ser considerada um caso particular de identificação com entrada de espectro racional. Isto nos permitirá tratar ambos os casos de maneira unificada nas seções seguintes. Depois, verificaremos que, nos exemplos apresentados no Capítulo 3, a convergência lenta com entrada não-branca não se deve a uma convergência lenta dos zeros. Motivados por este resultado, seguiremos com a determinação de uma base para o espaço $\mathcal{H}_2$ que permita escrever o erro quadrático de maneira análoga à (2.42). Isto nos levará a uma condição para maior velocidade de convergência de filtros adaptativos IIR com entrada não-branca, abrindo caminho para uma generalização do algoritmo de aproximações sucessivas.

4.1 Unificação das configurações de identificação inversa e identificação com entrada não-branca

Consideremos o caso particular de identificação com entrada não-branca da Figura 4.1-(A), onde o sistema sendo identificado é

$$H_I(z) = \frac{1}{H(z)} = \frac{D(z)}{C(z)}$$

e o sinal de entrada $u(n)$ é da forma $u(n) = G(z)w(n)$ com

$$G(z) = H(z) = \frac{C(z)}{D(z)}. \quad (4.1)$$

Consideraremos ainda que $H(z)$ é de mínima fase e, portanto, $H_I(z)$ é estável. Claramente, se tivermos $\deg[\hat{H}_I(z)] = \deg[H_I(z)]$, o valor ótimo de $\hat{H}_I(z)$ é $\hat{H}_I(z) = H_I(z)$. Com isto teremos

$$\begin{aligned} G(z)H_I(z) &= H(z)H_I(z) = 1 \\ G(z)\hat{H}_I(z) &= H(z)\hat{H}_I(z). \end{aligned}$$

Com as modificações indicadas na Figura 4.1-(B) e as funções acima, chegamos à configuração de identificação inversa com entrada branca de (C) que é igual à da Figura 2.3. Além disto, para o caso de configuração de identificação inversa com entrada não-branca dada por $u(n) = G'(z)w(n)$, basta tomarmos

$$G(z) = H(z)G'(z),$$

ao invés de 4.1. Podemos considerar então a identificação inversa (de mínima fase) um caso particular da configuração de identificação com entrada não-branca, o que nos permite analisar os dois casos de maneira conjunta.

4.2 Cálculo dos zeros ótimos

A condição para otimização dos zeros, utilizando uma realização na forma direta é que

$$\frac{\partial E[e^2(n)]}{\partial b_k} = 0.$$

Figura 4.1: (A) - Caso particular da configuração de identificação com entrada não-branca dada por $u(n) = H(z)w(n)$ com sistema sendo identificado igual a $H_I(z) = 1/H(z)$. (B) - Passo intermediário para a obtenção da configuração de identificação inversa. (C) - Configuração de identificação inversa.

A partir de (2.55) temos

$$\frac{\partial E[e^2(n)]}{\partial b_k} = -2 \left\langle \frac{\partial \hat{H}(z)}{\partial b_k}, S_u(z) (H(z) - \hat{H}(z)) \right\rangle = 0,$$

que, junto com (2.51) e usando a Propriedade 2.2, nos leva a

$$\left\langle z^k \frac{G(z)}{A(z)}, G(z) (H(z) - \hat{H}(z)) \right\rangle = 0,$$

ou seja,

$$\left\langle z^k \frac{G(z)}{A(z)}, G(z) \frac{B(z)}{A(z)} \right\rangle = \left\langle z^k \frac{G(z)}{A(z)}, G(z) H(z) \right\rangle$$

e finalmente,

$$\left\langle z^k \frac{G(z)G(z^{-1})}{A(z)A(z^{-1})}, B(z) \right\rangle = \left\langle z^k \frac{G(z)}{A(z)}, G(z) H(z) \right\rangle \triangleq p_k,$$

para $k = 0, 1, \dots, M$. Escrevendo

$$\frac{G(z)G(z^{-1})}{A(z)A(z^{-1})} = T(z) + T(z^{-1})$$

com $T(z) = t_0 + t_1 z + t_2 z^2 + \dots$, e agrupando os coeficientes p_k no vetor $\mathbf{p} = [p_0 \ p_1 \ \dots \ p_M]^t$, estas condições podem ser escritas na forma matricial como

$$\mathbf{T} \mathbf{b}_o = \mathbf{p},$$

onde $\mathbf{T}$ é uma matriz Toeplitz com a primeira linha dada por $[2t_0 \ t_1 \ \dots \ t_M]$.

Portanto, o vetor $\mathbf{b}_o = [b_0 \ b_1 \ \dots \ b_M]^t$ será dado por

$$\mathbf{b}_o = \mathbf{T}^{-1} \mathbf{p}. \quad (4.2)$$

Esta equação pode ser aplicada a cada iteração da identificação com entrada não-branca do sistema passa-tudo da Seção 3.1 (Tabela 3.1, Figuras 3.2 e 3.3), sendo o resultado mostrado na Figura 4.2, junto com os valores obtidos na adaptação. Vemos que estes valores se encontram bem próximos dos valores ótimos dados por (4.2), o que nos permite afirmar que a adaptação na Figura 3.3 é lenta devido aos valores de a_k obtidos, e não aos valores dos parâmetros dos zeros, situação similar à observada em [23],[24] para entrada branca. Isto sugere que uma abordagem analítica também similar à do caso de entrada branca possa ser adequada. Tal abordagem é desenvolvida a seguir.

Figura 4.2: Parâmetros $b_k(n)$ (linhas sólidas) e seus valores ótimos $b_{ok}(n)$ (quadrados e linhas pontilhadas), deslocados como indicado para melhor visualização.

4.3 Base para a decomposição do erro quadrático

Para um melhor entendimento das características do erro quadrático médio no caso de entrada não-branca faremos uma decomposição do erro quadrático nos moldes de (2.39). Vemos em (2.48) que para o caso de entrada colorida

$$E[e^2(n)] = \left\langle H(z) - \hat{H}(z), S_u(z) \left(H(z) - \hat{H}(z) \right) \right\rangle.$$

Admitindo que $u(n) = G(z)w(n)$, onde $w(n)$ é branco com variância unitária, $S_u(z)$ é dado por (2.51), o que nos permite reescrever a equação acima como

$$E[e^2(n)] = \left\langle G(z) \left(H(z) - \hat{H}(z) \right), G(z) \left(H(z) - \hat{H}(z) \right) \right\rangle.$$

Logo, precisamos de uma decomposição apropriada para as funções $G(z)H(z)$ e $G(z)\hat{H}(z)$.

Se $G(z) \in \mathcal{H}_2$ for da forma $G(z) = E(z)/F(z)$ com $\deg[G(z)] = L$, temos que uma extensão natural de (2.31) é uma base ortonormal completa para $\mathcal{H}_2$ do tipo

$$\left\{ \frac{P_k(z)}{A(z)F(z)} \right\}_{k=0}^{M+L} \cup \left\{ z^k \frac{\bar{A}(z)\bar{F}(z)}{A(z)F(z)} \right\}_{k=1}^{\infty}, \quad (4.3)$$

onde $\deg[P_k(z)] \leq M + L$ e $\overline{F}(z) = z^L F(z^{-1})$. Como $G(z)H(z)$ e $G(z)\hat{H}(z)$ pertencem a $\mathcal{H}_2$, podemos escrever

$$\begin{aligned} G(z)\hat{H}(z) &= \sum_{k=0}^{M+L} \nu_k \frac{P_k(z)}{A(z)F(z)} \\ G(z)H(z) &= \sum_{k=0}^{M+L} \rho_k \frac{P_k(z)}{A(z)F(z)} + \\ &\quad + \sum_{k=1}^{\infty} \tau_k z^k V(z) \frac{\overline{F}(z)}{F(z)}, \end{aligned}$$

com

$$\begin{aligned} \nu_k &= \left\langle G(z)\hat{H}(z), \frac{P_k(z)}{A(z)F(z)} \right\rangle \\ \rho_k &= \left\langle G(z)H(z), \frac{P_k(z)}{A(z)F(z)} \right\rangle \\ \tau_k &= \left\langle G(z)H(z), z^k V(z) \frac{\overline{F}(z)}{F(z)} \right\rangle. \end{aligned} \quad (4.4)$$

O erro quadrático

$$\left\| H(z) - \hat{H}(z) \right\|_{S_u}^2 = \left\| G(z) \left(H(z) - \hat{H}(z) \right) \right\|^2$$

pode então ser decomposto como

$$\left\| H(z) - \hat{H}(z) \right\|_{S_u}^2 = \sum_{k=0}^{M+L} (\rho_k - \nu_k)^2 + \sum_{k=1}^{\infty} \tau_k^2, \quad (4.5)$$

que tem uma forma semelhante a (2.39).

No caso particular da identificação inversa de $H_I(z)$, temos, como visto anteriormente, que a função sendo identificada é da forma $H_I(z) = 1/H(z)$, o filtro $G(z)$ é dado por $G(z) = H(z)$ e, portanto, $G(z)H_I(z) = 1$. Isto, usando (4.4), resulta em

$$\tau_k = \left\langle 1, z^k V(z) \frac{\overline{F}(z)}{F(z)} \right\rangle \equiv 0,$$

pois $z^k V(z) \overline{F}(z)/F(z)$ é estritamente causal. Apesar da simplicidade desta equação, a continuação da análise a partir deste ponto é mais difícil. É particularmente complicada a obtenção de uma expressão tratável para os parâmetros α_k dos pólos, como em (2.64), obtida para o caso de identificação com entrada branca. Sem

esta expressão não é possível determinar uma condição para uma maior velocidade de convergência para a configuração de identificação inversa. Como estamos analisando conjuntamente este caso e o de identificação com entrada não-branca, uma decomposição mais adequada se faz necessária.

Generalizando o procedimento seguido em [19] e [20], consideraremos a decomposição dos produtos de $H(z)$ e $\hat{H}(z)$ pela componente de fase máxima de $G(z)$, dada por $\overline{E}(z)/F(z) = G(z)\overline{E}(z)/E(z)$, produtos que também pertencem a $\mathcal{H}_2$:

$$\frac{\overline{E}(z)}{F(z)}\hat{H}(z) = \sum_{k=0}^{M+L} \overline{\nu}_k \frac{P_k(z)}{A(z)F(z)} \quad (4.6)$$

$$\begin{aligned} \frac{\overline{E}(z)}{F(z)}H(z) &= \sum_{k=0}^{M+L} \overline{\rho}_k \frac{P_k(z)}{A(z)F(z)} + \\ &+ \sum_{k=1}^{\infty} \overline{\tau}_k z^k V(z) \frac{\overline{F}(z)}{F(z)}, \end{aligned} \quad (4.7)$$

onde $\overline{E}(z) = z^L E(z^{-1})$. Este procedimento não altera o erro quadrático médio obtido com $G(z)$ de fase mínima, pois, como $\overline{E}(z)/E(z)$ é passa-tudo,

$$\begin{aligned} \|H(z) - \hat{H}(z)\|_{S_u}^2 &= \|G(z) [H(z) - \hat{H}(z)]\|^2 \\ &= \left\| \frac{\overline{E}(z)}{E(z)} G(z) [H(z) - \hat{H}(z)] \right\|^2 \\ &= \left\| \frac{\overline{E}(z)}{F(z)} [H(z) - \hat{H}(z)] \right\|^2. \end{aligned}$$

Com (4.6) e (4.7) podemos escrever então

$$\|H(z) - \hat{H}(z)\|_{S_u}^2 = \sum_{k=0}^{M+L} (\overline{\rho}_k - \overline{\nu}_k)^2 + \sum_{k=1}^{\infty} \overline{\tau}_k^2,$$

onde os coeficientes $\overline{\tau}_k$ são dados por

$$\begin{aligned} \overline{\tau}_k &= \left\langle \frac{\overline{E}(z)}{F(z)} H(z), z^k V(z) \frac{\overline{F}(z)}{F(z)} \right\rangle \\ &= \left\langle \frac{z^{-L} F(z) z^L E(z^{-1})}{F(z^{-1}) F(z)} H(z), z^k V(z) \right\rangle \\ &= \left\langle \frac{E(z^{-1})}{F(z^{-1})} H(z), z^k V(z) \right\rangle \\ &= \langle G(z^{-1}) H(z), z^k V(z) \rangle, \end{aligned}$$

o que, considerando que $z^k V(z)$ é estritamente causal para $k > 0$, nos permite enunciar:

Propriedade 4.1 *Com um sinal de entrada de espectro racional, o erro quadrático médio na identificação de $H(z)$ pode ser escrito como*

$$\left\| H(z) - \hat{H}(z) \right\|_{S_u}^2 = \sum_{k=0}^{M+L} (\bar{\rho}_k - \bar{\nu}_k)^2 + \sum_{k=1}^{\infty} \langle H_p(z), z^k V(z) \rangle^2 \quad (4.8)$$

onde $H_p(z) = [G(z^{-1})H(z)]_{\oplus}$.

Notamos que os coeficientes $\bar{\tau}_k = \langle H_p(z), z^k V(z) \rangle$ possuem uma forma similar à dos coeficientes τ_k do caso de entrada branca (equação (2.39)), somente substituindo $H(z)$ por $H_p(z)$. Além disto a equação (4.8) é similar a (2.39), só que agora com $M + L$ termos na primeira parte desta decomposição, ao invés de M termos no caso de entrada branca.

No caso de entrada branca, a hipótese de que a adaptação dos coeficientes de $\hat{H}(z)$ se dava próxima à superfície de erro reduzida levou ao modelo de adaptação (2.67). Partindo então de (4.8), podemos seguir um caminho semelhante.

4.4 Velocidade de convergência determinada por $H_p(z)$

Partindo da Propriedade 4.1, se os pontos do espaço de parâmetros em que, para todos os parâmetros adaptados w_j , valer

$$\frac{\partial}{\partial w_j} \sum_{k=0}^{M+L} (\bar{\rho}_k - \bar{\nu}_k)^2 = 0 \quad (4.9)$$

formarem uma variedade (*manifold*) e se a adaptação naturalmente se restringir a esta variedade, então segue de (4.8) que todas as propriedades da velocidade de convergência de um algoritmo de gradiente exato (determinístico) obtidas para o caso de entrada branca serão válidas para a entrada colorida em questão, com $H_p(z)$ fazendo o papel de $H(z)$. Isto também aconteceria se a adaptação fosse *forçada* a se restringir àquela variedade, o que teria motivação prática, por exemplo, nos casos em que $H_p(z)$, através da segunda parte da decomposição (4.8),

garantissem uma convergência rápida. Por outro lado, se os pontos em questão não formarem uma variedade (por exemplo, se forem um conjunto enumerável de pontos), então sua relevância para o processo de adaptação provavelmente será menor.

Até o momento não conhecemos uma condição necessária e suficiente para que (4.9) seja válida. Consideraremos a seguir, então, uma condição que é suficiente (porém, possivelmente não necessária) para a validade de (4.9) e analisaremos a natureza do conjunto de pontos no espaço de parâmetros que atendem esta condição.

4.5 A superfície de erro restrita

Vemos que, se

$$\bar{\rho}_k = \bar{\nu}_k, \quad k = 0, 1, \dots, M + L, \quad (4.10)$$

então a condição expressa em (4.9) é satisfeita. Neste caso chamaremos de *superfície de erro restrita* o erro

$$\left\| H(z) - \hat{H}(z) \right\|_{S_u}^2 = \sum_{k=1}^{\infty} \langle H_p(z), z^k V(z) \rangle^2$$

como função dos parâmetros dos pólos. A seguir a caracterizaremos em mais detalhes.

4.5.1 Forma da superfície de erro restrita

Inicialmente, de (4.6) e (4.7), se valer (4.10), então

$$\frac{\bar{E}(z)}{F(z)} [H(z) - \hat{H}(z)] = zQ(z)V(z)\frac{\bar{F}(z)}{F(z)} \quad (4.11)$$

para algum $Q(z) \in \mathcal{H}_2$. Nota-se que a superfície de erro restrita está contida na superfície de erro reduzida, e, quando $\deg[H(z)] = M$, é composta ao menos pelo ponto $\hat{H}(z) = H(z)$.

Supondo que $H(z) = C(z)/D(z)$, $\hat{H}(z) = B(z)/A(z)$ e $V(z) = \bar{A}(z)/A(z)$, segue que $Q(z)$ é da forma $Q(z) = R(z)/D(z)$, para algum $R(z)$. Se reescrevermos

$\overline{E}(z)$ como

$$\overline{E}(z) = z^l \overline{E}'(z), \quad \text{onde } l = \begin{cases} 0, & \text{se } \deg[E(z)] = L \\ 1, & \text{se } \deg[E(z)] < L \end{cases},$$

então podemos reescrever (4.11) como

$$z^l \overline{E}'(z) C(z) A(z) - z^l \overline{E}'(z) B(z) D(z) = z R(z) \overline{F}(z) \overline{A}(z). \quad (4.12)$$

Supondo que $\overline{A}(z)$ e $\overline{E}'(z)$ tenham exatamente K raízes em comum, podemos escrever

$$\begin{aligned} \overline{A}(z) &= z^M + a_1 z^{M-1} + \dots + a_M \\ &= \underbrace{(z^K + p_1 z^{K-1} + \dots + p_K)}_{\triangleq \overline{P}(z)} \times \underbrace{(z^{M-K} + a'_1 z^{M-K-1} + \dots + a'_{M-K})}_{\triangleq \overline{A}'(z)} \\ &= \overline{P}(z) \overline{A}'(z). \end{aligned} \quad (4.13)$$

De (4.13) e de $\overline{A}(z) = z^M A(z)$, segue

$$\begin{aligned} A(z) &= z^M \overline{P}(z^{-1}) \overline{A}'(z^{-1}) \\ &= (1 + p_1 z + \dots + p_K z^K) \times (1 + a'_1 z + \dots + a'_{M-K} z^{M-K}) \\ &= P(z) A'(z). \end{aligned} \quad (4.14)$$

De maneira similar temos $\overline{E}'(z) = \overline{P}(z) \overline{E}''(z)$ e $E(z) = P(z) E''(z)$.

Usando as equações acima em (4.12), temos

$$z^l \overline{E}''(z) C(z) P(z) A'(z) - z^l \overline{E}''(z) B(z) D(z) = z R(z) \overline{F}(z) \overline{A}'(z) \quad (4.15)$$

e como nenhuma das raízes de $\overline{E}''(z)$ pode ser compartilhada por $\overline{A}'(z)$ (por construção), nem por $\overline{F}(z)$ (o que contrariaria a hipótese adotada de que $\deg[G(z)] = L$), decorre que $R(z)$ é da forma $T(z) \overline{E}''(z)$, fazendo com que (4.15) simplifique ainda mais para

$$C(z) P(z) A'(z) - B(z) D(z) = z^{1-l} T(z) \overline{F}(z) \overline{A}'(z). \quad (4.16)$$

Como o grau do polinômio do lado esquerdo da equação é menor ou igual a $M+N$, e o do lado direito não ultrapassa $M - K + L + 1 - l + \deg[T(z)]$, temos que

$$\deg[T(z)] \leq N + K - L - 1 + l. \quad (4.17)$$

A equação (4.16) pode ser escrita em forma matricial como

$$\Theta_{CP} \begin{bmatrix} 1 \\ \mathbf{a}' \end{bmatrix} - \Theta_D \mathbf{b} = \Theta_{T\bar{F}} \Pi \begin{bmatrix} 1 \\ \mathbf{a}' \end{bmatrix} \quad (4.18)$$

onde Θ_{CP} e Θ_D são as matrizes de convolução de dimensão $(M + N + 1) \times (M - K + 1)$ formadas a partir dos coeficientes de $C(z)P(z)$ e de $D(z)$, respectivamente, $\Theta_{T\bar{F}}$ é a matriz de convolução de dimensão $(M + N + 1) \times (M - K + 1)$ formada a partir dos coeficientes de $z^{1-l}T(z)\bar{F}(z)$, Π é a matriz antidiagonal de permutação, e $\mathbf{a}' = [a'_1 \ a'_2 \ \dots \ a'_{M-K}]$ e $\mathbf{b}$ são os vetores formados a partir dos coeficientes de $A'(z) = 1 + a'_1 z + \dots + a'_{M-K} z^{M-K}$ e $B(z)$, respectivamente. Separando a primeira coluna de Θ_{CP} e $\Theta_{T\bar{F}}$, podemos escrever $\Theta_{CP} = [\boldsymbol{\theta}_{CP} \ \Phi_{CP}]$ e $\Theta_{T\bar{F}} \Pi = [\boldsymbol{\theta}_{T\bar{F}} \ \Phi_{T\bar{F}}]$, o que nos leva ao seguinte conjunto de equações:

$$\begin{bmatrix} -\Theta_D & \Phi_{CP} - \Phi_{T\bar{F}} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{a}' \end{bmatrix} = \boldsymbol{\theta}_{T\bar{F}} - \boldsymbol{\theta}_{CP}. \quad (4.19)$$

Para cada $T(z)$, (4.19) fornece um sistema de $M + N + 1$ equações e $2M - K + 1$ incógnitas. Então, se tivermos

$$M \geq N + K,$$

as soluções do sistema formam um espaço linear de dimensão $M - N - K$. Assumindo ainda, a partir de (4.17), que $T(z)$ têm

$$\deg[T(z)] + 1 = N + K - L + l > 0$$

graus de liberdade, as soluções para o problema original (4.10), ou em outras palavras, o domínio da superfície de erro restrita, é um conjunto de variedades, cada uma com dimensão

$$(M - N - K) + (N + K - L + l) = M - L + l. \quad (4.20)$$

Cada variedade é associada a um conjunto de zeros compartilhados por $\bar{A}(z)$ e $\bar{E}'(z)$.

Como discutido na Seção 2.5.6, adotar $M = N + L$ tem a vantagem de eliminar a possibilidade de mínimos locais. Neste caso, com (4.20), a dimensão das variedades acima será $M - L + l = N + l$.

Figura 4.3: Pólos (“o”) e zeros (“x”) de $H(z^{-1})$ (esquerda) e $H_p(z^{-1})$ (direita).

4.5.2 Exemplos Numéricos

Aplicaremos as idéias discutidas até aqui na identificação, com entrada não-branca, do sistema $H(z)$ passa-tudo da Tabela 3.1. Os pólos e zeros das funções $H(z)$ e $H_p(z)$ estão na Figura 4.3. Vemos que o efeito de $G(z)$ neste caso é o de fazer com que um par de zeros de $H_p(z)$ seja quase cancelado por um par de seus pólos.

Na Figura 4.4 são apresentados o erro médio quadrático $E[e^2(n)]$ para o algoritmo de gradiente exato, usando os mesmos passos para a adaptação que o algoritmo de gradiente estocástico da Tabela 3.1. Vemos que é uma boa aproximação do erro quadrático $e^2(n)$ do algoritmo estocástico. Além disto, como também mostrado na Figura 4.4, a parcela $\sum_{k=0}^{M+L} (\bar{\rho}_k - \bar{\nu}_k)^2$ do erro quadrático médio $E[e^2(n)]$ é bem menor que o erro total (mais de 20dB, ou seja, menos de 1%), de onde concluímos que a adaptação se dá próxima à superfície de erro restrita. Como discutido anteriormente, isto sugere que o modelo (2.67) é uma boa aproximação do processo de adaptação, com $H_p(z)$ no lugar de $H(z)$.

De fato, como mostrado nas figuras (4.5) e (4.6), o modelo (2.67) explica a baixa velocidade de convergência observada: o espalhamento dos autovalores de $\mathbf{J}\mathbf{J}^t\mathbf{\Sigma}^2$ é grande e o vetor $\mathbf{\alpha}(n)$, após um certo número de iterações, está principalmente nas direções dos autovetores associados aos menores autovalores de $\mathbf{J}\mathbf{J}^t\mathbf{\Sigma}^2$. Isto sugere que podemos estabelecer condições que levem a uma convergência mais rápida, o que será feito adiante na Seção 4.7.

Figura 4.4: $E[e^2(n)]$ para o algoritmo de gradiente exato e $e^2(n)$ para o algoritmo de gradiente estocástico (traços superiores), parcela $\sum_{k=0}^{M+L} (\bar{\rho}_k - \bar{\nu}_k)^2$ de $E[e^2(n)]$ (traço inferior).

Figura 4.5: Evolução dos autovalores $\lambda_k(n)$ de $\mathbf{J}\mathbf{J}^t \Sigma^2$ associada a $H_p(z)$

Figura 4.6: Evolução dos componentes $\beta_k(n)$ de $\alpha(n)$ nas direções dos autovetores associados aos autovalores $\lambda_k(n)$.

4.6 Propriedades do sistema projetado $H_p(z)$

Na seção anterior, o sistema projetado $H_p(z)$ se mostrou útil na explicação da baixa velocidade de convergência observada. Nesta seção discutiremos algumas propriedades do sistema projetado $H_p(z)$ para termos um melhor entendimento da influência do espectro da entrada nas propriedades de convergência de filtros adaptativos IIR. Iniciaremos mostrando que a forma de Hankel associada ao sistema $H_p(z)$ é uma versão ponderada da forma de Hankel de $H(z)$. Depois analisaremos a relação entre as normas de $G(z)H(z)$ e $H_p(z)$, o que terá consequências relevantes no passo de adaptação máximo do algoritmo adaptativo.

4.6.1 Forma de Hankel ponderada

Mostramos inicialmente que o sistema projetado $H_p(z)$ está relacionado com a forma de Hankel de $H(z)$ ponderada, definida em [5].

Com $H(z) = \sum_{k=0}^{\infty} h_k z^k$, $G(z^{-1}) = \sum_{l=0}^{\infty} g_l z^{-l}$ e $H_p(z) = \sum_{k=0}^{\infty} h_{p,k} z^k$, temos

$$\begin{aligned} \sum_{k=0}^{\infty} h_{p,k} z^k &= \left[\left(\sum_{l=0}^{\infty} g_l z^{-l} \right) \left(\sum_{m=0}^{\infty} h_m z^m \right) \right]_{\oplus} \\ &= \left[\sum_{l=0}^{\infty} \sum_{m=0}^{\infty} g_l h_m z^{m-l} \right]_{\oplus} \\ &= \left[\sum_{l=0}^{\infty} \sum_{k=-l}^{\infty} g_l h_{k+l} z^k \right]_{\oplus} \\ &= \sum_{l=0}^{\infty} \sum_{k=0}^{\infty} g_l h_{k+l} z^k. \end{aligned}$$

Os coeficientes $h_{p,k}$ são então dados por

$$h_{p,k} = \sum_{l=0}^{\infty} g_l h_{k+l}. \quad (4.21)$$

Com isto podemos escrever

$$\mathbf{h}_p^+ = \mathbf{\Gamma}_H \mathbf{\varsigma}, \quad (4.22)$$

onde $\mathbf{\Gamma}_H$ é a forma de Hankel de $H(z)$,

$$\mathbf{h}_p^+ = \begin{bmatrix} h_{p,1} \\ h_{p,2} \\ h_{p,3} \\ \vdots \end{bmatrix}$$

é um vetor cujos elementos são os coeficientes $h_{p,k}$ para $k > 0$ e

$$\mathbf{\varsigma} = \begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ \vdots \end{bmatrix}.$$

Utilizando (4.22) vemos que a forma de Hankel de $H_p(z)$ será dada por

$$\mathbf{\Gamma}_{H_p} = \begin{bmatrix} h_{p,1} & h_{p,2} & h_{p,3} & \cdots \\ h_{p,2} & h_{p,3} & h_{p,4} & \cdots \\ h_{p,3} & h_{p,4} & h_{p,5} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} = \mathbf{\Gamma}_H \mathbf{\Theta}_G$$

com

$$\Theta_G = \begin{bmatrix} g_0 & 0 & 0 & \cdots \\ g_1 & g_0 & 0 & \cdots \\ g_2 & g_1 & g_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

sendo a matriz de convolução formada a partir dos coeficientes g_k . Portanto, esta forma de Hankel é igual à forma de Hankel de $H(z)$ ponderada por Θ_G definida em [5].

Considerando que $G(z)$ é de fase mínima, temos que a matriz Θ_G admite uma matriz inversa

$$\Theta_Q = \begin{bmatrix} q_0 & 0 & 0 & \cdots \\ q_1 & q_0 & 0 & \cdots \\ q_2 & q_1 & q_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

formada a partir dos coeficientes da função causal

$$Q(z) = \frac{1}{G(z)} = \sum_{k=0}^{\infty} q_k z^k.$$

Neste caso, se Γ_H tiver posto N , Γ_{H_p} também terá posto N , pois Θ_G é uma transformação inversível. Pelo teorema de Kronecker segue que $H_p(z)$ tem ordem N , igual à ordem de $H(z)$.

4.6.2 Norma de $G(z)H(z)$ e relação com os passos de adaptação

Nesta seção vamos verificar uma propriedade da norma de $G(z)H(z)$, grandeza que terá uma relação com o valor máximo permitido para os passos de adaptação, como discutido adiante na Seção 4.7.

Começaremos reescrevendo 4.21 de uma maneira que inclua o termo direto h_0

como

$$\mathbf{h}_p = \begin{bmatrix} h_{p,0} \\ h_{p,1} \\ h_{p,2} \\ \vdots \end{bmatrix} = \underbrace{\begin{bmatrix} g_0 & g_1 & g_2 & \cdots \\ 0 & g_0 & g_1 & \cdots \\ 0 & 0 & g_0 & \cdots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix}}_{\mathbf{\Theta}_G^t} \underbrace{\begin{bmatrix} h_0 \\ h_1 \\ h_2 \\ \vdots \end{bmatrix}}_{\mathbf{h}}, \quad (4.23)$$

onde $\mathbf{h}$ contém os coeficientes de $H(z) = \sum_{k=0}^{\infty} h_k z^k$.

Com (4.23) podemos escrever

$$\|H_p(z)\|^2 = \sum_{k=0}^{\infty} h_{p,k}^2 = \mathbf{h}_p^t \mathbf{h}_p = \mathbf{h}^t \mathbf{\Theta}_G \mathbf{\Theta}_G^t \mathbf{h}. \quad (4.24)$$

Por sua vez, com $S_u(z) = \sum_{k=-\infty}^{\infty} r_{uu,k} z^k$ e $r_{uu,k} = r_{uu,-k}$, o quadrado da norma L_2 de $H(z)$ pode ser escrita como

$$\langle H(z), H(z) \rangle_{S_u} = \mathbf{h}^t \underbrace{\begin{bmatrix} r_{uu,0} & r_{uu,1} & r_{uu,2} & \cdots \\ r_{uu,1} & r_{uu,0} & r_{uu,1} & \ddots \\ r_{uu,2} & r_{uu,1} & r_{uu,0} & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix}}_{=\mathbf{\Theta}_G^t \mathbf{\Theta}_G} \mathbf{h}. \quad (4.25)$$

Subtraindo (4.24) de (4.25) temos

$$\langle H(z), H(z) \rangle_{S_u} - \|H_p(z)\|^2 = \mathbf{h}^t (\mathbf{\Theta}_G^t \mathbf{\Theta}_G - \mathbf{\Theta}_G \mathbf{\Theta}_G^t) \mathbf{h}. \quad (4.26)$$

O fator central do lado direito desta equação aparece ao se subtrair duas maneiras diferentes de elevar $\mathbf{\Theta}_G$ ao quadrado. Esta diferença é positiva semi-definida, como a seguinte propriedade mostra [5]:

Propriedade 4.2 *Seja*

$$\mathbf{\Gamma}_G = \begin{bmatrix} g_1 & g_2 & g_3 & \cdots \\ g_2 & g_3 & g_4 & \cdots \\ g_3 & g_4 & g_5 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

a forma de Hankel de $G(z)$. Então

$$\mathbf{\Theta}_G^t \mathbf{\Theta}_G - \mathbf{\Theta}_G \mathbf{\Theta}_G^t = \mathbf{\Gamma}_G^2.$$

Para ver que esta propriedade é verdadeira basta calcular os elementos das matrizes de cada lado da equação. Do lado direito, cada elemento $\gamma_{k,l}$ de $\mathbf{\Gamma}_G^2$ é da forma

$$\gamma_{k,l} = \sum_{i=0}^{\infty} g_{i+k}g_{i+l}, \quad k, l > 0.$$

Do lado esquerdo, com $g_k = 0$, $k < 0$, cada elemento é dado por

$$\begin{aligned} \gamma'_{k,l} &= \sum_{i=0}^{\infty} g_i g_{i+l-k} - \sum_{i=0}^{k-1} g_i g_{i+l-k} \\ &= \sum_{m=-k}^{\infty} g_{m+k} g_{m+l} - \sum_{m=-k}^{-1} g_{m+k} g_{m+l}, \quad m = i - k \\ &= \sum_{m=0}^{\infty} g_{m+k} g_{m+l} = \gamma_{k,l}. \end{aligned}$$

Retornando à comparação (4.26), temos então

$$\langle H(z), H(z) \rangle_{S_u} - \|H_p(z)\|^2 = \mathbf{h}^t \mathbf{\Gamma}_G^2 \mathbf{h} \geq 0, \quad (4.27)$$

com a igualdade ocorrendo apenas se $\mathbf{h}$ estiver no espaço nulo de $\mathbf{\Gamma}_G$. Notando que

$$\langle H(z), H(z) \rangle_{S_u} = \langle G(z)H(z), G(z)H(z) \rangle,$$

vemos que com o resultado da equação (4.27) temos que a norma de $G(z)H(z)$ será sempre maior ou igual à norma do sistema projetado $H_p(z)$.

Como exemplo, sejam as funções utilizadas da Tabela 3.1 e da Figura 3.2. Os pólos e zeros destas funções se encontram reproduzidos na Tabela 4.1 junto com os do sistema projetado $H_p(z)$ correspondente. Vemos que a norma de $H_p(z)$ é cerca de 0,77 vezes a norma de $H(z)$ ponderada por $S_u(z) = G(z^{-1})G(z)$.

Esta diferença nas normas ocorre mesmo quando o sistema $H(z)$ sendo identificado propicia melhores propriedades de convergência no caso de entrada não-branca. Observando o que ocorre com o sistema da Tabela 3.2 com seus parâmetros reproduzidos na Tabela 4.2 junto com os do sistema $H_p(z)$ correspondente, vemos que a norma de $H_p(z)$ é cerca de 0,6 vezes a norma de $G(z)H(z)$.

Em relação ao valor máximo admissível para μ_A , notamos que os sinais usados para a adaptação dos parâmetros a_k são os sinais (ver a Seção 2.7.1)

$$\nabla_{ak}(n) = -\frac{z^k}{A(z)} \hat{H}(z)u(n) = -\frac{z^k}{A(z)} G(z) \hat{H}(z)w(n).$$

Funções utilizadas		
Função	Pólos	Zeros
$H(z^{-1})$	$0,8\angle \pm 30^\circ$ e $0,9\angle \pm 100^\circ$	$1/0,8\angle \pm 30^\circ$ e $1/0,9\angle \pm 100^\circ$
$G(z^{-1})$	$0,7\angle \pm 110^\circ$	$0,75\angle \pm 30^\circ$
$H_p(z^{-1})$	$0,8\angle \pm 30^\circ$ e $0,9\angle \pm 100^\circ$	$1,01\angle \pm 82,15^\circ$ e $0,87\angle \pm 26,3^\circ$
Normas		
$\ H(z)\ ^2$	1,0	
$\ G(z)H(z)\ ^2$	1,0	
$\ H_p(z)\ ^2$	0,77	

Tabela 4.1: Pólos e zeros das funções utilizadas no exemplo da Figura 3.2 junto com suas respectivas normas.

Funções utilizadas		
Função	Pólos	Zeros
$H(z^{-1})$	$0,8\angle \pm 30^\circ$ e $0,9\angle \pm 100^\circ$	$1,6\angle \pm 55^\circ$ e $0,9\angle \pm 115^\circ$
$G(z^{-1})$	$0,7\angle \pm 110^\circ$	$0,75\angle \pm 30^\circ$
$H_p(z^{-1})$	$0,8\angle \pm 30^\circ$ e $0,9\angle \pm 100^\circ$	$1,41\angle \pm 120,6^\circ$ e $1,01\angle \pm 29,3^\circ$
Normas		
$\ H(z)\ ^2$	1,0	
$\ G(z)H(z)\ ^2$	0,13	
$\ H_p(z)\ ^2$	0,08	

Tabela 4.2: Pólos e zeros das funções utilizadas no exemplo da Figura 3.5 junto com suas respectivas normas.

Perto do ponto de ótimo temos $\hat{H}(z) \approx H(z)$ e resulta, portanto, que um aumento da norma $\|G(z)H(z)\|$ tende a fazer cair o valor máximo de μ_A . Portanto, como $\|G(z)H(z)\| \geq \|H_p(z)\|$, o passo de adaptação μ_A máximo tenderá a ser menor do que no caso da identificação de $H_p(z)$ com entrada branca.

4.7 Condição para pequeno espalhamento dos autovalores de $\mathbf{J}\mathbf{J}^t\Sigma^2$

No modelo da adaptação dado por (2.67), se os valores singulares de Hankel de $H_p(z)$ forem todos iguais então o espalhamento dos autovalores de $\mathbf{J}\mathbf{J}^t\Sigma^2$ tenderá a diminuir, aumentando a velocidade de convergência. A seguir, determinaremos uma condição para esta igualdade.

4.7.1 Igualdade dos valores singulares

Notando que, com $H(z) = C(z)/D(z)$ e com $G(z^{-1})$ sendo anti-causal, $H_p(z)$ tem a forma

$$H_p(z) \triangleq [G(z^{-1})H(z)]_{\oplus} = \frac{S(z)}{D(z)}$$

com $\deg[S(z)] \leq N$. Então, se

$$S(z) = s [\overline{D}(z) + cD(z)] \quad (4.28)$$

para qualquer c , todos os autovalores da forma de Hankel de $H_p(z)$ serão iguais a s .

Para encontrarmos um $C_s(z)$ de forma que $H_s(z) = C_s(z)/D(z)$ leve a um $S(z)$ como em (4.28), começamos separando

$$\begin{aligned} G(z^{-1})H_s(z) &= \frac{E(z^{-1})C_s(z)}{F(z^{-1})D(z)} = \frac{z^L E(z^{-1})C_s(z)}{z^L F(z^{-1})D(z)} \\ &= \frac{\overline{E}(z)C_s(z)}{\overline{F}(z)D(z)} \end{aligned}$$

em suas parcelas causal e estritamente anti-causal

$$\frac{\overline{E}(z)C_s(z)}{\overline{F}(z)D(z)} = \frac{S(z)}{D(z)} + \frac{\overline{Q}(z)}{\overline{F}(z)}, \quad (4.29)$$

para algum $\overline{Q}(z) = z^L Q(z^{-1})$, onde $\deg[Q(z)] < L$. Substituindo (4.28) em (4.29), com alguma manipulação algébrica, chegamos a

$$C_s(z)\overline{E}(z) - D(z)\overline{Q}(z) = s [\overline{D}(z) + cD(z)] \overline{F}(z),$$

que pode ser escrita, em forma matricial, como

$$\begin{bmatrix} \Theta_{\overline{E}} & -\Theta_D \end{bmatrix} \begin{bmatrix} \mathbf{c} \\ \mathbf{q} \end{bmatrix} = \mathbf{p} \quad (4.30)$$

onde $\Theta_{\overline{E}}$ é a matriz de convolução de dimensões $(N + L + 1) \times (N + 1)$ formada a partir dos coeficientes de $\overline{E}(z)$, Θ_D é a matriz de convolução de dimensões $(N + L + 1) \times L$ formada a partir dos coeficientes de $D(z)$, $\mathbf{c}$ contém os $N + 1$ coeficientes de $C_s(z)$, $\mathbf{q}$ contém os L coeficientes de $\overline{Q}(z)$ e $\mathbf{p}$ contém os $N + L + 1$ coeficientes de $s[\overline{D}(z) + cD(z)]\overline{F}(z)$. A solução deste sistema fornece o polinômio $C_s(z)$ desejado.

4.7.2 Exemplos numéricos

4.7.2.1 Teste do efeito de $H_p(z)$ passa-tudo

Consideraremos agora as funções da Tabela 3.1, caso em que a identificação de um sistema passa-tudo com entrada não-branca era lenta. Para evitar ambiguidades, denominamos aqui este sistema específico de $H_1(z)$. Calculamos $C_s(z)$ de acordo com (4.30), com $s = 1$ e $c = 0$. Os diagramas de pólos e zeros de $H_1(z) = C(z)/D(z)$ e de $H_s(z) = C_s(z)/D(z)$ estão na Figura 4.7 e as respostas em frequência de $G(z)$ e $H_s(z)$ estão na Figura 4.8. Vemos que, neste caso, $|H_s(e^{j\omega})|$ têm o efeito de aproximadamente inverter $|G(e^{j\omega})|$.

O novo sistema a ser identificado $H_s(z)$, em combinação com a entrada colorida determinada por $G(z)$, leva a um sistema projetado $H_p(z)$ passa-tudo igual ao sistema original $H_1(z)$ considerado no exemplo. Portanto, conforme a análise anterior, se a adaptação ocorrer na superfície de erro restrita, seu comportamento deve ser similar ao da identificação de $H_1(z)$ com entrada branca e mais rápido que o da identificação de $H_1(z)$ com a entrada colorida determinada por $G(z)$.

Figure 4.7: Pólos (“o”) e zeros (“x”) de $H(z^{-1})$ (esquerda) e $H_s(z^{-1})$ (direita).

Figure 4.8: Módulo das respostas em frequência de $G(z)$ (linha sólida) e de $H_s(z)$ (linha tracejada).

Começamos considerando passos de adaptação $\mu_A = 0,001$ e $\mu_B = 0,04$ e condições iniciais $b_k = 0,4 \times c_{a,k}$ e $a_k = 0,4 \times d_k$. Os resultados para $H_1(z)$ com entrada branca e $H_s(z)$ com entrada colorida são mostrados nas figuras 4.9 a 4.11 e nas figuras 4.12 a 4.14, respectivamente. Notamos que, em ambos os casos, a parcela $\sum_{k=0}^{M+L} (\bar{\rho}_k - \bar{\nu}_k)^2$ é consideravelmente menor que $E[e^2(n)]$, ou seja, a adaptação ocorre próxima à superfície de erro restrita. Além disto, corroborando a análise teórica, o comportamento geral da adaptação nos dois casos é bastante similar.

O valor $\mu_B = 0,04$ é próximo do valor que faz, em ambos os casos, o algoritmo estocástico divergir e será mantido nas simulações seguintes. Devemos notar que fazer a adaptação dos zeros ser muito mais rápida do que a adaptação dos pólos (resultado desta escolha $\mu_B \gg \mu_A$) pode não ser suficiente, em outras situações, para fazer a adaptação ocorrer próxima à superfície restrita, uma vez que tanto os zeros quanto os pólos têm efeito na parcela $\sum_{k=0}^{M+L} (\bar{\rho}_k - \bar{\nu}_k)^2$.

Figura 4.9: Identificação de $H_1(z)$ com entrada branca: $E[e^2(n)]$ para o algoritmo de gradiente exato e $e^2(n)$ para o algoritmo de gradiente estocástico (traços superiores), parcela $\sum_{k=0}^{M+L} (\bar{\rho}_k - \bar{\nu}_k)^2$ de $E[e^2(n)]$ (traço inferior).

Figura 4.10: Identificação de $H_1(z)$ com entrada branca: autovalores $\lambda_k(n)$ de $\mathbf{J}\mathbf{J}^t\mathbf{\Sigma}^2$ associado a $H_p(z)$

Figura 4.11: Identificação de $H_1(z)$ com entrada branca: componentes $\beta_k(n)$ de $\boldsymbol{\alpha}(n)$ nas direções dos autovetores associados aos autovalores $\lambda_k(n)$.

Figura 4.12: Identificação de $H_s(z)$ com entrada colorida: $E[e^2(n)]$ para o algoritmo de gradiente exato e $e^2(n)$ para o algoritmo de gradiente estocástico (traços superiores), parcela $\sum_{k=0}^{M+L} (\bar{\rho}_k - \bar{\nu}_k)^2$ de $E[e^2(n)]$ (traço inferior).

Figura 4.13: Identificação de $H_s(z)$ com entrada colorida: autovalores $\lambda_k(n)$ de $\mathbf{J}\mathbf{J}^t \mathbf{\Sigma}^2$ associado a $H_p(z)$

Figura 4.14: Identificação de $H_s(z)$ com entrada colorida: componentes $\beta_k(n)$ de $\alpha(n)$ nas direções dos autovetores associados aos autovalores $\lambda_k(n)$.

Aumentemos agora μ_A para 0,005, mantendo as demais condições. Este valor de μ_A é próximo do valor que faz o algoritmo estocástico divergir no caso da identificação de $H_s(z)$ com entrada colorida. Os resultados são mostrados nas figuras 4.15 a 4.20. Notamos que, no segundo caso, a parcela $\sum_{k=0}^{M+L} (\bar{\rho}_k - \bar{\nu}_k)^2$ não é muito menor que $E[e^2(n)]$, ou seja, a adaptação não ocorre próxima à superfície de erro restrita. Podemos ver também que isto, neste caso, tem o efeito de retardar a adaptação, que é mais lenta do que a adaptação para $H_1(z)$ e entrada branca. Mesmo assim, a velocidade ainda é muito maior do que a velocidade para $H_1(z)$ passa-tudo e entrada colorida, o que mostra a utilidade da condição (4.30).

Figura 4.15: Identificação de $H_1(z)$ com entrada branca, $\mu_A = 0,005$: $E[e^2(n)]$ para o algoritmo de gradiente exato e $e^2(n)$ para o algoritmo de gradiente estocástico (traços superiores), parcela $\sum_{k=0}^{M+L} (\bar{\rho}_k - \bar{\nu}_k)^2$ de $E[e^2(n)]$ (traço inferior).

Figura 4.16: Identificação de $H_1(z)$ com entrada branca, $\mu_A = 0,005$: autovalores $\lambda_k(n)$ de $\mathbf{J}\mathbf{J}^t\mathbf{\Sigma}^2$ associado a $H_p(z)$

Figura 4.17: Identificação de $H_1(z)$ com entrada branca, $\mu_A = 0,005$: componentes $\beta_k(n)$ de $\boldsymbol{\alpha}(n)$ nas direções dos autovetores associados aos autovalores $\lambda_k(n)$.

Figura 4.18: Identificação de $H_s(z)$ com entrada colorida, $\mu_A = 0,005$: $E[e^2(n)]$ para o algoritmo de gradiente exato e $e^2(n)$ para o algoritmo de gradiente estocástico (traços superiores), parcela $\sum_{k=0}^{M+L} (\bar{\rho}_k - \bar{\nu}_k)^2$ de $E[e^2(n)]$ (traço inferior).

Figura 4.19: Identificação de $H_s(z)$ com entrada colorida, $\mu_A = 0,005$: autovalores $\lambda_k(n)$ de $\mathbf{J}\mathbf{J}^t \Sigma^2$ associado a $H_p(z)$

Figura 4.20: Identificação de $H_s(z)$ com entrada colorida, $\mu_A = 0,005$: componentes $\beta_k(n)$ de $\alpha(n)$ nas direções dos autovetores associados aos autovalores $\lambda_k(n)$.

Aumentemos finalmente μ_A para 0,02, mantendo outra vez as demais condições. Agora, este valor de μ_A é próximo do valor que faz o algoritmo estocástico divergir no caso da identificação de $H_1(z)$ com entrada branca. Os resultados são mostrados nas figuras 4.21 a 4.23, onde notamos que a adaptação não ocorre próxima à superfície de erro restrita.

Na Seção 4.6.2 mostramos que sempre vale $\|G(z)H(z)\| \geq \|H_p(z)\|$. Aplicando isto ao caso em questão ($H(z) = H_s(z)$ que leva $H_p(z) = [G(z^{-1})H_s(z)]_+$ a ser igual a $H_1(z)$, passa-tudo), resulta em

$$\|G(z)H_s(z)\| \geq \|H_1(z)\|,$$

o que sugere, como de fato verificado, que o valor máximo de μ_A na identificação de $H_s(z)$ com entrada colorida é menor do que na identificação de $H_1(z)$ com entrada branca.

Figura 4.21: Identificação de $H_1(z)$ com entrada branca, $\mu_A = 0,02$: $E[e^2(n)]$ para o algoritmo de gradiente exato e $e^2(n)$ para o algoritmo de gradiente estocástico (traços superiores), parcela $\sum_{k=0}^{M+L} (\bar{p}_k - \bar{v}_k)^2$ de $E[e^2(n)]$ (traço inferior).

Figura 4.22: Identificação de $H_1(z)$ com entrada branca, $\mu_A = 0,02$: autovalores $\lambda_k(n)$ de $\mathbf{J}\mathbf{J}^t\mathbf{\Sigma}^2$ associado a $H_p(z)$

4.7.2.2 Variação de $C_s(z)$ com a mudança dos pólos de $H_s(z)$

Como vemos na equação (4.30), $C_s(z)$ depende de $D(z)$ e de $G(z)$. Então, se mudarmos os pólos de $H_s(z)$ mudaremos o polinômio $C_s(z)$ correspondente. Na

Figura 4.23: Identificação de $H_1(z)$ com entrada branca, $\mu_A = 0,02$: componentes $\beta_k(n)$ de $\alpha(n)$ nas direções dos autovetores associados aos autovalores $\lambda_k(n)$.

Figura 4.24 temos a resposta em frequência de um novo sistema $H_s(z)$, agora com pólos em $0.8\angle \pm 120^\circ$ e $0.9\angle \pm 160^\circ$, junto com a resposta em frequência de $G(z)$. No exemplo anterior, tínhamos $H_s(z)$ aproximadamente igual ao inverso de $G(z)$. Agora, $H_s(z)$ já não apresenta esta característica tão acentuada, possuindo inclusive valores maiores do que o do exemplo anterior em toda a faixa de frequência.

Neste exemplo temos $\|G(z)H_s(z)\| = 2.2$ e portanto devemos esperar que os valores máximos para os passos de adaptação sejam menores que os do caso anterior. Simulações numéricas mostraram que o valor máximo dos passos de adaptação para este caso são $\mu_A = 0,002$ e $\mu_B = 0,01$ (Figura 4.25). No exemplo anterior tínhamos $H_1(z)$ com pólos em $0,8\angle \pm 30^\circ$ e $0,9\angle \pm 100^\circ$, a norma $\|G(z)H_s(z)\| = 1$ e o valores máximos dos passos de adaptação $\mu_A = 0,005$ e $\mu_B = 0,04$. Comparando os valores máximos dos passos de adaptação e a norma $\|G(z)H_s(z)\|$ vemos que os resultados obtidos se encontram coerentes com a análise feita na Seção 4.6.2.

Figure 4.24: Módulo das respostas em frequência de $G(z)$ (linha sólida) e de $H_s(z)$ (linha tracejada) com um novo conjunto de pólos.

Figura 4.25: Identificação de $H_s(z)$ com entrada não-branca: $e^2(n)$ para o algoritmo de gradiente estocástico.

4.8 Algoritmo de aproximações sucessivas com entrada não-branca

Os resultados das seções anteriores nos permitem modificar o algoritmo de aproximações sucessivas da Seção 2.9 de modo que haja uma convergência mais rápida em filtros adaptativos submetidos a uma entrada de espectro racional, o que é feito a seguir. O caso de identificação com entrada branca e identificação inversa passam a ser casos particulares deste algoritmo, como será mostrado. Concluiremos com exemplos numéricos.

4.8.1 Descrição do algoritmo

A idéia do algoritmo é a de fazer com que o filtro adaptativo enxergue uma função que seja mais fácil de se identificar. Desta forma, se anteriormente o que se desejava era que o sistema $H(z)$ se parecesse com uma função passa-tudo de modo a possuir boas características de velocidade de convergência, agora o que queremos é que $H_p(z)$ tenha esta característica, sendo então o algoritmo modificado da seguinte maneira:

1. A entrada é considerada da forma $u(n) = G(z)w(n)$ com $w(n)$ branco e de variância unitária. Com isto, $S_u = G(z^{-1})G(z)$, e a partir de $u(n)$ obtêm-se uma estimativa da função $G(z)$ de mínima fase.
2. A partir de uma estimativa $\hat{H}(z) = B(z)/A(z)$ do sistema original $H(z)$ constrói-se a função

$$H_a(z) = \hat{H}(z) - H_s(z)$$

onde $H_s(z) = C_s(z)/A(z)$ é a função para a qual

$$[G(z^{-1})H_s(z)]_{\oplus} = \frac{s(\overline{A}(z) + cA(z))}{A(z)}.$$

3. Um filtro adaptativo faz a identificação do sistema

$$H_b(z) = H(z) - H_a(z)$$

Figura 4.26: Diagrama em blocos do algoritmo de aproximações sucessivas para o caso de entrada não-branca.

durante um número pré-estabelecido n_a de amostras, obtendo-se a função $\hat{H}_b(z)$.

4. A partir do valor de $\hat{H}_b(z)$, determina-se $H'(z) = H_a(z) + \hat{H}_b(z)$ e uma nova estimativa $\hat{H}'(z) = B'(z)/A'(z)$ é obtida através da minimização do erro quadrático de equação $\|A'(z)H'(z) - B'(z)\|_{S_u}^2$ (ver Apêndice D). Notar que esta minimização leva em conta a correlação da entrada. Para este procedimento usa-se um número n_x de amostras, durante as quais a adaptação de $\hat{H}_b(z)$ é congelada.
5. Durante n_t amostras os parâmetros do bloco auxiliar $H_a(z)$ e do filtro adaptativo sofrem uma transição suave para os valores decorrentes de $H'_a(z)$ obtidos a partir da estimativa $\hat{H}'(z)$ do passo anterior, desta maneira evitando transitórios intensos. A adaptação continua congelada.
6. A adaptação continua congelada por mais n_w amostras, de modo a não ser influenciada por transitórios residuais. Volta-se então ao passo 3.

Um diagrama em blocos do algoritmo é mostrado na Figura (4.26). Caso tenhamos $E(z) = 1$, $F(z) = 1$, $s = 1$ e $c = 0$, o algoritmo descrito recai no caso

Funções utilizadas		
Função	Pólos	Zeros
$H(z^{-1})$	$0,8\angle \pm 30^\circ$ e $0,9\angle \pm 100^\circ$	$1/0,8\angle \pm 30^\circ$ e $1/0,9\angle \pm 100^\circ$
$G(z^{-1})$	$0,7\angle \pm 110^\circ$	$0,75\angle \pm 30^\circ$
Parâmetros da simulação numérica (forma direta)		
Ordem de $\hat{H}(z)$	$M = 4$	
Passos μ_A, μ_B	2×10^{-2}	
$B_0(z), A_0(z)$	0,4 vezes os valores ótimos	
Parâmetros da simulação numérica (algoritmo SA)		
Ordem de $\hat{H}(z)$	$M = 4$	
Passo μ_A	5×10^{-3}	
Passo μ_B	2×10^{-2}	
$B_0(z), A_0(z)$	0,4 vezes os valores ótimos	
n_a	2200	
n_x	150	
n_t	50	
n_w	100	
s	0,5	
c	0,0	

Tabela 4.3: Inverso dos pólos e zeros do sistema $H(z)$ sendo identificado e do filtro $G(z)$ utilizado para gerar a entrada não-branca junto com os parâmetros do filtro adaptativo IIR na forma direta e com o algoritmo SA usado na simulação numérica da Figura 4.27.

de entrada branca com variância unitária.

4.8.2 Exemplo numérico

O algoritmo de aproximações sucessivas foi aplicado para o sistema passa-tudo da Seção 3.1. O erro obtido com este algoritmo se encontra na Figura 4.27 e os parâmetros utilizados na Tabela 4.3.

Vemos que para atingir um erro de -60dB são necessárias cerca de 9×10^3 amostras. Comparando o valor dos passos de adaptação no caso do algoritmo SA vemos que estes estão próximos aos máximos utilizados na Seção 4.7.2.1 para identificar o sistema $H_s(z)$. Isto indica que o algoritmo realmente faz com que o sistema sendo identificado se pareça com $H_s(z)$.

Figura 4.27: Algoritmo de aproximações sucessivas (A) e algoritmo de gradiente na forma direta (B).

4.9 Conclusão

A análise de filtros adaptativos IIR em configuração de identificação feita neste capítulo indica que, no caso de entrada não-branca, as características do sistema projetado $H_p(z)$ controlam as propriedades de convergência da adaptação de $\hat{H}(z)$ para identificar o sistema $H(z)$. Analisamos algumas propriedades deste sistema projetado apontando eventuais problemas que surgem com a utilização de entrada não-branca.

Os resultados apontados explicam porque sistemas fáceis de se identificar com entrada branca não o são com outro tipo de entrada e vice-versa, e com isto determinamos uma condição para que haja um menor espalhamento dos autovetores de $\mathbf{J}\mathbf{J}^t\mathbf{\Sigma}^2$. Com esta condição, o algoritmo de aproximações sucessivas foi alterado de forma a levar em conta a função densidade espectral da entrada. Estas modificações levaram a uma maior velocidade de convergência, do mesmo modo que o algoritmo original para o caso de entrada branca.

Além disto, os resultados obtidos nos permitem afirmar que branquear a en-

trada nem sempre proporcionará uma maior velocidade de convergência, apesar de ser uma situação em que as propriedades são bem conhecidas.

CAPÍTULO 5 CONCLUSÃO E TRABALHOS FUTUROS

Este capítulo encerra o trabalho, apresentando uma discussão final sobre os resultados obtidos e sugestões para futuros desenvolvimentos.

5.1 Resultados obtidos

Neste trabalho a análise efetuada para filtros adaptativos IIR com entrada branca de [23] e [24] foi estendida para o caso de entrada não-branca. Verificamos também que a configuração de identificação inversa tratada em [19] e [20] é um caso particular da configuração de identificação com entrada não-branca, o que nos permitiu analisar as duas configurações de maneira unificada.

O principal resultado obtido foi mostrar que no caso de entrada não-branca, as características do sistema projetado $H_p(z)$ controlam as propriedades de convergência da adaptação de $\hat{H}(z)$ para identificar o sistema $H(z)$. Este sistema projetado depende tanto da função densidade espectral da entrada quanto do sistema sendo identificado e tem a mesma ordem que $H(z)$ quando há modelamento suficiente. A existência do sistema projetado $H_p(z)$ controlando as propriedades de convergência de $\hat{H}(z)$ explica porque sistemas fáceis de se identificar com entrada branca não o são com entrada não-branca e vice-versa. Também explica porque branquear a entrada nem sempre fornece uma maior velocidade de convergência, apesar de ser uma situação em que vários resultados e propriedades são conhecidos.

Outro resultado importante obtido foi mostrar que quando a adaptação é lenta, esta se dá próxima à superfície de erro restrita, onde é mais acentuada a

influência de $H_p(z)$. Concluimos então que neste caso a aproximação de primeira ordem de (2.67), a qual descreve localmente a adaptação dos pólos de $\hat{H}(z)$, poderia ser utilizada substituindo o sistema $H(z)$ por $H_p(z)$ e que as conclusões de [23] também se aplicariam a este caso.

Determinamos, então, uma condição para que haja um menor espalhamento dos autovetores da matriz $\mathbf{J}\mathbf{J}^t\mathbf{\Sigma}^2$ associada a $H_p(z)$. Com esta condição, o algoritmo de aproximações sucessivas foi alterado de forma a levar em conta a função densidade espectral da entrada. Agora, todos os cálculos e funções envolvidas dependem de $S_u(z)$, e quando $S_u(z) = \sigma_u^2$, ou seja, a entrada é branca, o algoritmo recai no que foi apresentado em [24]. Estas modificações levaram a uma maior velocidade de convergência, do mesmo modo que o algoritmo original para o caso de entrada branca.

5.2 Trabalhos futuros

Uma limitação da análise efetuada diz respeito ao que ocorre quando a adaptação de $\hat{H}(z)$ não se dá próxima a superfície de erro restrita. Neste caso, a influência de $H_p(z)$ nas propriedades de convergência do filtro adaptativo é menor, impossibilitando um desempenho semelhante ao do caso de entrada branca. Desta forma, um trabalho futuro ligado à análise efetuada que seria de grande interesse é um melhor entendimento das propriedades da superfície de erro restrita e sua ligação com a superfície de erro reduzida. Com isto seria possível determinar condições para que a adaptação ocorresse mais próxima a superfície de erro restrita, o que poderia aumentar a velocidade de convergência.

Outra sugestão para futuros desenvolvimentos é explorar a possibilidade de se filtrar o sinal de entrada e o sinal de saída do sistema desconhecido $H(z)$ de modo que a adaptação ocorra como se a entrada fosse não-branca mas com uma densidade espectral apropriada para se obter uma maior velocidade de convergência. Alguns resultados iniciais promissores neste sentido foram obtidos, mas não foi obtido nada conclusivo.

APÊNDICE A ALGORITMO DE FILTRAGEM ADAPTATIVA IIR NA FORMA DIRETA

A seguir descreveremos o algoritmo usado para adaptar os coeficientes de $B(z)$ e $A(z)$ em filtros adaptativos na forma direta utilizando o gradiente estocástico. Esta descrição segue a de [5, pág. 271]. A função $\hat{H}(z, n)$ corresponde à função de transferência do filtro sendo adaptado no instante n

$$\hat{H}(z, n) = \frac{b_0(n) + b_1(n)z + \cdots + b_M(n)z^M}{1 + a_1(n)z + \cdots + a_M(n)z^M}.$$

Para facilitar o entendimento do algoritmo, na Figura A.1 temos uma realização onde todos os sinais necessários para a atualização dos parâmetros estão disponíveis (reprodução da Figura 2.7).

Algoritmo adaptativo IIR na forma direta

Variáveis disponíveis na iteração n :

- Estados dos filtros:

$$\nabla_{b_k}(n-1), \quad k = 0, \dots, M-1, \quad (\text{filtro adaptativo IIR})$$

$$\nabla_{a_k}(n), \quad k = 1, \dots, M, \quad (\text{pós-filtro de saída})$$

A variável $\nabla_{b_M}(n)$ será calculada mas não precisa ser armazenada.

- Valor dos coeficientes:

$$a_1(n), a_2(n), \dots, a_M(n), b_0(n), b_1(n), \dots, b_M(n)$$

- Novos valores:

$$u(n), \quad \text{amostra da entrada}$$

$$y(n), \quad \text{amostra da saída.}$$

Figura A.1: Geração dos sinais regressores filtrados no algoritmo na forma direta.

Cálculo dos filtros:

- Recursão dos estados do filtro adaptativo:

$$\begin{bmatrix} \nabla_{b_0}(n) \\ \vdots \\ \nabla_{b_{M-1}}(n) \\ \nabla_{b_M}(n) \end{bmatrix} = \begin{bmatrix} -a_1(n) & \cdots & -a_M(n) & 1 \\ 1 & \cdots & 0 & 0 \\ 0 & \ddots & \vdots & \vdots \\ 0 & \ddots & 0 & 0 \\ 0 & \ddots & 1 & 0 \end{bmatrix} \begin{bmatrix} \nabla_{b_0}(n-1) \\ \vdots \\ \nabla_{b_{M-1}}(n-1) \\ u(n) \end{bmatrix}$$

- Saída $\hat{y}(n)$ do filtro adaptativo $\hat{H}(z, n)$ e o erro da saída $e(n)$:

$$\begin{aligned} \hat{y}(n) &= \sum_{k=0}^M b_k(n) \nabla_{b_k}(n) \\ e(n) &= y(n) - \hat{y}(n) \end{aligned}$$

- Atualização dos coeficientes:

$$\begin{aligned} b_k(n+1) &= b_k(n) + \mu_B e(n) \nabla_{b_k}(n), & k = 0, 1, \dots, M \\ a_k(n+1) &= a_k(n) + \mu_A e(n) \nabla_{a_k}(n), & k = 1, \dots, M \end{aligned}$$

- Recursão dos estados do pós-filtro da saída:

$$\begin{bmatrix} \nabla_{a_1}(n+1) \\ \vdots \\ \nabla_{a_{M-1}}(n+1) \\ \nabla_{a_M}(n+1) \end{bmatrix} = \begin{bmatrix} -a_1(n+1) & \cdots & -a_M(n+1) & 1 \\ 1 & \cdots & 0 & 0 \\ 0 & \ddots & \vdots & \vdots \\ 0 & \ddots & 0 & 0 \\ 0 & \ddots & 1 & 0 \end{bmatrix} \begin{bmatrix} \nabla_{a_1}(n) \\ \vdots \\ \nabla_{a_M}(n) \\ -\hat{y}(n) \end{bmatrix}.$$

APÊNDICE B ALGORITMO DE FILTRAGEM ADAPTATIVA IIR NA FORMA EM TRELIÇA NORMALIZADA

No caso de filtros adaptativos IIR na forma em treliça normalizada, o algoritmo utilizado é o que se segue. Este algoritmo é o mesmo de [5, pág. 332]. Agora o filtro adaptativo $\hat{H}(z, n)$ é descrito pelos coeficientes $\nu_k(n)$ e $\theta_k(n)$. Para facilitar o entendimento do algoritmo, reproduzimos na Figura B.1, onde $\gamma_M = 1$ e

$$\gamma_j \triangleq \prod_{i=j+1}^M \cos \theta_i, \quad j = 1, \dots, M-1,$$

como os regressores são obtidos.

Algoritmo de filtros adaptativos na forma em treliça normalizada (gradiente parcial simplificado)

Sinais disponíveis no instante n :

- Coeficientes do filtro $\hat{H}(z, n)$:

$$\nu_k(n), \quad k = 0, \dots, M$$

$$\theta_k(n), \quad k = 1, \dots, M$$

- Estados do filtro adaptativo $\hat{H}(z, n)$:

$$\nabla_{\nu_k}(n-1), \quad k = 0, \dots, M-1.$$

A variável $\nabla_{\nu_M}(n)$ será calculada mas não precisa ser armazenada.

- Estado do pós-filtro:

$$\xi_k(n), \quad k = 0, 1, \dots, M-1.$$

Figura B.1: Obtenção dos regressores em filtros adaptativos na forma em treliça normalizada.

A variável $\xi_M(n)$ será calculada mas não precisa ser armazenada.

- Novos valores:

$u(n)$, amostra da entrada

$y(n)$, amostra da saída.

Cálculo dos filtros:

- Seja $g_M = u(n)$.
- Para $k = M, M - 1, \dots, 1$ faça

$$\begin{bmatrix} g_{k-1} \\ \nabla_{\nu_k}(n) \end{bmatrix} = \begin{bmatrix} \cos \theta_k(n) & -\sin \theta_k(n) \\ \sin \theta_k(n) & \cos \theta_k(n) \end{bmatrix} \begin{bmatrix} g_k \\ \nabla_{\nu_{k-1}}(n-1) \end{bmatrix}.$$

O valor de g_{k-1} pode sobrepor o valor de g_k na memória do sistema sendo utilizado para processamento.

- $\nabla_{\nu_0}(n) = g_0$.
- Saída $\hat{y}(n)$ do filtro adaptativo $\hat{H}(z, n)$:

$$\hat{y}(n) = \sum_{k=0}^M \nu_k(n) \nabla_{\nu_k}(n).$$

- Erro da saída:

$$e(n) = y(n) - \hat{y}(n).$$

Regressores filtrados:

- Seja $\gamma_M = 1$.
- Para $k = M, M - 1, \dots, 1$, faça

$$\nabla_{\theta_k} = \gamma_k \xi_{k-1}(n)$$

$$\gamma_{k-1} = \gamma_k \cos \theta_k(n).$$

O valor de γ_{k-1} pode sobrepor o valor γ_k na memória.

Atualização dos coeficientes:

$$\begin{aligned}\nu_k(n+1) &= \nu_k(n) + \mu_\nu e(n) \nabla_{\nu_k}(n), \quad k = 0, 1, \dots, M \\ \theta_k(n+1) &= \theta_k(n) + \nu_\theta e(n) \nabla_{\theta_k}(n), \quad k = 1, \dots, M.\end{aligned}$$

Teste de estabilidade:

- Se $|\theta_k(n+1)| > \pi/2$, $k = 1, 2, \dots, M$, então $\theta_k(n+1) = \theta_k(n)$.

Cálculo do pós-filtro:

- Seja $t = -\hat{y}(n)$.
- Para $k = M, M-1, \dots, 1$ faça

$$\begin{bmatrix} t \\ \xi_k(n+1) \end{bmatrix} = \begin{bmatrix} \cos \theta_k(n+1) & -\sin \theta_k(n+1) \\ \sin \theta_k(n+1) & \cos \theta_k(n+1) \end{bmatrix} \begin{bmatrix} t \\ \xi_{k-1}(n) \end{bmatrix}.$$

Vemos que o novo valor de t sobrepõem o antigo.

- $\xi_0(n+1) = t$.

APÊNDICE C ALGORITMO DE FILTRAGEM ADAPTATIVA IIR NA FORMA DIRETA EM CONFIGURAÇÃO DE IDENTIFICAÇÃO INVERSA

O algoritmo que segue é similar ao do Apêndice A, sendo $\hat{H}(z, n)$ da mesma forma. É utilizado o gradiente estocástico descrito em [19]. A Figura C.1 mostra a geração dos sinais regressores.

Algoritmo adaptativo IIR forma direta em configuração de identificação inversa

Variáveis disponíveis no instante n :

- Estados dos filtros:

$$\nabla_{b_k}(n-1), \quad k = 0, \dots, M-1, \quad (\text{filtro adaptativo IIR})$$

$$\nabla_{a_k}(n), \quad k = 1, \dots, M, \quad (\text{pós-filtro de saída})$$

A variável $\nabla_{b_M}(n)$ será calculada mas não precisa ser armazenada.

- Valor dos coeficientes:

$$a_1(n), a_2(n), \dots, a_M(n), b_0(n), b_1(n), \dots, b_M(n)$$

- Novos valores:

$$u(n), \quad \text{amostra da entrada}$$

$$y(n), \quad \text{amostra da saída.}$$

Figura C.1: Geração dos sinais regressores filtrados no algoritmo na forma direta.

Cálculo dos filtros:

- Recursão dos estados do filtro adaptativo:

$$\begin{bmatrix} \nabla_{b_0}(n) \\ \vdots \\ \nabla_{b_{M-1}}(n) \\ \nabla_{b_M}(n) \end{bmatrix} = \begin{bmatrix} -a_1(n) & \cdots & -a_M(n) & 1 \\ 1 & \cdots & 0 & 0 \\ 0 & \ddots & \vdots & \vdots \\ 0 & \ddots & 0 & 0 \\ 0 & \ddots & 1 & 0 \end{bmatrix} \begin{bmatrix} \nabla_{b_0}(n-1) \\ \vdots \\ \nabla_{b_{M-1}}(n-1) \\ y(n) \end{bmatrix}$$

- Saída $\hat{u}(n)$ do filtro adaptativo $\hat{H}(z, n)$ e o erro da saída $e(n)$:

$$\begin{aligned} \hat{u}(n) &= \sum_{k=0}^M b_k(n) \nabla_{b_k}(n) \\ e(n) &= u(n) - \hat{u}(n) \end{aligned}$$

- Atualização dos coeficientes:

$$\begin{aligned} b_k(n+1) &= b_k(n) + \mu_B e(n) \nabla_{b_k}(n), & k &= 0, 1, \dots, M \\ a_k(n+1) &= a_k(n) + \mu_A e(n) \nabla_{a_k}(n), & k &= 1, \dots, M \end{aligned}$$

- Recursão dos estados do pós-filtro da saída:

$$\begin{bmatrix} \nabla_{a_1}(n+1) \\ \vdots \\ \nabla_{a_{M-1}}(n+1) \\ \nabla_{a_M}(n+1) \end{bmatrix} = \begin{bmatrix} -a_1(n+1) & \cdots & -a_M(n+1) & 1 \\ 1 & \cdots & 0 & 0 \\ 0 & \ddots & \vdots & \vdots \\ 0 & \ddots & 0 & 0 \\ 0 & \ddots & 1 & 0 \end{bmatrix} \begin{bmatrix} \nabla_{a_1}(n) \\ \vdots \\ \nabla_{a_M}(n) \\ -\hat{u}(n) \end{bmatrix}.$$

APÊNDICE D MÉTODO DE IDENTIFICAÇÃO UTILIZANDO O ERRO DE EQUAÇÃO

Seja $y(n) = H(z)u(n)$ a saída do sistema

$$H(z) = \frac{C(z)}{D(z)} = \frac{c_0 + c_1z + \cdots + c_Nz^N}{1 + d_1z + \cdots + d_Nz^N}$$

com sinal de entrada $u(n)$. Com os vetores

$$\begin{aligned}\mathbf{u}(n) &= \begin{bmatrix} u(n) & u(n-1) & \cdots & u(M) \end{bmatrix}^t \\ \mathbf{y}(n) &= \begin{bmatrix} y(n) & y(n-1) & \cdots & y(M) \end{bmatrix}^t\end{aligned}$$

construimos a matriz

$$\mathbf{R}_M = E \left\{ \begin{bmatrix} \mathbf{u}(n) \\ \mathbf{y}(n) \end{bmatrix} \begin{bmatrix} \mathbf{u}(n) & \mathbf{y}(n) \end{bmatrix} \right\} = \begin{bmatrix} \mathbf{R}_{uu} & \mathbf{R}_{uy} \\ \mathbf{R}_{uy}^t & \mathbf{R}_{yy} \end{bmatrix} \quad (\text{D.1})$$

de dimensões $2(M+1) \times 2(M+1)$. O seguinte teorema [5, sec. 3.4] nos ajudará a mostrar um método para identificação de $H(z)$:

Teorema D.1 *Sejam os sinais $u(n)$ e $y(n) = H(z)u(n)$. A matriz $\mathbf{R}_M$ em (D.1) é singular se e somente se $H(z)$ for racional de ordem menor ou igual a M .*

Para verificar o Teorema D.1, note que $\mathbf{R}_M$ é por construção positiva semi-definida, e então é singular se e somente se existir um vetor $\mathbf{w}$ diferente de zero para o qual

$$\mathbf{w}^t \mathbf{R}_M \mathbf{w} = 0. \quad (\text{D.2})$$

Se os elementos de $\mathbf{w}$ forem

$$\mathbf{w} = \begin{bmatrix} -b_0 \\ -b_1 \\ \vdots \\ -b_M \\ a_0 \\ a_1 \\ \vdots \\ a_M \end{bmatrix} \quad (\text{D.3})$$

então (D.2) equivale à

$$E \left\{ \left(\sum_{k=0}^M a_k y(n-k) - \sum_{k=0}^M b_k u(n-k) \right)^2 \right\} = 0.$$

Isto nos leva a

$$\sum_{k=0}^M a_k y(n-k) = \sum_{k=0}^M b_k u(n-k), \quad \text{para qualquer } n,$$

que pode ser modificada para

$$a_0 y(n) = \sum_{k=0}^M b_k u(n-k) - \sum_{k=1}^M a_k y(n-k)$$

correspondendo a equação de diferença obtida quando a entrada $u(n)$ é aplicada ao sistema

$$\hat{H}(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1 z + \cdots + b_M z^M}{a_0 + a_1 z + \cdots + a_M z^M}$$

produzindo a mesma saída $y(n)$ do sistema $H(z)$ original. Isto só será possível se

$$\hat{H}(z) = \frac{B(z)}{A(z)} = \frac{C(z)P(z)}{D(z)P(z)},$$

onde $\deg[P(z)] = M - N$. Isto mostra a necessidade de $H(z)$ ser racional de ordem menor igual a M para o Teorema D.1. Para mostrar que esta condição é suficiente, basta escrever a equação de diferenças que descreve a saída $y(n) = H(z)u(n)$

$$\begin{aligned} d_0 y(n) &= \sum_{k=0}^N c_k u(n-k) - \sum_{k=1}^N d_k y(n-k) \\ 0 &= \sum_{k=0}^N c_k u(n-k) - \sum_{k=0}^N d_k y(n-k) \end{aligned}$$

e calcular sua variância

$$E \left\{ \left(\sum_{k=0}^N d_k y(n-k) - \sum_{k=0}^N c_k u(n-k) \right)^2 \right\} = 0,$$

que em forma matricial equivale a (D.2) com $\mathbf{w} = [-c_0 \quad \dots \quad -c_N \quad d_0 \quad \dots \quad d_N]$ e $M = N$.

Na verdade, em (D.2) o vetor $\mathbf{w}$ se encontra no espaço nulo de $\mathbf{R}_M$. Com isto e o Teorema D.1 podemos realizar a identificação de $H(z)$ construindo a matriz $\mathbf{R}_M$ e encontrando um vetor em seu espaço nulo. O vetor encontrado minimizará a variância da equação de erro, ou seja,

$$\begin{aligned} E[e^2(n)] &= E \left\{ \left(\sum_{k=0}^M a_k y(n-k) - \sum_{k=0}^M b_k u(n-k) \right)^2 \right\} \\ &= E \left\{ \left(\sum_{k=0}^M a_k H(z) u(n-k) - \sum_{k=0}^M b_k u(n-k) \right)^2 \right\} \\ &= E \{ [(A(z)H(z) - B(z)) u(n-k)]^2 \} \end{aligned}$$

que, pela equação (2.20), pode ser escrita como

$$E[e^2(n)] = \|A(z)H(z) - B(z)\|_{S_u}^2.$$

Para o caso em que $N \leq M$ (modelamento suficiente), evitando a solução trivial $A(z) = B(z) = 0$, temos que a estimativa $\hat{H}(z)$ obtida será da forma

$$\hat{H}(z) = \frac{B(z)}{A(z)} = \frac{C(z)P(z)}{D(z)P(z)}$$

com $P(z) = p_0 + p_1 z + \dots + p_{M-N} z^{M-N}$. Logo vemos que temos $M - N + 1$ graus de liberdade na escolha do vetor $\mathbf{w}$ e dos coeficientes de $B(z)$ e $A(z)$. Um destes graus de liberdade será utilizado para impor $a_0 = 1$, de modo que o polinômio $A(z)$ obtido seja mônico. Particionando a matriz

$$\mathbf{R}_{uy} = \begin{bmatrix} \mathbf{r}_{uy} & \mathbf{R}_{uy,2} \end{bmatrix},$$

onde $\mathbf{r}_{uy}$ é formado a partir dos valores de $r_{uy}(k) = E[u(n)y(n-k)]$, $k = 0, \dots, M$, e $\mathbf{R}_{uy,2}$ é a matriz formada com as colunas de 2 a M de $\mathbf{R}_{uy}$, e a matriz

$$\mathbf{R}_{yy} = \begin{bmatrix} \mathbf{r}_{yy}^t \\ \mathbf{R}_{yy,2} \end{bmatrix},$$

com $\mathbf{r}_{yy}$ é formado a partir dos valores de $r_{yy}(k) = E[y(n)y(n-k)]$, $k = 0, \dots, M$, e $\mathbf{R}_{yy,2}$ é a matriz formada com as linhas de 2 a M de $\mathbf{R}_{yy}$, temos

$$\begin{bmatrix} \mathbf{R}_{uu} & \mathbf{R}_{uy} \\ \mathbf{R}_{uy,2} & \mathbf{R}_{yy,2} \end{bmatrix} \mathbf{w} = \begin{bmatrix} \mathbf{r}_{uy} \\ \mathbf{r}_{yy} \end{bmatrix}. \quad (\text{D.4})$$

A solução deste sistema fornece os coeficientes de $B(z)$ e $A(z)$. Em [5, cap. 3] é mostrado que no caso de submodelamento ($N > M$), a solução de (D.4) nos fornece uma boa aproximação para $H(z)$.

No caso de $u(n)$ branco, temos

$$\begin{bmatrix} \mathbf{I}_M & \boldsymbol{\Theta}_H \\ \theta_H & \mathbf{R}_{yy,2} \end{bmatrix} \mathbf{w} = \begin{bmatrix} \mathbf{h} \\ \mathbf{r}_{yy} \end{bmatrix}, \quad (\text{D.5})$$

com $\mathbf{h} = [h_0 \ h_1 \ \dots \ h_M]^t$ e

$$\begin{aligned} \boldsymbol{\Theta}_H &= \begin{bmatrix} h_0 & 0 & \dots & 0 \\ h_1 & h_0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_M & h_{M-1} & \dots & h_0 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{h} & \theta_H \end{bmatrix}. \end{aligned}$$

APÊNDICE E RESULTADOS DA TEORIA DE MATRIZES

Neste apêndice abordamos alguns resultados da teoria de matrizes que foram utilizados no decorrer do texto. Também é mostrado como calcular os gramianos $\mathbf{K}$ e $\mathbf{W}$ usando ambientes compatíveis com o MatLab[©] e como ortogonalizar estas matrizes de forma a se obter uma realização balanceada.

E.1 Invariância de autovalores à permutação

Se $\mathbf{M}$ é uma matriz $m \times n$ e $\mathbf{N}$ é uma matriz $n \times m$, com $m \leq n$, então $\mathbf{NM}$ tem os mesmos autovalores que $\mathbf{MN}$, além de $n - m$ autovalores adicionais iguais a zero. Uma demonstração pode ser encontrada em [27, p. 53].

E.2 Positividade de autovalores

Se uma matriz real $\mathbf{R}$ é simétrica é simples mostrar que seus autovalores são reais [27]. Segue que seus autovetores podem ser sempre tomados como reais também. Se, além disto, $\mathbf{R}$ for simétrica positiva semidefinida (isto é, $\mathbf{x}^t \mathbf{R} \mathbf{x} \geq 0$ para qualquer $\mathbf{x}$) então seus autovalores λ_k são positivos:

$$\mathbf{R} \mathbf{x} = \lambda_k \mathbf{x} \Rightarrow \mathbf{x}^t \mathbf{R} \mathbf{x} = \lambda_k \|\mathbf{x}\|^2 \geq 0 \Rightarrow \lambda_k \geq 0.$$

E.3 Fatoração e raiz quadrada

Qualquer matriz real e simétrica $\mathbf{R}$ pode ser fatorada como

$$\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^t,$$

onde $\mathbf{Q}$ é ortogonal ($\mathbf{Q}\mathbf{Q}^t = \mathbf{Q}^t\mathbf{Q} = \mathbf{I}$) e $\mathbf{\Lambda}$ é diagonal e composta pelos autovalores de $\mathbf{R}$ (ainda que repetidos) [27]. Se $\mathbf{R}$ for também positiva semidefinida então podemos definir sua raiz quadrada real, simétrica e positiva semidefinida

$$\mathbf{R}^{1/2} = \mathbf{Q}\mathbf{\Lambda}^{1/2}\mathbf{Q}^t,$$

para a qual vale

$$\mathbf{R}^{1/2}\mathbf{R}^{1/2} = \mathbf{Q}\mathbf{\Lambda}^{1/2}\mathbf{Q}^t\mathbf{Q}\mathbf{\Lambda}^{1/2}\mathbf{Q}^t = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^t = \mathbf{R}.$$

E.4 Positividade de autovalores de produto de matrizes

Os autovalores λ_k do produto de duas matrizes $\mathbf{R}$ e $\mathbf{P}$ reais, simétricas e positivas semidefinidas, são positivos:

$$\mathbf{R}\mathbf{P}\mathbf{x} = \lambda_k\mathbf{x} \Rightarrow \mathbf{P}^{1/2}\mathbf{R}\mathbf{P}^{1/2}\mathbf{P}^{1/2}\mathbf{x} = \lambda_k\mathbf{P}^{1/2}\mathbf{x},$$

logo λ_k é autovalor de $\mathbf{P}^{1/2}\mathbf{R}\mathbf{P}^{1/2}$, que é real, simétrica e positiva semi-definida. Como visto acima, resulta então que $\lambda_k \geq 0$.

E.5 SVD

A decomposição em valores singulares (SVD) [27] de uma matriz quadrada real $\mathbf{M}$ é

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^t,$$

onde $\mathbf{U}$ e $\mathbf{V}$ são ortogonais. Ela pode ser computada a partir da computação de autovalores: observando os produtos reais, simétricos e positivos semidefinidos

$$\mathbf{M}^t\mathbf{M} = \mathbf{V}\mathbf{\Sigma}\mathbf{U}^t\mathbf{U}\mathbf{\Sigma}\mathbf{V}^t = \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^t$$

e

$$\mathbf{M}\mathbf{M}^t = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^t\mathbf{V}\mathbf{\Sigma}\mathbf{U}^t = \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^t,$$

vemos que $\mathbf{\Sigma}^2$ contém os quadrados dos autovalores de $\mathbf{M}^t\mathbf{M}$ (que são iguais aos de $\mathbf{M}\mathbf{M}^t$), $\mathbf{U}$ contém os autovetores de $\mathbf{M}\mathbf{M}^t$ e $\mathbf{V}$ contém os autovetores de $\mathbf{M}^t\mathbf{M}$.

E.6 Computação de $\mathbf{K}$ e $\mathbf{W}$

A computação dos gramianos $\mathbf{K}$ e $\mathbf{W}$ pode ser realizada resolvendo as equações de Lyapunov discretas associadas a cada um. Em ambientes compatíveis com MatLab[®] isto é realizado com a função `dlyap`.

E.7 Computação da realização balanceada

Os gramianos $\mathbf{K}$ e $\mathbf{W}$ de uma realização mínima de $\hat{H}(z)$ são positivos definidos. Segue que suas raízes quadradas $\mathbf{K}^{1/2}$ e $\mathbf{W}^{1/2}$ também o são. Consideremos a decomposição em valores singulares do produto $\mathbf{K}^{1/2}\mathbf{W}^{1/2}$:

$$\mathbf{K}^{1/2}\mathbf{W}^{1/2} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^t.$$

Notamos que, pela propriedade de invariância à permutação vista acima, os autovalores de $\mathbf{K}^{1/2}\mathbf{W}^{1/2} (\mathbf{K}^{1/2}\mathbf{W}^{1/2})^t = \mathbf{K}^{1/2}\mathbf{W}^{1/2}\mathbf{W}^{1/2}\mathbf{K}^{1/2}$ são iguais aos autovalores de $\mathbf{KW}$. Portanto, a matriz $\mathbf{\Sigma}$ acima é composta pelos valores singulares de Hankel de $\hat{H}(z)$. Notamos também que vale $\mathbf{W}^{1/2}\mathbf{K}^{1/2} = \mathbf{V}\mathbf{\Sigma}\mathbf{U}^t$.

Consideremos agora a transformação

$$\mathbf{T} = \mathbf{\Sigma}^{-1/2}\mathbf{V}^t\mathbf{W}^{1/2},$$

cuja inversa, pode-se constatar diretamente, vale $\mathbf{T}^{-1} = \mathbf{K}^{1/2}\mathbf{U}\mathbf{\Sigma}^{-1/2}$. A aplicação desta transformação diagonaliza os gramianos $\mathbf{K}$ e $\mathbf{W}$, podendo ser usada, portanto, para obter a realização balanceada. Constatamos a diagonalização

usando 2.14 e 2.15:

$$\begin{aligned}
\mathbf{K}_b &= \mathbf{T}\mathbf{K}\mathbf{T}^t \\
&= \mathbf{\Sigma}^{-1/2}\mathbf{V}^t\mathbf{W}^{1/2}\mathbf{K}^{1/2}\mathbf{K}^{1/2}\mathbf{W}^{1/2}\mathbf{V}\mathbf{\Sigma}^{-1/2} \\
&= \mathbf{\Sigma}^{-1/2}\mathbf{V}^t\mathbf{V}\mathbf{\Sigma}\mathbf{U}^t\mathbf{U}\mathbf{\Sigma}\mathbf{V}^t\mathbf{V}\mathbf{\Sigma}^{-1/2} \\
&= \mathbf{\Sigma},
\end{aligned}$$

procedendo analogamente para $\mathbf{W}$.

Referências Bibliográficas

- [1] S. Haykin, *Adaptive Filter Theory*. Upper Saddle River: Prentice-Hall, 2004.
- [2] J. Treichler, C. R. Johnson, Jr., and M. G. Larimore, *Theory and design of adaptive filters*. Upper Saddle River: Prentice-Hall, 2001.
- [3] J. J. Shynk, “Adaptive IIR filtering,” *IEEE ASSP Magazine*, vol. 6, pp. 4–21, Apr. 1989.
- [4] L. Ljung and T. Söderström, *Theory and practice of recursive identification*. Cambridge: MIT Press, 1983.
- [5] P. A. Regalia, *Adaptive IIR Filtering in Signal Processing and Control*. New York: Marcel Dekker, 2005.
- [6] S. D. Stearns, “Error surfaces of recursive adaptive filters,” *IEEE Transactions on Circuits and Systems*, vol. 28, pp. 603–606, June 1981.
- [7] C. R. Johnson, Jr., “Adaptive IIR filtering: current results and open issues,” *IEEE Transactions on Information Theory*, vol. 30, pp. 237–250, Mar. 1984.
- [8] H. Fan, “A structural view of asymptotic convergence speed of adaptive IIR filtering algorithms. I. Infinite precision implementation,” *IEEE Transactions on Signal Processing*, vol. 41, pp. 1493–1517, Apr. 1993.
- [9] V. L. Stonick and M. H. Cheng, “Adaptive IIR filtering: composite pre-filtered regressor method,” in *Proc. IEEE ICASSP*, (Adelaide), pp. 501–504, 1994.
- [10] P. Campolucci and F. Piazza, “Intrinsic stability control method for recursive filters and neural networks,” *IEEE Transactions on Circuits and Systems*, vol. 47, pp. 797–802, Aug. 2000.
- [11] R. López-Valcarce and F. Pérez-González, “Adaptive IIR lattice filtering revisited: convergence issues and new algorithms with improved stability properties,” *IEEE Transactions on Signal Processing*, vol. 49, pp. 811–821, Apr. 2001.
- [12] A. David and T. Aboulnasr, “A globally convergent adaptive IIR filter,” in *Proc. IEEE ISCAS*, vol. 3, (geneva), pp. 531–534, 2000.

- [13] P. M. S. Burt, "A switched regressor algorithm for adaptive IIR filtering," in *Proc. IEEE 38th Midwest Symposium on Circuits and Systems*, vol. 2, (Rio de Janeiro), pp. 990–993, 1995.
- [14] P. M. S. Burt and M. Gerken, "A polyphase IIR adaptive filter: error surface analysis and application," in *Proc. IEEE ICASSP*, vol. 3, (München), pp. 2285–2288, 1997.
- [15] P. M. S. Burt and M. Gerken, "Um filtro polifásico adaptativo," in *Proc. SBT*, (Recife), pp. 511–514, 1997.
- [16] P. M. S. Burt and M. Gerken, "A polyphase IIR adaptive filter," *IEEE Transactions on Circuits and Systems*, vol. 49, pp. 356–359, May 2002.
- [17] P. M. S. Burt and M. Gerken, "On the uniqueness of stationary points in L_2 approximation with a rational polyphase function," *Signal Processing*, vol. 82, pp. 1707–1725, Nov. 2002.
- [18] P. M. S. Burt and P. A. Regalia, "IIR adaptive filtering: convergence speed properties in the undermodelled case," in *Proc. IEEE ICASSP*, vol. 4, (Philadelphia), pp. 45–48, 2005.
- [19] P. M. S. Burt, "On the convergence speed of inverse identification adaptive IIR filters," in *Proc. SBrT/IEEE International Telecommunications Symposium*, (Fortaleza-CE, Brazil), pp. 51–56, 2006.
- [20] P. M. S. Burt, "Inverse identification adaptive IIR filtering: convergence speed analysis and successive approximation algorithm," in *Proc. IEEE ICASSP*, (Honolulu), 2007.
- [21] P. M. S. Burt, "Filtragem adaptativa IIR: análise da velocidade de convergência baseada na teoria de realização balanceada," Fevereiro 2005. Tese de Livre Docência.
- [22] T. Söderström and P. Stoica, "Some properties of the output error method," *Automatica*, vol. 18, no. 8, pp. 93–99, 1982.
- [23] P. M. S. Burt and P. A. Regalia, "A new framework for convergence analysis and algorithm development of adaptive IIR filters," *IEEE Transactions on Signal Processing*, vol. 53, pp. 3129–3140, Aug 2005.
- [24] P. M. S. Burt and P. A. Regalia, "A new framework for convergence analysis and algorithm development of adaptive IIR filters," in *Proc. IEEE ICASSP*, vol. 2, (Montreal), pp. 441–444, 2004.
- [25] T. E. Filgueiras F. and P. M. S. Burt, "On the convergence speed of adaptive IIR filters with rational spectrum input signals," *EUSIPCO 2008*, Aug. 2008.
- [26] A. W. Oppenheim and R. Schaffer, *Discrete-time signal processing*. Prentice-Hall, 1985.

- [27] R. G. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, UK: Cambridge University Press, 1990.

ON THE CONVERGENCE SPEED OF ADAPTIVE IIR FILTERS WITH RATIONAL SPECTRUM INPUT SIGNALS

Thomas E. Filgueiras, F. and Phillip M. S. Burt

Dept. of Telecommunications and Control Engineering
Escola Politécnica, Universidade de São Paulo
Emails: tefil@lcs.poli.usp.br, phillip@lcs.poli.usp.br

ABSTRACT

A previously presented framework for the convergence speed analysis of adaptive IIR filters considered white input signals and an identification configuration. In this framework, a local model was built and the Hankel singular values of the system being identified had a fundamental role. It also led to the successive approximations (SA) algorithm, attaining a large speed gain. Here, an approach to extend this framework is presented, allowing more general rational spectrum input signals and unifying the analysis of the identification and inverse identification configurations. It leads to a model that has the same form and depends on the same number of Hankel singular values as in the white input case. The singular values, however, refer now to the result of a certain degree-preserving transformation of the system being identified. Conditions for faster convergence as well as the application of the SA algorithm follow naturally from the extended framework.

1. INTRODUCTION

A framework for the analysis of the convergence speed of adaptive IIR filters [1]-[3] with white input signals was presented in [4],[5]. It provides a better understanding of the effect of factors such as the order of the rational system being identified, the angular and radial distribution of its poles and zeros and the parameterization (direct form, lattice, etc.) that is employed for the adaptive filter. It also leads to an adaptive algorithm with faster convergence, the successive approximations algorithm (SA). In [6],[7] the same framework was extended to the case of inverse identification with white input and it was suggested that a similar approach could be applied to identification or inverse identification adaptive IIR filtering with non-white input signals. This idea is pursued in this paper, which is organized as follows.

In Section 2, properties of adaptive IIR filtering and of the framework for analysis of its convergence speed are reviewed. In Section 3, an approach to extend this framework to more general rational spectrum input signals is presented. Based on this, a condition for faster convergence is presented in Section 4 and the application of the SA algorithm is discussed in Section 5. In Section 6, perspectives for future work are discussed.

2. REVIEW OF ADAPTIVE IIR FILTERING PROPERTIES

In identification configuration adaptive IIR filtering, a rational function $\hat{H}(z) = B(z)/A(z) = \sum_{k=0}^M b_k z^k / (1 + \sum_{k=1}^M a_k z^k)$ is adapted so as to minimize the mean square error $E[e^2(n)]$, where the error is $e(n) = y(n) - \hat{y}(n)$, with $y(n) = H(z)u(n)$ being the system output for a known input $u(n)$ and $\hat{y}(n) = \hat{H}(z)u(n)$ being the adaptive filter output for the same input. In this mixed notation, z is the unit delay operator and $zu(n) = u(n-1)$. Using $\deg[\cdot]$ to denote the minimum number of unit delays needed to implement a system, we assume throughout the paper that $\deg[H(z)] = N$.

For $X(z) = \sum_{n=-\infty}^{\infty} x_n z^n$ and $Y(z) = \sum_{n=-\infty}^{\infty} y_n z^n$, the standard inner product is $\langle X(z), Y(z) \rangle = \sum_{n=-\infty}^{\infty} x_n y_n$. The mean square error

is given then by

$$E[e^2(n)] = \langle H(z) - \hat{H}(z), S_u(z)[H(z) - \hat{H}(z)] \rangle \quad (1)$$

$$\triangleq \left\| H(z) - \hat{H}(z) \right\|_{S_u}^2, \quad (2)$$

where $S_u(z) = \sum_{k=-\infty}^{\infty} E[u(n)u(n-k)]z^k$.

2.1 Conditions for mean square error minimization

Given the poles of $\hat{H}(z)$, the mean square error is minimized in relation to the zeros of $\hat{H}(z)$ if and only if [3]

$$\left[S_u(z)[H(z) - \hat{H}(z)] \right]_{\oplus} = zg(z)V(z), \quad (3)$$

where $[\cdot]_{\oplus}$ is the causal projection operator (which eliminates the negative powers of z), $V(z) = z^M A(z^{-1})/A(z)$ is all-pass (that is, $|V(e^{j\omega})| \equiv 1$) and $g(z)$ belongs to the Hardy space $\mathcal{H}_2$ of the stable and causal functions of the complex variable z .

If the input signal has a rational spectrum, then $S_u(z) = G(z)G(z^{-1})$, where $G(z)$ is a rational function in $\mathcal{H}_2$ and can also be constrained to have minimum phase. In particular, when the input is white and has unitary variance then $S_u(z) \equiv 1$. In this case, from (3) and from $V(z)$ being all-pass, the mean square error is given by $E[e^2(n)] = \|H(z) - \hat{H}(z)\|^2 = \|g(z)\|^2$, where $\|g(z)\|^2$ as a function of the poles of $\hat{H}(z)$ is called the "reduced error surface".

Considering now the optimization of the zeros and the poles of $\hat{H}(z)$, it can be shown [3] that $\hat{H}(z)$ is at a stationary point (minimum or saddle point) of the mean square error if and only if $[S_u(z)[H(z) - \hat{H}(z)]_{\oplus} = zQ(z)V^2(z)$, for some $Q(z) \in \mathcal{H}_2$. From this and $S_u(z) = G(z)G(z^{-1})$, with $\deg[G(z)] = L$, it can be shown (see also [8]) that the condition $M \geq N + L$ ensures there are no local minima of the mean square error. For a white input signal ($L = 0$) this requires at least sufficient modelling ($M = N$), while for non-white inputs ($L > 0$) it requires over-modelling ($M > N$).

2.2 Mean square error orthonormal decomposition

As follows from [3, p.122], with $\hat{H}(z) = B(z)/A(z) = \sum_{k=0}^M b_k z^k / (1 + \sum_{k=1}^M a_k z^k)$, we can always obtain a set of functions

$$\left\{ \frac{P_k(z)}{A(z)} \right\}_{k=0}^M \cup \left\{ z^k V(z) \right\}_{k=1}^{\infty}, \quad (4)$$

where $\deg[P_k(z)] \leq M$, which is a complete orthonormal basis for $\mathcal{H}_2$. In this basis, $\hat{H}(z) = \sum_{k=0}^M v_k P_k(z)/A(z)$ and $H(z) = \sum_{k=0}^M \rho_k P_k(z)/A(z) + \sum_{k=1}^{\infty} \tau_k z^k V(z)$, which, with $\tau_k = \langle H(z), z^k V(z) \rangle$, leads to

$$\left\| H(z) - \hat{H}(z) \right\|^2 = \sum_{k=0}^M (\rho_k - v_k)^2 + \sum_{k=1}^{\infty} \left\langle H(z), z^k V(z) \right\rangle^2. \quad (5)$$

T. E. Filgueiras, F. and P. M. S. Burt were supported by CNPq and FAPESP (Proc. 06/01113-0), respectively.

Optimizing the zeros of $\hat{H}(z)$ given its poles makes $\sum_{k=0}^M (\rho_k - v_k)^2 = 0$. With $V(z) = \sum_{k=0}^{\infty} v_k z^k$ and $\mathbf{v}^t = [v_0 \ v_1 \ \dots]$, the reduced error surface is given then by [3]

$$\|H(z) - \hat{H}(z)\|^2 = \sum_{k=1}^{\infty} \langle H(z), z^k V(z) \rangle^2 = \mathbf{v}^t \mathbf{\Gamma}_H^2 \mathbf{v}, \quad (6)$$

where $\mathbf{\Gamma}_H$ is the Hankel form of $H(z)$.

2.3 Balanced form factorization

The Hankel form $\mathbf{\Gamma}_H$ can be factored into the controllability and observability forms of a balanced realization of $H(z)$, which, with $\deg[H(z)] = N$, allows us to rewrite (6) so that the reduced error surface is given by

$$\|H(z) - \hat{H}(z)\|^2 = \boldsymbol{\alpha}^t \boldsymbol{\Sigma}^2 \boldsymbol{\alpha},$$

where $\boldsymbol{\Sigma}$ is diagonal and composed of the N Hankel singular values σ_k of $H(z)$ and the elements of $\boldsymbol{\alpha}$ are $\alpha_k = \langle \zeta_k(z), V(z) \rangle$, $k = 1, \dots, N$, with $\zeta_k(z)$ being the normalized controllability transfer functions of a balanced realization of $H(z)$ [4], [5].

If we consider a true gradient adaptation algorithm working on the reduced error surface, a local description of the adaptation process when the input is white is then obtained:

$$\boldsymbol{\alpha}(n+1) \approx [\mathbf{I} - \mu_a \mathbf{J} \mathbf{J}' \boldsymbol{\Sigma}^2] \boldsymbol{\alpha}(n), \quad (7)$$

where μ_a is the adaptation stepsize of the pole parameters and $\mathbf{J}$ is a Jacobian sensitivity matrix, with elements $[\mathbf{J}]_{i,j} = \partial \alpha_i / \partial a_j$. It varies with pole parameters a_j , but for greater simplicity we do not explicitly denote this. If the eigenvalue spread of $\mathbf{J} \mathbf{J}' \boldsymbol{\Sigma}^2$ is large, then convergence is slow at points where $\boldsymbol{\alpha}(n)$ is mainly in the directions of eigenvectors associated with small eigenvalues of $\mathbf{J} \mathbf{J}' \boldsymbol{\Sigma}^2$. In these slowly converging situations, the true gradient adaptation algorithm is usually a good approximation of the stochastic gradient algorithm.

2.4 Faster Convergence and SA Algorithm

Still assuming a white input signal, expression (7) suggests a condition for faster convergence. Since the direct term of $H(z)$ does not affect the associated Hankel form, if $H(z)$ has the form $s[z^N D(z^{-1})/D(z) + c]$ (where we note that $z^N D(z^{-1})/D(z)$ is all-pass), then all Hankel singular values of $H(z)$ are equal to s , which tends to contribute to a smaller eigenvalue spread of $\mathbf{J} \mathbf{J}' \boldsymbol{\Sigma}^2$, and thus to a faster convergence of the adaptation. In the successive approximations (SA) algorithm proposed in [4], an auxiliary block is used in an iterative manner to turn $H(z)$ into an all-pass function and thus speed up convergence.

3. CONVERGENCE SPEED WITH RATIONAL SPECTRUM INPUT SIGNALS

Adaptive filtering in the identification configuration and with a rational spectrum input signal is represented in Fig. 1, where $w(n)$ is a unitary variance white signal. It can be seen that the particular case in which, for some minimum-phase $H_I(z) \in \mathcal{H}_2$, we have $H(z) = 1/H_I(z)$ and $G(z) = H_I(z)$ is equivalent to the inverse identification configuration with white input, also represented in Fig. 1 and which was analyzed in [6], [7]. Moreover, inverse identification with non-white input can be obtained with $G(z) = G'(z)H_I(z)$. Thus, both configurations can be analyzed in an unified manner.

3.1 Example of the effect of a non-white input

In order to motivate the following analysis, we consider now a numerical example. A standard direct-form stochastic gradient algorithm [3, p.271] is applied to the identification of an all-pass system for which $H(z^{-1})$ has poles at $0.8 \angle \pm 30^\circ$ and $0.9 \angle \pm 100^\circ$

Figure 1: Identification (top) and inverse identification (bottom) adaptive filter configurations.

and zeros at $1/0.8 \angle \pm 30^\circ$ and $1/0.9 \angle \pm 100^\circ$. The input $u(n)$ is white Gaussian noise with unitary variance. With $M = N = 4$ and $H(z) = \sum_{k=0}^N c_k z^k / (1 + \sum_{k=1}^N d_k z^k)$, the parameters were initialized at $b_k = 0.4 \times c_k$ and $a_k = 0.4 \times d_k$. Adaptation steps $\mu_a = 0.02$ and $\mu_b = 0.04$ were used for the pole and zero adapted parameters, respectively. They are within a factor of 2 of the values that lead the adaptation to diverge (considering a larger set of initializations), which also holds for the other simulations presented here. The resulting squared error $e^2(n)$ is shown in Fig. 2(A), in logarithmic scale and smoothed for better visualization. As could be expected from the discussion in Section 2, convergence is relatively fast: an error of -60 dB, for instance, is reached in less than 2×10^3 iterations.

We consider now that the input is given by $u(n) = G(z)w(n)$, where $w(n)$ is white Gaussian noise with unitary variance and $G(z^{-1})$ has poles at $0.7 \angle \pm 110^\circ$, zeros at $0.75 \angle \pm 30^\circ$ and unitary L_2 norm. The frequency response $|G(e^{j\omega})|$ is high-pass. With $\mu_a = 0.02$, $\mu_b = 0.02$, $b_k = 0.9 \times c_k$ and $a_k = 0.9 \times d_k$, the resulting squared error $e^2(n)$ is shown in Fig. 2(B). It can be seen that despite starting closer to the optimum, convergence is much slower than for the white input case. It can be also shown that parameters b_k remain in this case very close to their optimum values given parameters a_k .

Figure 2: Squared error, $H(z)$ with white (A) and non-white (B) inputs, $H_s(z)$ with non-white input (C) and $H(z)$ with non-white input and SA algorithm (D).

3.2 Error decomposition

Assuming the minimum-phase function $G(z) \in \mathcal{H}_2$ associated with the input has the form $G(z) = E(z)/F(z)$ and $\deg[G(z)] = L$, a nat-

ural extension to (4) is a complete orthonormal basis for $\mathcal{H}_2$ with the form

$$\left\{ \frac{P_k(z)}{A(z)F(z)} \right\}_{k=0}^{M+L} \cup \left\{ z^k V(z) \frac{\bar{F}(z)}{F(z)} \right\}_{k=1}^{\infty}, \quad (8)$$

where $\deg[P_k(z)] \leq M+L$ and $\bar{F}(z) = z^L F(z^{-1})$. Now, since $G(z)H(z)$ and $G(z)\hat{H}(z)$ are both in $\mathcal{H}_2$, we can decompose them using this basis as $G(z)\hat{H}(z) = \sum_{k=0}^{M+L} v_k P_k(z)/[A(z)F(z)]$ and

$$G(z)H(z) = \sum_{k=0}^{M+L} \rho_k \frac{P_k(z)}{A(z)F(z)} + \sum_{k=1}^{\infty} \tau_k z^k V(z) \frac{\bar{F}(z)}{F(z)}. \quad (9)$$

The mean square error $\|H(z) - \hat{H}(z)\|_{S_u}^2 = \|G(z)[H(z) - \hat{H}(z)]\|^2$ can then be decomposed as

$$\|H(z) - \hat{H}(z)\|_{S_u}^2 = \sum_{k=0}^{M+L} (\rho_k - v_k)^2 + \sum_{k=1}^{\infty} \tau_k^2. \quad (10)$$

The particular case in which $H(z) = 1/H_I(z)$ and $G(z)H(z) = 1$ results in $\tau_k = \langle 1, z^k V(z) \bar{F}(z)/F(z) \rangle \equiv 0$ and, as seen, is equivalent to the inverse identification of $H_I(z)$, assuming $H_I(z)$ is minimum-phase. As discussed in more detail in [6], the decomposition of the mean square error obtained in this case is not very useful. The case $G(z)H(z) \neq 1$ would deserve further attention, but we postpone it at this moment. Instead, generalizing the procedure in [6],[7] we consider the decomposition of the maximum-phase component $\bar{E}(z)/F(z)$ of $G(z)$, where $\bar{E}(z) = z^L E(z^{-1})$, multiplied by $H(z)$ or $\hat{H}(z)$, products which are also in $\mathcal{H}_2$:

$$\frac{\bar{E}(z)}{F(z)} \hat{H}(z) = \sum_{k=0}^{M+L} \bar{v}_k \frac{P_k(z)}{A(z)F(z)}, \quad (11)$$

$$\begin{aligned} \frac{\bar{E}(z)}{F(z)} H(z) &= \sum_{k=0}^{M+L} \bar{\rho}_k \frac{P_k(z)}{A(z)F(z)} + \\ &= + \sum_{k=1}^{\infty} \bar{\tau}_k z^k V(z) \frac{\bar{F}(z)}{F(z)}. \end{aligned} \quad (12)$$

Now, since $\bar{E}(z)/E(z)$ is all-pass it does not affect the mean square error and we can write then

$$\|H(z) - \hat{H}(z)\|_{S_u}^2 = \sum_{k=0}^{M+L} (\bar{\rho}_k - \bar{v}_k)^2 + \sum_{k=1}^{\infty} \bar{\tau}_k^2, \quad (13)$$

where the coefficients $\bar{\tau}_k$ are given by

$$\begin{aligned} \bar{\tau}_k &= \left\langle \frac{\bar{E}(z)}{F(z)} H(z), z^k V(z) \frac{\bar{F}(z)}{F(z)} \right\rangle \\ &= \left\langle \frac{E(z^{-1})}{F(z^{-1})} H(z), z^k V(z) \right\rangle. \end{aligned}$$

Moreover, since $z^k V(z)$ is (strictly) causal we can write $\bar{\tau}_k = \langle H_p(z), z^k V(z) \rangle$, where

$$H_p(z) \triangleq \left[G(z^{-1}) H(z) \right]_{\oplus}. \quad (14)$$

Synthesizing, then:

Property 1 With a rational spectrum input signal, the mean square error when identifying $H(z)$ can be written as

$$\|H(z) - \hat{H}(z)\|_{S_u}^2 = \sum_{k=0}^{M+L} (\bar{\rho}_k - \bar{v}_k)^2 + \sum_{k=1}^{\infty} \langle H_p(z), z^k V(z) \rangle^2, \quad (15)$$

where $H_p(z)$ is given by (14).

Comparing with (5), we note that the first part of this decomposition has $M+L$ terms instead of M terms and the second part involves $H_p(z)$ instead of $H(z)$. We note also that $H_p(z)$ is equivalent to the function associated to the weighted Hankel form considered in [3, p. 623], where it is shown that from $G(z) \in \mathcal{H}_2$ being minimum-phase it follows that

$$\deg[H_p(z)] = \deg[H(z)].$$

That is, the transformation in (14) is degree-preserving. Finally, the function $H_p(z) = [H_I(z^{-1})/H_I(z)]_{\oplus}$ considered in [6],[7] (this particular form suggested in [9]) is a special case of (14) with, as already discussed, $H(z) = 1/H_I(z)$ and $G(z) = H_I(z)$.

3.3 The restricted error surface

In the white input case, the assumption that the zeros of $\hat{H}(z)$ are optimized leads to (6) which, in turn, was used to analyze the convergence speed properties discussed in [4],[5]. We will consider now a similar procedure, starting from (15).

We note, initially, that if the points for which, for all adapted parameters p_j ,

$$\frac{\partial}{\partial p_j} \sum_{k=0}^{M+L} (\bar{v}_k - \bar{\rho}_k)^2 = 0 \quad (16)$$

form a manifold in the parameter space, and if the adaptation takes place on this manifold, then, from (15), all the convergence speed properties of a true gradient algorithm already obtained for the white input case would apply with $H_p(z)$ being used in place of $H(z)$. At this moment we don't know a necessary and sufficient condition for (16), so we consider in the following a condition which is sufficient but maybe not necessary.

Clearly, if

$$\bar{v}_k = \bar{\rho}_k, \quad k = 0, 1, \dots, M+L, \quad (17)$$

then (16) is satisfied. In this case, we will use the term "restricted error surface" to refer to the mean square error $\|H(z) - \hat{H}(z)\|_{S_u}^2 = \sum_{k=1}^{\infty} \langle H_p(z), z^k V(z) \rangle^2$ as a function of the pole parameters. We note that the restricted error surface is contained in the reduced error surface.

Now, from (11) and (12), if (17) is true then

$$\frac{\bar{E}(z)}{F(z)} [H(z) - \hat{H}(z)] = zQ(z)V(z) \frac{\bar{F}(z)}{F(z)}, \quad (18)$$

for some $Q(z) \in \mathcal{H}_2$. With $H(z) = C(z)/D(z)$, $\hat{H}(z) = B(z)/A(z)$ and $V(z) = \bar{A}(z)/A(z)$ it follows that $Q(z)$ has the form $R(z)/D(z)$, for some $R(z)$. If we also rewrite $\bar{E}(z)$ as

$$\bar{E}(z) = z^l \bar{E}'(z), \quad \text{where } l = \begin{cases} 0, & \text{if } \deg[E(z)] = L \\ 1, & \text{if } \deg[E(z)] < L \end{cases},$$

then we can rewrite (18) as

$$z^l \bar{E}'(z) C(z) A(z) - z^l \bar{E}'(z) B(z) D(z) = zR(z) \bar{F}(z) \bar{A}(z). \quad (19)$$

Assuming now $\bar{A}(z)$ and $\bar{E}'(z)$ have exactly K roots in common, we can group these roots in $\bar{P}(z)$ and write $\bar{A}(z) = \bar{P}(z) \bar{A}'(z)$, where $\deg[\bar{A}'(z)] = M - K$ and $A(z) = P(z) A'(z)$, where $\deg[A'(z)] \leq M -$

K . Proceeding similarly, we can also write $\overline{E}'(z) = \overline{P}(z)\overline{E}''(z)$ and $E'(z) = E''(z)P(z)$. Using these expressions in (19) simplifies it to

$$\begin{aligned} z^l \overline{E}''(z)C(z)P(z)A'(z) - z^l \overline{E}''(z)B(z)D(z) = \\ = zR(z)\overline{F}(z)\overline{A}'(z). \end{aligned} \quad (20)$$

Since none of the roots of $\overline{E}''(z)$ can be shared by $\overline{A}'(z)$ (by construction) or $\overline{F}(z)$ (since $\deg[G(z)] = L$ is assumed), it follows that $R(z)$ has the form $T(z)\overline{E}''(z)$ and (20) simplifies further to

$$C(z)P(z)A'(z) - B(z)D(z) = z^{1-l}T(z)\overline{F}(z)\overline{A}'(z). \quad (21)$$

Since the degrees of the left and right hand sides of this expression don't exceed $M+N$ and $M-K+L+1-l+\deg[T(z)]$, respectively, it follows that

$$\deg[T(z)] \leq N+K-L-1+l. \quad (22)$$

Writing (21) in matrix form we have

$$\Theta_{CP} \begin{bmatrix} 1 \\ \mathbf{a}' \end{bmatrix} - \Theta_D \mathbf{b} = \Theta_{T\overline{F}} \Pi \begin{bmatrix} 1 \\ \mathbf{a}' \end{bmatrix},$$

where Θ_{CP} and Θ_D are the $(M+N+1) \times (M-K+1)$ convolution matrices formed from the coefficients of $C(z)P(z)$ and $D(z)$ respectively, $\Theta_{T\overline{F}}$ is the $(M+N+1) \times (M-K+1)$ convolution matrix formed from the coefficients of $z^{1-l}T(z)\overline{F}(z)$, Π is the antidiagonal permutation matrix, and $\mathbf{a}'$ and $\mathbf{b}$ are the vectors formed from the coefficients of $A'(z)$ and $B(z)$ respectively. Separating their first columns we can write $\Theta_{CP} = [\theta_{CP} \ \Phi_{CP}]$ and $\Theta_{T\overline{F}} \Pi = [\theta_{T\overline{F}} \ \Phi_{T\overline{F}}]$. We arrive then at the following system of equations:

$$[-\theta_D \quad \Phi_{CP} - \Phi_{T\overline{F}}] \begin{bmatrix} \mathbf{b} \\ \mathbf{a}' \end{bmatrix} = \theta_{T\overline{F}} - \theta_{CP}. \quad (23)$$

For each $T(z)$, (23) is a system of $M+N+1$ equations and $2M-K+1$ unknowns. Therefore, as long as

$$M \geq N+K,$$

the solutions to this system are a linear space of $M-N-K$ dimensions. Assuming, additionally, from (22), that $T(z)$ has

$$N+K-L+1 > 0$$

degrees of freedom, the solution of the original problem (17), or in other terms, the domain of the restricted error surface, is a set of manifolds, each one with

$$(M-N-K) + (N+K-L+1) = M-L+1 \quad (24)$$

dimensions. Each manifold is associated to a particular set of roots shared by $\overline{A}(z)$ and $\overline{E}'(z)$.

As discussed in Section 2.1, adopting $M = N+L$ has the advantage of ruling out local minima. In this case, with (24), the dimension of the manifolds above is $M-L+1 = N+1$.

3.4 Numerical example

Let us now apply the ideas discussed in this section to the non-white input case of the example presented in Subsection 3.1. Initially, the zero/pole diagrams of $H(z)$ and $H_p(z)$ are shown in Fig. 3. It can be seen that the effect of $G(z)$ in this case is to make one pair of zeros of $H_p(z)$ almost cancel one pair of its poles.

In Fig. 4 we show the mean square error $E[e^2(n)]$ that results from the true gradient algorithm, employing the same adaptation steps as the stochastic gradient algorithm. As seen, it is a good estimate of the squared error $e^2(n)$ resulting from the stochastic gradient algorithm. Moreover, as also shown in the figure, the part

Figure 3: Zeros ("o") and poles ("x") of $H(z^{-1})$ (left) and $H_p(z^{-1})$ (right).

Figure 4: $E[e^2(n)]$ and $e^2(n)$ of true and stochastic gradient algorithms (top traces), respectively, and the part $\sum_{k=0}^{M+L} (\overline{p}_k - \overline{v}_k)^2$ of $E[e^2(n)]$ (bottom trace).

$\sum_{k=0}^{M+L} (\overline{p}_k - \overline{v}_k)^2$ in (15) of $E[e^2(n)]$ is quite smaller than the total (more than 20 dB, that is, less than 1 %). In other terms, the adaptation occurs close to the restricted error surface. As discussed above, this suggest that in this case the model (7) is a good approximation of the adaptation process, with $H_p(z)$ used in place of $H(z)$.

Indeed, the model (7) accounts for the slow convergence that is observed: the eigenvalue spread of $\mathbf{J}\mathbf{J}'\Sigma^2$ is quite large (the eigenvalues are almost constant during the adaptation, with $\lambda_1 = 3.2$, $\lambda_2 = 2.2$, $\lambda_3 = 0.032$ and $\lambda_4 = 0.005$) and $\alpha(n)$ is mainly in the directions of the eigenvectors associated to the smaller eigenvalues, as shown in Fig. 5. This suggests that model (7) can be used to establish conditions that contribute to a faster convergence, which is carried out in the next section.

Figure 5: Components $\beta_k(n)$ of $\alpha(n)$ in the directions of the eigenvectors associated to the eigenvalues $\lambda_k(n)$.

4. SMALL SINGULAR VALUE SPREAD CONDITION

4.1 Condition on the zeros of $H(z)$

In our model (7) of the adaptation, if the Hankel singular values of $H_p(z)$ are all equal this will tend to reduce the spread of the eigenvalues of $\mathbf{J}\mathbf{J}^H\mathbf{\Sigma}^2$. Initially, we note that with $H(z) = C(z)/D(z)$ and since $G(z^{-1})$ is anti-causal, $H_p(z)$ has the form

$$H_p(z) \triangleq \left[G(z^{-1})H(z) \right]_{\oplus} = \frac{S(z)}{D(z)},$$

where $\deg[S(z)] \leq N$. So, if

$$S(z) = s [\overline{D}(z) + cD(z)] \quad (25)$$

for any c , all the Hankel singular values of $H_p(z)$ will be equal to s . To find a $C_s(z)$ such that $H_s(z) = C_s(z)/D(z)$ leads to $S(z)$ with the form above, let us separate

$$\begin{aligned} G(z^{-1})H(z) &= \frac{E(z^{-1}) C_s(z)}{F(z^{-1}) D(z)} = \frac{z^L E(z^{-1}) C_s(z)}{z^L F(z^{-1}) D(z)} \\ &= \frac{\overline{E}(z) C_s(z)}{\overline{F}(z) D(z)} \end{aligned}$$

in its causal and strictly anti-causal parts, writing

$$\frac{\overline{E}(z) C_s(z)}{\overline{F}(z) D(z)} = \frac{S(z)}{D(z)} + \frac{\overline{Q}(z)}{\overline{F}(z)}, \quad (26)$$

for some $\overline{Q}(z) = z^L Q(z^{-1})$, where $\deg[Q(z)] < L$. Substituting (25) in (26), and with some algebraic manipulation we get

$$C_s(z) \overline{E}(z) - D(z) \overline{Q}(z) = s [\overline{D}(z) + cD(z)] \overline{F}(z),$$

which can be written in matrix form as

$$\begin{bmatrix} \Theta_{\overline{E}} & -\Theta_D \end{bmatrix} \begin{bmatrix} \mathbf{c} \\ \mathbf{q} \end{bmatrix} = \mathbf{b}, \quad (27)$$

where $\Theta_{\overline{E}}$ is the $(N+L+1) \times (N+1)$ convolution matrix formed from the coefficients of $\overline{E}(z)$, Θ_D is the $(N+L+1) \times L$ convolution matrix formed from the coefficients of $D(z)$, $\mathbf{c}$ contains the $N+1$ coefficients of $C_s(z)$, $\mathbf{q}$ contains the L coefficients of $\overline{Q}(z)$ and $\mathbf{b}$ contains the $N+L+1$ coefficients of $s [\overline{D}(z) + cD(z)] \overline{F}(z)$. Solving the system gives the $C_s(z)$ we seek.

4.2 Examples

Let us return to the non-white input case of the example presented in Subsection 3.1, for which we compute $C_s(z)$ according to (27), with $s = 0.5$ and $c = 0$. The result of the identification of $H_s(z)$ with the stochastic gradient algorithm with adaptation steps $\mu_a = 0.02$ and $\mu_b = 0.04$, and initial parameters $b_k = 0.4 \times c_{s,k}$ and $a_k = 0.4 \times d_k$ is shown in Fig. 2(C). It can be seen that, indeed, the convergence is much faster for $H_s(z)$ than for $H(z)$, although not as fast as in the white input case.

5. SUCCESSIVE APPROXIMATIONS ALGORITHM

The condition for small Hankel singular value spread obtained in the previous section can be used now to extend for rational spectrum inputs the Successive Approximations (SA) algorithm, which was originally presented in [4] for the white input case. The resulting procedure is :

1) Starting from an initial estimate $H_*(z) = C_*(z)/D_*(z)$ of $H(z) = C(z)/D(z)$, $C_s(z)$ is calculated with (27) and an auxiliary block $H_a(z) = [C_*(z) - C_s(z)]/D_*(z)$ is used in parallel with the adaptive IIR filter, which is initialized at $C_s(z)/D_*(z)$.

(Observation: with such an auxiliary block, if $H_*(z) \approx H(z)$ then $H(z) - H_a(z) \approx C_s(z)/D_*(z)$, which has equal singular values, and the adaptation of $\hat{H}(z)$ will tend to be fast, as discussed in the previous section. Differently from [4], a direct form adaptive filter is used for comparison with previous simulations.)

2) After a certain number n_a of samples, the transfer function $\hat{H}(z)$ of the adaptive filter is used to obtain a new estimate $H_*(z)$ such that $\|\hat{H}(z) + H_a(z) - H_*(z)\|^2$ is approximately minimized. This computation is carried out in a certain number n_x of samples, and with the new $H_*(z)$ we go back to Step 1.

The result of the SA algorithm with $n_a = 2000$, $n_x = 200$ and applied to the example as above is presented in Fig. 2(D). It can be seen that once in the neighbourhood of the optimum it converges with the same rate of the identification of $H_s(z)$, as would be expected.

6. CONCLUSION

The analytical framework presented in [4],[5] was extended to the more general case of rational spectrum inputs, which also allowed a unification with the inverse identification of a minimum-phase system, considered in [6],[7]. Desirable further developments would be a deeper understanding of the role of the restricted error surface, which was introduced in this work, and inclusion of the non-minimum phase case of inverse identification.

REFERENCES

- [1] J. Treichler, C. R. Johnson, Jr., and M. G. Larimore, *Theory and design of adaptive filters*, Prentice-Hall, Upper Saddle River, 2001.
- [2] J. J. Shynk, "Adaptive IIR filtering," *IEEE ASSP Magazine*, vol. 6, no. 2, pp. 4–21, Apr. 1989.
- [3] P. A. Regalia, *Adaptive IIR filtering in signal processing and control*, Marcel Dekker, New York, 1995.
- [4] P. M. S. Burt and P. A. Regalia, "A new framework for convergence analysis and algorithm development of adaptive IIR filters," in *Proc. IEEE ICASSP*, Montreal, 2004, vol. 2, pp. 441–444.
- [5] P. M. S. Burt and P. A. Regalia, "A new framework for convergence analysis and algorithm development of adaptive IIR filters," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3129–3140, Aug. 2005.
- [6] P. M. S. Burt, "On the convergence speed of inverse identification adaptive IIR filters," in *Proc. SBrT/IEEE International Telecommunications Symposium*, Fortaleza, 2006, pp. 51–56.
- [7] P. M. S. Burt, "Inverse identification adaptive IIR filtering: convergence speed analysis and successive approximations algorithm," in *Proc. IEEE Icassp*, Honolulu, 2007.
- [8] T. Söderström and P. Stoica, "Some properties of the output error method," *Automatica*, vol. 18, pp. 93–99, 1982.
- [9] M. Mboup, (*private communication*), 2007.