

PHILLIP M. S. BURT

**UM ESTUDO SOBRE FILTROS IIR ADAPTATIVOS  
COM APLICAÇÃO A UMA ESTRUTURA  
POLIFÁSICA**

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

São Paulo, abril de 1997

PHILLIP M. S. BURT

**UM ESTUDO SOBRE FILTROS IIR ADAPTATIVOS  
COM APLICAÇÃO A UMA ESTRUTURA  
POLIFÁSICA**

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

**Área de Concentração:**  
Engenharia de Sistemas Eletrônicos

**Orientador:**  
Prof. Dr. Max Gerken

São Paulo, abril de 1997

## **Agradecimentos**

Ao Prof. Dr. Max Gerken, pela orientação e sugestões - em particular, a de se investigar a estrutura polifásica - pela leitura meticulosa das versões finais do trabalho e pela ajuda.

Ao Prof. Dr. Jacyntho J. Angerami e ao Prof. Dr. Paul Jean E. Jeszensky, por todo o apoio recebido.

Aos colegas Maria D. Miranda, Luiz Antonio Baccalá e Luiz Antonio B. Coelho, pelo incentivo e ajuda.

A toda minha família, pelo constante estímulo e apoio.

A Tânia, Patricia e Renata, que compreenderam minha ausência e a quem este trabalho é dedicado.

# RESUMO

Neste trabalho faz-se um estudo sobre filtros IIR adaptativos e é apresentada uma estrutura polifásica para filtragem IIR adaptativa, que, em troca de um aumento de complexidade computacional, pode apresentar características mais favoráveis do que a estrutura direta comumente usada. O aumento da complexidade computacional, relativamente a um algoritmo do tipo Newton, por exemplo, é pequeno.

Apresenta-se uma análise dos efeitos da proximidade ao círculo unitário dos polos do sistema sendo modelado. Um dos efeitos considerados é o comportamento limite do condicionamento da matriz de estados associada ao algoritmo de adaptação. São considerados algoritmos de adaptação de passo constante de uso comum para filtros IIR adaptativos. O método utilizado é particularmente útil para a verificação do efeito da posição dos polos do sistema sendo modelado e também para a introdução de certas restrições ao mesmo, como por exemplo, norma  $L_2$  fixa e resposta em frequência passa-tudo. Um resultado interessante é que a única situação, entre as testadas, em que o condicionamento da matriz mencionada não tende a infinito quando um número qualquer de polos do sistema sendo modelado  $H(z)$  se aproxima da circunferência unitária, é quando  $H(z)$  é passa-tudo e emprega-se o algoritmo PLR.

São analisadas também a superfície de erro e a superfície de erro reduzida para filtros IIR adaptativos. Mostra-se que, quando o sistema sendo modelado possui polos próximos à circunferência unitária, a superfície de erro reduzida apresenta regiões planas com erro quadrático médio elevado. A existência destas regiões resulta em uma baixa velocidade de convergência global de algoritmos de passo constante.

A partir da decomposição em valores singulares (SVD) da forma de Hankel do sistema sendo modelado, é apresentada também uma decomposição da superfície de erro reduzida, a partir da qual pode-se obter uma separação parcial dos efeitos do sistema sendo modelado e da forma de realização do filtro adaptativo.

Uma estrutura polifásica para filtragem IIR adaptativa é apresentada e seu desempenho é comparado com o de filtros IIR adaptativos na forma direta. Mostra-se o possível ganho da estrutura polifásica quanto à velocidade de convergência local e quanto às características



da superfície de erro reduzida e à velocidade de convergência global. Demonstra-se, para a estrutura polifásica, que, com entrada branca e modelamento suficiente, todos os pontos estacionários da superfície de erro são mínimos globais da mesma. Este resultado não decorre diretamente de propriedades análogas relativas à estrutura direta, já conhecidas.

Tanto para a estrutura direta quanto para a estrutura polifásica, são apresentados os resultados de várias simulações dos algoritmos de adaptação considerados.

# ABSTRACT

A study on IIR adaptive filters and a polyphase structure for IIR adaptive filtering are presented. In exchange for an increase in computational complexity, which is small if compared to Newton algorithms, the polyphase structure may exhibit a better performance than direct structures.

An analysis of the effects of the proximity to the unit circle of the modelled system's poles is presented. One of the considered points is the limiting behaviour of the condition of the state matrix related to the adaptive algorithm. Commonly used constant gain algorithms are considered. The method of analysis is specially useful for verifying the effects of the position of the system's poles and also for introducing certain restrictions to the system, as fixed  $L_2$  norm and all-pass frequency response. An interesting result is that, among the situations that were tested, the only one in which the condition of the aforementioned matrix does not tend to infinity as the poles of the modelled system  $H(z)$  tend to the unit circle is when  $H(z)$  is all-pass and the PLR algorithm is employed.

The error surface and the reduced error surface for IIR adaptive filters are also analysed. It is shown that when the modelled system has poles close to the unit circle the reduced error surface presents flat regions with high mean square error. The presence of these flat regions results in low global convergence speed for constant gain adaptive algorithms.

Based on the singular value decomposition (SVD) of the modelled system's Hankel form, a decomposition of the reduced error surface is also presented. In it there exists a partial separation of the effects of the system and the adaptive filter's structure.

A polyphase structure for IIR adaptive filtering is presented and its performance is compared to the performance of the direct structure. The gain in local convergence and global convergence speed, as well as the better behaviour of the reduced error surface which may be attained, are shown. It is demonstrated, for the polyphase structure, that, with white input and sufficient modelling, all the stationary points of the error surface are global minima. This result does not follow directly from similar well known results for the direct structure.

Simulation results for the considered algorithms are also presented.

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>5</b>
1.1	Resumo dos capítulos . . . . .	6
1.2	Contribuições apresentadas . . . . .	7
1.3	Formulação genérica do problema de filtragem adaptativa . . . . .	9
1.4	Aplicações de filtros IIR adaptativos . . . . .	10
1.4.1	Identificação de sistemas . . . . .	10
1.4.2	Cancelamento de eco . . . . .	13
1.4.3	Equalização adaptativa de canal em transmissão digital . . . . .	15
1.4.4	Cancelamento de ruído sonoro . . . . .	16
1.5	Algumas definições e propriedades . . . . .	17
1.5.1	Espaços e propriedades de funções em $z$ . . . . .	17
1.5.2	Propriedades de autovalores e valores singulares . . . . .	21
1.5.3	Excitação persistente . . . . .	23
1.5.4	Ortogonalidade a um vetor de regressão . . . . .	23
<b>2</b>	<b>Características de algoritmos de adaptação para filtros IIR na forma direta</b>	<b>26</b>
2.1	Notação para filtros com parâmetros variantes no tempo . . . . .	28
2.2	Convergência local de algoritmos de passo constante . . . . .	29
2.2.1	Formulação genérica de algoritmos de adaptação e equação diferencial associada . . . . .	29
2.2.2	Condição suficiente para convergência local . . . . .	30
2.2.3	Limites para o passo de adaptação . . . . .	36
2.2.4	Velocidade de convergência local . . . . .	42
2.3	Algoritmo do gradiente recursivo . . . . .	48

2.3.1	Obtenção do algoritmo . . . . .	49
2.3.2	Aplicação à estrutura direta de implementação . . . . .	51
2.3.3	Condição suficiente para convergência . . . . .	53
2.3.4	Sobremodelamento . . . . .	56
2.3.5	Influência da posição dos polos sobre a região de positividade e o domínio de atração . . . . .	56
2.4	Algoritmo Steiglitz-McBride . . . . .	58
2.4.1	Obtenção do algoritmo recursivo . . . . .	58
2.4.2	Condição para convergência . . . . .	62
2.5	Algoritmos PLR e SHARF . . . . .	64
2.5.1	Expressões de adaptação e condição para convergência do algoritmo PLR . . . . .	64
2.5.2	Influência da posição dos polos sobre a região de positividade . . . . .	67
2.5.3	Expressões de adaptação e condição para convergência do algoritmo SHARF . . . . .	67
2.6	Efeito da posição dos polos de $H(z)$ no condicionamento da matriz de estados da ODE . . . . .	70
2.6.1	Expressões para a análise dos valores singulares . . . . .	70
2.6.2	Efeito do sinal de entrada . . . . .	73
2.6.3	Restrições a $H(z)$ . . . . .	74
2.6.4	Valores singulares extremos de $\mathbf{R}(\mathbf{w}^*)$ considerando a Restrição 1 . . . . .	75
2.6.5	Valores singulares extremos de $\mathbf{R}(\mathbf{w}^*)$ considerando a Restrição 2 . . . . .	80
2.6.6	Valores singulares extremos de $\mathbf{R}(\mathbf{w}^*)$ considerando a Restrição 3 . . . . .	82
2.6.7	Resumo dos resultados da análise de valores singulares . . . . .	83
2.6.8	Exemplos numéricos do comportamento dos valores singulares . . . . .	84
2.7	Exemplos da simulação dos algoritmos de adaptação . . . . .	89
2.7.1	Verificação da análise local . . . . .	104
2.7.2	Discussão dos resultados . . . . .	119
<b>3</b>	<b>Análise da superfície de erro reduzida para filtros IIR adaptativos</b> . . . . .	<b>127</b>
3.1	Tópicos de sistemas lineares . . . . .	128
3.1.1	Matriz e gramiano de controlabilidade . . . . .	128

3.1.2	Vetor de controlabilidade e complemento passa-tudo . . . . .	129
3.2	Análise da superfície de erro com entrada branca . . . . .	131
3.2.1	Forma do erro de aproximação . . . . .	131
3.2.2	Modelamento suficiente . . . . .	134
3.2.3	Superfície de erro reduzida . . . . .	134
3.3	Outras formas de implementação . . . . .	136
3.3.1	Gradiente e Hessiana após mudança de variáveis . . . . .	136
3.3.2	Aproximação local da superfície de erro . . . . .	140
3.3.3	Extensão de propriedades do erro de aproximação . . . . .	140
3.4	Análise SVD da superfície de erro reduzida . . . . .	141
3.4.1	Decomposição SVD de $\ g(z)\ $ . . . . .	141
3.4.2	Forma da Hessiana no mínimo global . . . . .	144
3.5	Diagonalização da Hessiana . . . . .	147
3.5.1	Realização direta . . . . .	147
3.5.2	Uma realização hipotética ideal . . . . .	148
3.6	Sistemas com polos próximos ao círculo unitário . . . . .	149
3.7	Exemplos da simulação de algoritmos de adaptação . . . . .	153
<b>4</b>	<b>Um filtro IIR polifásico adaptativo</b>	<b>178</b>
4.1	Expansão polifásica de uma função racional . . . . .	179
4.1.1	Determinação da expansão polifásica pela descrição no espaço de estados	179
4.1.2	Possível não-unicidade da expansão polifásica . . . . .	182
4.2	Filtro polifásico adaptativo . . . . .	185
4.2.1	Abordagens para filtragem IIR polifásica adaptativa . . . . .	185
4.2.2	Forma de $\hat{H}(z)$ . . . . .	187
4.2.3	Algoritmos de passo constante . . . . .	190
4.2.4	Complexidade computacional de filtros adaptativos polifásicos . . . . .	192
4.3	Análise da convergência e do condicionamento da matriz de estados . . . . .	195
4.3.1	Matrizes de estado da ODE no caso polifásico . . . . .	195
4.3.2	Condições para convergência e domínio de atração . . . . .	198
4.3.3	Expressões para a análise dos valores singulares de $\mathbf{R}_P(\mathbf{w}_P^*)$ . . . . .	201
4.3.4	Valores singulares extremos de $\mathbf{R}_P(\mathbf{w}_P^*)$ para os algoritmos RG e SMM	205

4.3.5	Valores singulares extremos de $\mathbf{R}_P(\mathbf{w}_P^*)$ para o algoritmo PLR . . . . .	212
4.3.6	Valores singulares extremos de $\mathbf{R}_P(\mathbf{w}_P^*)$ para o algoritmo SHARF . . . . .	215
4.3.7	Exemplos numéricos do comportamento dos valores singulares . . . . .	215
4.4	Análise dos pontos estacionários da superfície de erro (entrada branca) . . . . .	216
4.4.1	Forma polifásica do erro de aproximação . . . . .	219
4.4.2	Otimização dos zeros com os polos fixados . . . . .	220
4.4.3	Otimização dos zeros e polos . . . . .	223
4.4.4	Forma das funções $g_l(z)$ . . . . .	225
4.4.5	Pontos estacionários da superfície de erro . . . . .	227
4.5	Efeito da posição dos polos na superfície de erro reduzida . . . . .	231
4.5.1	Expressão para superfície de erro reduzida . . . . .	231
4.5.2	Efeito da posição dos polos de $H(z)$ . . . . .	232
4.6	Exemplos da simulação dos algoritmos de adaptação para a estrutura polifásica	233
4.6.1	Comparação com estrutura direta: ponto inicial fora da origem . . . . .	261
4.6.2	Comparação com estrutura direta: ponto inicial na origem . . . . .	264
<b>A</b>	<b>Expressões usadas na análise local de algoritmos de adaptação</b>	<b>266</b>
A.1	Dedução de (2.156) . . . . .	266
A.2	Dedução de (2.161) . . . . .	268
<b>B</b>	<b>Lista dos principais símbolos empregados</b>	<b>272</b>

# Capítulo 1

## Introdução

Técnicas de processamento digital de sinais tem sido crescentemente utilizadas em diversos campos de atividade, tais como telecomunicações, automação e controle de processos industriais, radar e sonar, engenharia biomédica e sismologia[12]. Em todas estas aplicações é comum não se dispor de um conhecimento prévio das características dos sinais envolvidos, o que exige o emprego de técnicas adaptativas de processamento.

A utilização de filtros FIR (resposta ao pulso unitário de duração finita) adaptativos já é uma técnica bastante evoluída [12], [15], [2]. Tem o inconveniente, porém, de modelar o sistema real envolvido na aplicação como um sistema cuja função de transferência só possui zeros e não polos. Consequentemente, se o sistema real tiver polos pouco amortecidos será necessário um grande número de coeficientes para o filtro FIR adaptativo, o que reduz a eficiência geral da solução. Nestes casos, filtros IIR (resposta ao pulso unitário de duração infinita) adaptativos são potencialmente mais eficientes, por possuírem função de transferência com polos e zeros. Por outro lado, apresentam aspectos problemáticos como o da velocidade de convergência e o controle de estabilidade. Além disso, o seu tratamento analítico é consideravelmente mais difícil do que o de filtros FIR adaptativos, existindo aspectos importantes relativamente pouco explorados, como o das propriedades da superfície de erro. Devido a este fatores, o uso de filtros IIR adaptativos é ainda relativamente restrito.

Neste trabalho faz-se um estudo sobre filtros IIR adaptativos e é apresentada uma estrutura polifásica para filtragem IIR adaptativa, que, em troca de um aumento de complexidade computacional, apresenta características mais favoráveis do que a estrutura direta comumente usada. Nos dois itens a seguir é feito um resumo dos capítulos do trabalho e são destacadas as contribuições apresentadas (No início de cada capítulo há um resumo mais

detalhado do mesmo). Nos itens subsequentes algumas aplicações práticas importantes de filtros IIR adaptativos são enquadradas em um mesmo problema genérico, o qual servirá então de ponto de partida para as análises realizadas nos demais capítulos da tese. São introduzidas também certas definições e propriedades a serem utilizadas.

## 1.1 Resumo dos capítulos

No Capítulo 2 são descritos algoritmos de adaptação comumente usados para filtros IIR adaptativos implementados na forma direta e seu comportamento é analisado. São considerados: o algoritmo do gradiente recursivo (RG), o algoritmo baseado no método de identificação Steiglitz-McBride (SMM), o algoritmo baseado na regressão pseudo-linear (PLR) e o algoritmo SHARF. Apresenta-se uma análise do efeito da proximidade ao círculo unitário dos polos do sistema sendo modelado. São apresentadas várias simulações da execução dos algoritmos de adaptação em questão, onde são exemplificados os aspectos analisados do ponto de vista teórico nos itens precedentes.

No Capítulo 3 são analisadas a superfície de erro e a superfície de erro reduzida para filtros IIR adaptativos. Para algoritmos de adaptação do tipo gradiente, como o algoritmo RG as propriedades destas superfícies tem relações diretas tanto com a sua convergência global (ponto inicial distante do ótimo) quanto com a sua convergência local (nas imediações do ponto ótimo). Para outros tipos de algoritmos estas relações não são tão diretas, porém, pode-se observar, na prática, que são importantes: por exemplo, em regiões onde a superfície de erro é excessivamente plana a convergência de todos os algoritmos considerados é lenta. No Item 3.4, a partir da decomposição em valores singulares (SVD) da forma de Hankel do sistema sendo modelado, é apresentada uma decomposição da superfície de erro reduzida. A partir desta decomposição outros resultados são obtidos. No Item 3.6, mostra-se que, quando o sistema sendo identificado possui polos próximos à circunferência unitária, a superfície de erro reduzida apresenta regiões planas com erro quadrático médio elevado. A existência destas regiões resulta em uma baixa velocidade de convergência global de algoritmos de passo constante, como exemplificado pelos resultados de simulações apresentados no Item 4.5.

No Capítulo 4 apresenta-se uma estrutura polifásica para filtragem IIR adaptativa e seu desempenho é comparado com o de filtros IIR adaptativos na forma direta. No Item 4.1



a expansão polifásica de uma função racional é formulada e é analisada a questão de sua unicidade. No Item 4.3 a análise da convergência local de algoritmos de passo constante para filtros IIR adaptativos na forma direta, realizada nos Itens 2.2 a 2.6, é estendida para filtros IIR adaptativos na forma polifásica. Mostra-se como as condições de positividade relacionadas à convergência dos algoritmos podem ser menos restritivas no caso polifásico. A análise permite também uma compreensão do possível ganho da estrutura polifásica quanto à velocidade de convergência local. Em seguida passa-se a analisar a superfície de erro reduzida para o caso da estrutura polifásica. No Item 4.4 demonstra-se, para a estrutura polifásica, que, com entrada branca e modelamento suficiente, todos os pontos estacionários da superfície de erro são mínimos globais da mesma. Deve-se notar que esta propriedade não decorre diretamente de propriedades análogas relativas à estrutura direta, já bastante conhecidas, e trata-se portanto de um resultado novo. No Item 4.5, a partir da expressão para a superfície de erro reduzida obtida no item anterior, mostra-se a possível melhoria da mesma em relação ao caso da estrutura direta. Para um dado sistema sendo modelado  $H(z)$ , a planicidade da superfície de erro reduzida em regiões distantes do mínimo global, resultante da existência de polos de  $H(z)$  próximos à circunferência unitária, tende a ser menos pronunciada do que no caso da estrutura direta, o que resulta em uma maior velocidade de convergência global de algoritmos de passo constante. Finalmente, no Item 4.6 são mostrados exemplos práticos da aplicação da estrutura polifásica, e seu desempenho é comparado com o da estrutura direta.

## 1.2 Contribuições apresentadas

As contribuições apresentadas neste trabalho, na ordem em que aparecem no mesmo, estão relacionadas a seguir:

- No Item 2.2, considerando o caso genérico em que a matriz de estados do sistema associado ao algoritmo (matriz de estados da ODE) não é simétrica, são obtidas relações entre os valores singulares extremos desta matriz, o passo de adaptação máximo e a velocidade de convergência local do algoritmo de adaptação.
- No Item 2.2 mostra-se que a simetria desta matriz de estados favorece a validade da aproximação da convergência local do algoritmo de adaptação pelo decaimento dos modos normais de um sistema linear livre associado ao mesmo.

- No Item 2.3, mostra-se com exemplos, como, para o algoritmo RG, a proximidade à circunferência unitária dos polos do sistema sendo modelado pode reduzir a parcela que pode ser garantida do domínio de atração do mínimo global.
- No Item 2.4 é apresentada uma interpretação simples da passagem da forma *off-line* do algoritmo SMM para a sua forma recursiva.
- No Item 2.6 apresenta-se um método de análise do efeito da proximidade ao círculo unitário dos polos do sistema sendo modelado no condicionamento da matriz de estados da ODE. O método de análise apresentado é baseado em uma transposição ao domínio da frequência de certas formas quadráticas cujos limites superior e inferior correspondem aos valores singulares extremos da matriz em questão. Esta transposição ao domínio da frequência é bastante conveniente para a verificação do efeito da posição dos polos do sistema sendo modelado. É conveniente também para a introdução de certas restrições ao mesmo, como por exemplo, norma  $L_2$  fixa e resposta em frequência passa-tudo.
- A análise apresentada no Item 2.6 permite verificar que, das situações testadas, a única situação em que o espalhamento dos valores singulares da matriz de estados da ODE não tende a infinito quando um número qualquer de polos do sistema sendo modelado  $H(z)$  se aproxima da circunferência unitária, é quando  $H(z)$  é passa-tudo e emprega-se o algoritmo PLR.
- Também segue da mesma análise, que, para o algoritmo SHARF, o aparente ideal em termos do atendimento de uma condição de positividade necessária para convergência, não é, necessariamente, o ideal em termos da velocidade de convergência local do algoritmo.
- Ainda em referência ao Item 2.6, as relações obtidas entre as velocidades de crescimento ou decaimento dos valores singulares extremos da matriz de estados da ODE para os diferentes algoritmos permitem estabelecer relações entre os valores limites para estas mesmas grandezas, mesmo sem especificar completamente  $H(z)$ .
- No Item 3.4, a partir da decomposição em valores singulares (SVD) da forma de Hankel do sistema sendo modelado, é apresentada uma decomposição da superfície de erro

reduzida. A partir deste resultado, obtém-se então uma expressão para a matriz Hessiana no mínimo global da superfície de erro reduzida, sendo que nesta expressão existe uma separação parcial dos efeitos do sistema sendo modelado e da forma de realização do filtro adaptativo.

- No Item 3.6, mostra-se que quando o sistema sendo identificado possui polos próximos à circunferência unitária a superfície de erro reduzida apresenta regiões planas com erro quadrático médio elevado. A existência destas regiões resulta em uma baixa velocidade de convergência global de algoritmos de passo constante, como exemplificado pelos resultados de simulações apresentados no Item 2.7.
- No Item 4.1, os polinômios da expansão polifásica de uma função racional são obtidos a partir de uma descrição no espaço de estados da mesma.
- No Item 4.1 é analisada a questão da unicidade da expansão polifásica e é introduzido o conceito de expansão polifásica principal.
- No Item 4.3.4 é analisado o possível ganho da estrutura polifásica quanto ao condicionamento da matriz de estados associada a cada algoritmo.
- No Item 4.4 demonstra-se, para a estrutura polifásica, que, com entrada branca e modelamento suficiente, todos os pontos estacionários da superfície de erro são mínimos globais da mesma. Esta propriedade não decorre diretamente de propriedades análogas relativas à estrutura direta, já conhecidas.
- No Item 4.5, mostra-se a possível melhoria da superfície de erro reduzida para a estrutura polifásica em relação à da estrutura direta.

### 1.3 Formulação genérica do problema de filtragem adaptativa

Diferentes aplicações de filtros adaptativos podem ser formuladas em termos do seguinte problema genérico [1]: são dadas duas sequências  $u(n)$  e  $y(n)$ , relacionadas por

$$y(n) = H(z)u(n) + \zeta(n), \quad (1.1)$$

onde  $u(n)$  e  $\zeta(n)$  são sequências aleatórias estacionárias, possivelmente estatisticamente independentes entre si, e  $H(z)$  é uma função de transferência estável e causal<sup>1</sup>. Deseja-se, a partir de  $u(n)$  e  $y(n)$ , obter uma aproximação racional  $\hat{H}(z)$  de  $H(z)$ , de modo a minimizar uma determinada função custo que depende da aplicação específica.

Nos demais capítulos desta tese, a função de transferência  $H(z)$  que resulta da formulação do problema específico na forma geral dada por (1.1), é referida como a função de transferência do "sistema sendo modelado". O significado físico de  $H(z)$  dependerá da aplicação específica, como será visto a seguir.

Nos itens subsequentes, para maior simplicidade, o fato de, na prática,  $\hat{H}(z)$  ser variante no tempo (devido à natureza recursiva do processo de adaptação) não é ainda explicitado na notação. A indicação da dependência de  $\hat{H}(z)$  em relação ao tempo  $n$  é introduzida no Capítulo 2.

## 1.4 Aplicações de filtros IIR adaptativos

### 1.4.1 Identificação de sistemas

Consideremos inicialmente uma formulação genérica do problema de identificação de sistemas. Como representado na Figura 1-1,  $u(n)$  é a entrada do sistema  $S$  e  $\zeta(n)$  é uma perturbação que leva em conta a existência de entradas não conhecidas do sistema ou de ruído de medida. O sinal medido na saída é dado por

$$y(n) = S(\{u(n), \zeta(n)\}), \quad (1.2)$$

onde o conjunto  $\{u(n), \zeta(n)\}$  é formado por valores de  $u$  e  $\zeta$  em  $n$  e em instantes anteriores a  $n$  [16], [17], [18].

Formulado desta maneira genérica, o problema de identificação consiste em obter um modelo  $\hat{S}$  (pertencente a uma determinada classe de modelos admitidos) de tal modo que a estimativa

$$\hat{y}(n) = \hat{S}(\{u(n), y(n)\}) \quad (1.3)$$

aproxime  $y(n)$  de maneira ótima segundo um determinado critério, por exemplo, a mini-

---

<sup>1</sup> $H(z)$ , quando multiplica uma sequência em  $n$ , é interpretado como um operador que consiste na combinação de retardos. Notar que  $zu(n) = u(n-1)$ .

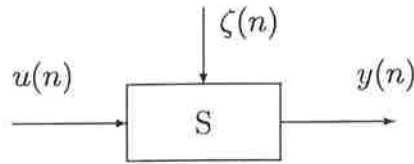


Figura 1-1: representação entrada/saída de sistema

mização do erro quadrático médio de estimação  $E[e^2(n)]$ , com  $e(n) = y(n) - \hat{y}(n)$ . Notar que, a princípio, o modelo  $\hat{S}$  leva em conta toda a informação disponível, isto é, o conjunto dos valores de entrada e saída do sistema.

No caso em que a classe de modelos admitidos for a dos sistemas lineares, a expressão (1.3) pode ser escrita como

$$\hat{y}(n) = \hat{H}_u(z)u(n) + \hat{H}_y(z)y(n). \quad (1.4)$$

Com imposições adicionais à classe de modelos podemos associar o problema de identificação aos seguintes problemas de filtros adaptativos:

- filtro FIR de uma entrada:

$$\begin{aligned} \hat{H}_u(z) &= B(z), \\ \hat{H}_y(z) &\equiv 0; \end{aligned}$$

- filtro FIR de duas entradas:

$$\begin{aligned} \hat{H}_u(z) &= B(z), \\ \hat{H}_y(z) &= 1 - A(z); \end{aligned}$$

- filtro IIR de uma entrada:

$$\hat{H}_u(z) = \frac{B(z)}{A(z)},$$

$$\widehat{H}_y(z) \equiv 0;$$

- filtro IIR de duas entradas:

$$\widehat{H}_u(z) = \frac{B(z)}{A(z)},$$

$$\widehat{H}_y(z) = \frac{F(z)}{G(z)}.$$

Os polinômios  $A(z)$  empregados nestas expressões tem a forma

$$A(z) = 1 + a_1z + \cdots + a_Mz^M,$$

e os demais tem a mesma forma que

$$B(z) = b_0 + b_1z + \cdots + b_Mz^M.$$

Muitas vezes dispõe-se de mais informações sobre a estrutura do problema do que a indicada por (1.2). Por exemplo, pode-se saber que sistema é linear e a perturbação  $\zeta(n)$  é apenas ruído de medida (não conhecido), estatisticamente independente de  $u(n)$ , ou seja, que  $y(n)$  é dado por

$$y(n) = H(z)u(n) + \zeta(n). \quad (1.5)$$

O problema se enquadra agora na formulação genérica dada por (1.1).

Neste caso é razoável que se queira usar como critério de identificação não mais o erro de estimação  $e(n) = y(n) - \widehat{y}(n)$ , como acima, mas o *erro de aproximação* definido aqui como

$$\varepsilon(n) = y(n) - \widehat{y}(n), \quad \zeta(n) \equiv 0, \quad (1.6)$$

o que equivale a dizer que, no caso geral  $\zeta(n) \neq 0$ , deseja-se aproximar em  $y(n)$  apenas a parcela devida ao sistema propriamente dito e não a parcela devida ao ruído de medida. O erro de aproximação, porém, ao contrário do erro de estimação, não é diretamente acessível para efeitos de adaptação, pois o ruído de medida  $\zeta(n)$  não é conhecido. Vejamos como isto se reflete para as diferentes classes de filtros adaptativos consideradas acima.

No caso de se usar um filtro IIR de uma entrada, fazendo  $\widehat{H}(z) = \widehat{H}_u(z)$ , temos  $\widehat{y}(n) =$

$\hat{H}(z)u(n)$  e portanto

$$e(n) = y(n) - \hat{y}(n) = [H(z) - \hat{H}(z)]u(n) + \zeta(n). \quad (1.7)$$

Com  $u(n)$  e  $\zeta(n)$  estatisticamente independentes, a função  $\hat{H}(z)$  que minimiza o erro médio quadrático de aproximação  $E[\varepsilon^2(n)]$  é a mesma que que minimiza o erro quadrático médio de estimação  $E[e^2(n)]$ . Portanto, o fato de  $\varepsilon(n)$  ser inacessível não impede a obtenção de um  $\hat{H}(z)$  que minimize seu valor quadrático médio.

Por outro lado, no caso de um filtro FIR de duas entradas, teríamos  $\hat{y}(n) = B(z)u(n) + [1 - A(z)]y(n)$  e o erro de estimação seria

$$\begin{aligned} e(n) &= y(n) - \hat{y}(n) = A(z)y(n) - B(z)u(n) \\ &= [A(z)H(z) - B(z)]u(n) + A(z)\zeta(n), \end{aligned} \quad (1.8)$$

ao passo que o erro de aproximação seria

$$\varepsilon(n) = [A(z)H(z) - B(z)]u(n). \quad (1.9)$$

As funções  $B(z)$  e  $A(z)$  que minimizam  $E[\varepsilon^2(n)]$ , portanto, não são as mesmas que minimizam  $E[e^2(n)]$ , o que constitui uma desvantagem do filtro FIR de duas entradas em relação ao filtro IIR, neste caso.

## 1.4.2 Cancelamento de eco

Nesta categoria de aplicações podemos encontrar o problema de cancelamento de eco em conexões telefônicas, representado na Figura 1-2, onde, na terminologia do item anterior, já é suposto que  $\hat{H}(z)$  é um filtro adaptativo de uma entrada. Nesta figura,  $u(n)$  é o sinal recebido da outra extremidade da conexão, a saída da híbrida é  $y(n) = H(z)u(n) + \zeta(n)$ , onde  $H(z)$  é a função de transferência de eco da híbrida e  $\zeta(n)$  é o sinal gerado localmente, e  $e(n)$  é o sinal enviado à outra extremidade. Esta formulação, portanto, se enquadra diretamente em (1.1).

O objetivo aqui é que  $\hat{y}(n) = \hat{H}(z)u(n)$  cancele a parcela  $H(z)u(n)$  refletida na híbrida, o que equivale a minimizar  $E[\varepsilon^2(n)]$  onde  $\varepsilon(n)$  é o erro de aproximação definido em (1.6). Como já está-se considerando um filtro adaptativo de uma entrada,  $\varepsilon(n)$  também pode ser

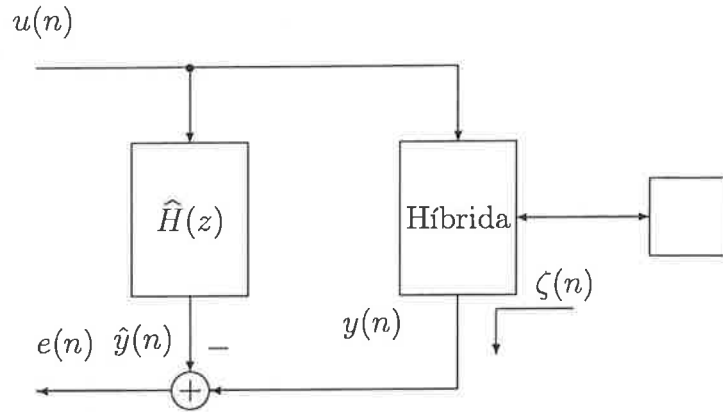


Figura 1-2: cancelamento de eco em telefonia

escrito como

$$\varepsilon(n) = [H(z) - \hat{H}(z)]u(n). \quad (1.10)$$

Com isso,  $e(n)$  deve corresponder, idealmente, apenas ao sinal local  $\zeta(n)$ . Como no item anterior, com  $u(n)$  e  $\zeta(n)$  independentes e de média nula, a minimização de  $E[\varepsilon^2(n)]$  é obtida com a minimização de  $E[e^2(n)]$ .

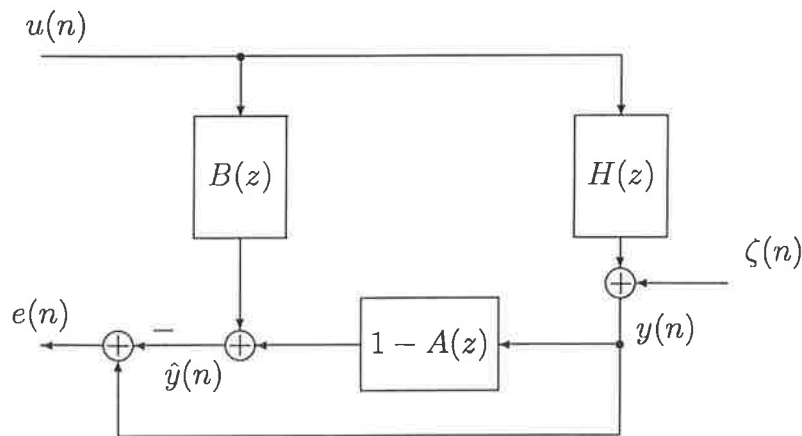


Figura 1-3: cancelamento com filtro FIR de duas entradas

Notar que, para esta aplicação, ao contrário do que foi visto no item anterior, não poderia ser usado para o cálculo de  $e(n)$ , ao invés de um filtro IIR, um filtro FIR de duas entradas, como representado na figura 1-3. Isto porque o sinal  $e(n)$  que passaria a ser mandado para a outra extremidade não conteria o sinal local  $\zeta(n)$  mas uma versão distorcida do mesmo,



$A(z)\zeta(n)$ . Por outro lado, ao introduzir-se o processo de adaptação dos coeficientes do filtro, estas classes de filtros adaptativos podem ser de certa forma misturadas: por exemplo, pode-se usar um filtro IIR de uma entrada para calcular o sinal  $e(n)$  enviado à outra extremidade, e um filtro FIR de duas entradas para obter

$$e_e(n) = A(z)y(n) - B(z)u(n), \quad (1.11)$$

que seria usado apenas no processo de adaptação dos coeficientes, que teria como objetivo então minimizar  $E[e_e^2(n)]$ . Como a adaptação é feita para um filtro FIR, alguns problemas importantes encontrados na adaptação de filtros IIR não ocorrem neste caso. Na literatura de filtros adaptativos este método é denominado "método do erro de equação" [3]. Há uma contrapartida, porém: escrevendo

$$e_e(n) = [A(z)H(z) - B(z)]u(n) + A(z)\zeta(n), \quad (1.12)$$

podemos ver que a minimização de  $E[e_e^2(n)]$  realizada no método do erro de equação não implica na desejada minimização do erro de aproximação (ver (1.10)).

### 1.4.3 Equalização adaptativa de canal em transmissão digital

Como representado na Figura 1-4,  $y(n)$  agora é a sequência de símbolos transmitidos,  $G(z)$  é a função de transferência do canal de transmissão,  $u(n)$  é a entrada do receptor,  $\eta(n)$  é o ruído de entrada do receptor e  $\hat{H}(z)$  é o equalizador adaptativo [20],[21]. Restringindo a análise aos casos em que  $G(z)$  é de fase mínima, podemos colocar o problema na forma de (1.1), escrevendo:

$$\begin{aligned} y(n) &= \frac{1}{G(z)}u(n) - \frac{1}{G(z)}\eta(n) \\ &= H(z)u(n) + \zeta(n), \end{aligned} \quad (1.13)$$

onde agora  $u(n)$  e  $\zeta(n)$  não são estatisticamente independentes. O objetivo da equalização é produzir um sinal  $\hat{y}(n)$  a partir do qual o símbolo transmitido  $y(n)$  possa ser detectado com uma baixa taxa de erro. Para tanto deve-se minimizar o erro quadrático médio de estimação

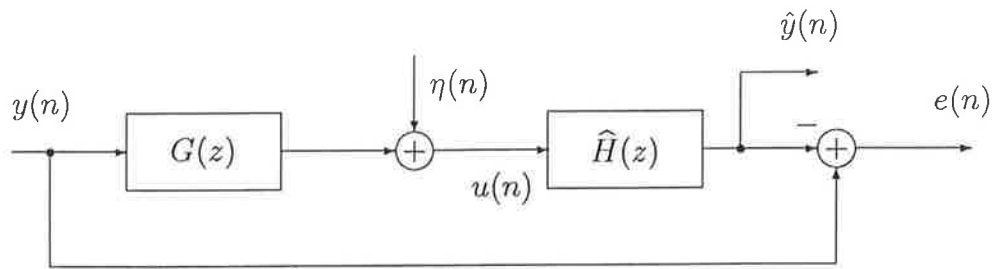


Figura 1-4: equalização adaptativa em transmissão digital

$E[e^2(n)]$ , com  $e(n) = y(n) - \hat{y}(n) = y(n) - \hat{H}(z)u(n)$ .

A Figura 1-4 representa a fase de treinamento do equalizador, na qual a sequência  $y(n)$  é conhecida. Nesta situação, um filtro FIR de duas entradas poderia ser usado como alternativa ao filtro IIR. Durante a fase de operação, porém, no lugar do sinal  $y(n)$  tem que usar-se uma estimativa, recaindo-se de qualquer modo em uma estrutura recursiva: se a estimativa usada for  $\hat{y}(n)$  tem-se exatamente o caso IIR; se a estimativa usada for o símbolo detectado a partir de  $\hat{y}(n)$  tem-se a estrutura chamada *Decision Feedback Equalizer* [21].

#### 1.4.4 Cancelamento de ruído sonoro

Outra aplicação ainda de filtros IIR adaptativos é no cancelamento de ruído sonoro [24], representado na Figura 1-5. Neste caso o objetivo é cancelar o efeito da fonte de ruído no sensor primário, utilizando para tanto uma fonte de ruído secundária gerada a partir da filtragem  $\hat{H}(z)$  do ruído obtido de um sensor de referência. Sobre o sensor primário incidem ainda outros sinais sonoros que não devem prejudicar a adaptação de  $\hat{H}(z)$ .

Em diagrama de blocos, o problema de cancelamento de ruído sonoro está representado na Figura 1-6, onde  $r(n)$  é o sinal produzido pela fonte de ruído,  $G_1(z)$  é a função de transferência entre a fonte de ruído e o sensor primário,  $G_2(z)$  é a função de transferência entre a fonte de ruído e o sensor de referência,  $G_3(z)$  é a função de transferência entre a fonte de ruído secundária e o sensor primário e  $\eta(n)$  é a soma dos demais sons do ambiente no sensor primário. Foi considerado que há um alto desacoplamento entre a fonte de ruído secundária e o sensor de referência.

Conceitualmente, podemos separar o problema em dois outros, já vistos acima: 1) identificação de  $G_3(z)$  (Figura 1-7) e 2) cancelamento de  $[G_1(z)/G_2(z)]u(n)$ , supondo  $G_2(z)$  de

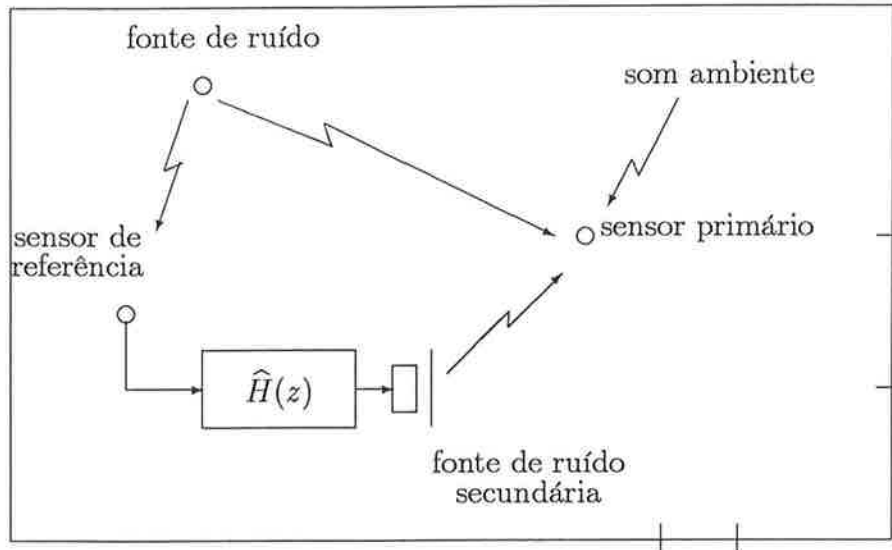


Figura 1-5: cancelamento de ruído sonoro

fase mínima (Figura 1-8). Na prática, a solução conjunta destes dois problemas dependerá das particularidades da aplicação, por exemplo, da mobilidade da fonte de ruído e do sensor primário, da viabilidade de usar um sinal internamente gerado para realizar a identificação de  $G_3(z)$ , etc.

## 1.5 Algumas definições e propriedades

Neste item são apresentadas definições e propriedades que serão utilizadas no restante do trabalho.

### 1.5.1 Espaços e propriedades de funções em $z$

Consideremos inicialmente o espaço  $L_2$  das funções  $f(z)$  que tem norma  $L_2$  finita:

$$\|f(z)\| \triangleq \left[ \int_{-\pi}^{\pi} |f(e^{j\omega})|^2 d\omega / 2\pi \right]^{1/2} < \infty.$$

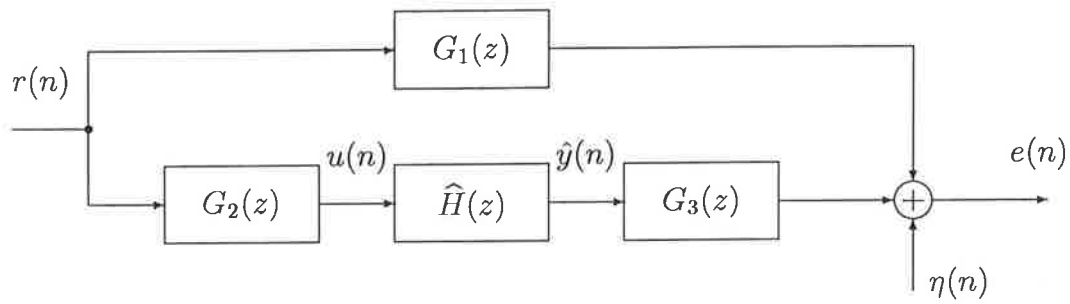


Figura 1-6: diagrama de blocos do cancelamento de ruído sonoro

$$\zeta_1(n) = \eta(n) + G_1(z)r(n)$$

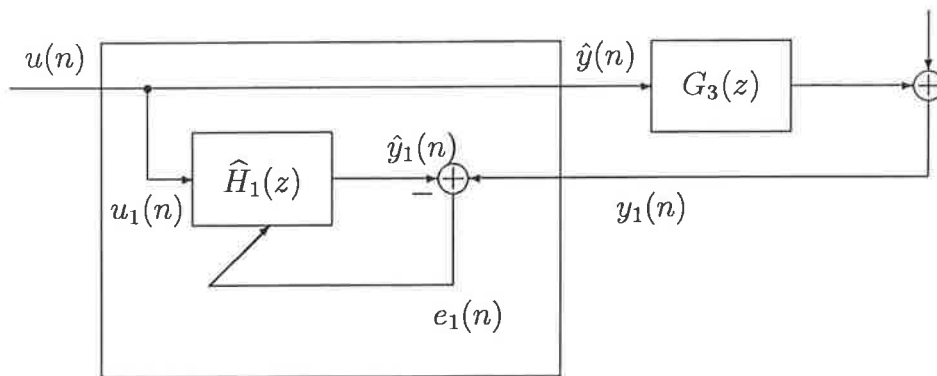


Figura 1-7: identificação da função de transferência entre a fonte secundária e o sensor primário

Uma função que pertence a  $L_2$  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 se escrever simplesmente

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

Notar que, de modo geral, a região de convergência de expansões como a acima não se limita ao círculo unitário  $|z| = 1$ . Na maioria dos casos, porém, não será necessário considerar explicitamente as regiões de convergência das expansões utilizadas, sendo suficiente que a

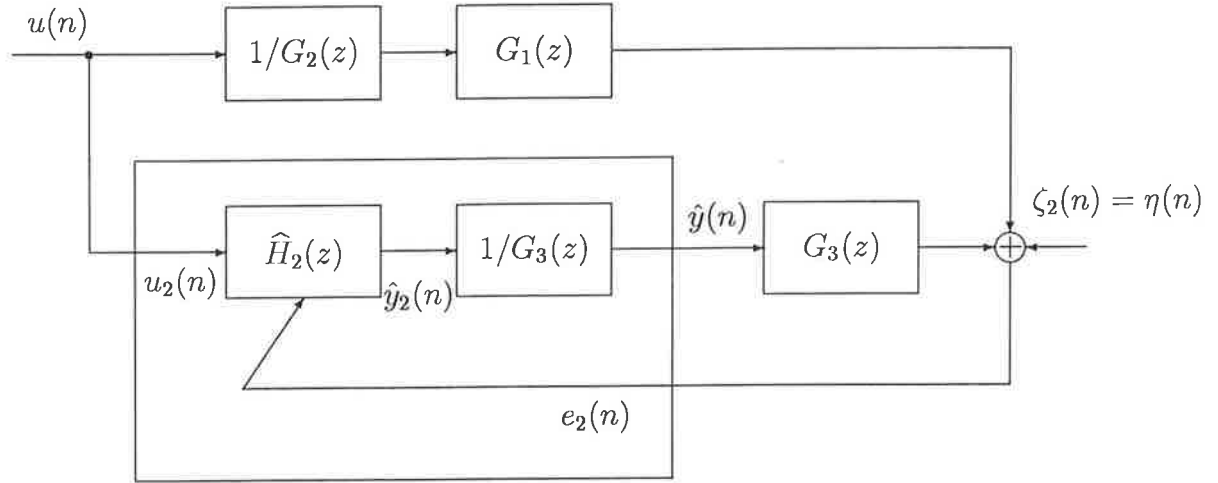


Figura 1-8: cancelamento de ruído após identificação de  $G_3(z)$

circunferência unitária faça parte destas regiões.

Será conveniente em certos casos separar uma função qualquer  $f(z) = \sum_{k=-\infty}^{\infty} f_k z^k$  em suas projeções estritamente causal  $[f(z)]_+ \triangleq \sum_{k=1}^{\infty} f_k z^k$  e anti-causal  $[f(z)]_- \triangleq \sum_{k=-\infty}^0 f_k z^k$ :

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

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

Um sub-espaço das funções  $L_2$  em  $z$  particularmente importante é o sub-espaço de Hardy  $\mathcal{H}_2$ , definido como o conjunto das funções  $f(z)$  de norma  $L_2$  finita e que admitem uma expansão em somatória causal:

$$\mathcal{H}_2 = \{f(z) \mid f(z) \in L_2, f(z) = \sum_{k=0}^{\infty} f_k z^k, \quad |z| = 1\} \quad (1.14)$$

(em outros termos,  $f(z)$  é estável e causal). Se  $f(z)$  for uma função racional

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

então  $f(z) \in \mathcal{H}_2$  se os polos de  $A(z)$  estiverem fora do círculo unitário. Notar que  $B(z^{-1})/A(z^{-1}) = f(z^{-1})$  tem os polos dentro do círculo unitário e sua expansão é dada

por

$$\frac{B(z^{-1})}{A(z^{-1})} = \sum_{k=0}^{\infty} f_k z^{-k} = \sum_{k=-\infty}^0 f_{-k} z^k \quad (1.15)$$

ou seja, se  $B(z)/A(z) \in \mathcal{H}_2$  então  $B(z^{-1})/A(z^{-1})$  é uma função estável (tem norma  $L_2$  finita) e anti-causal. Analogamente se  $B(z)/A(z) \in \mathcal{H}_2$  e for estritamente causal ( $b_0 = 0$ ) então  $B(z^{-1})/A(z^{-1})$  é uma função estável e estritamente anti-causal.

Dadas duas funções  $f(z)$  e  $g(z)$  de  $L_2$  define-se o seu produto interno como

$$\langle f(z), g(z) \rangle \triangleq \frac{1}{2\pi j} \oint_{|z|=1} f(z)g(z^{-1}) \frac{dz}{z} = \sum_{k=-\infty}^{\infty} f_k g_k \quad (1.16)$$

(subentende-se que a integral de linha é feita no sentido anti-horário). Pode-se verificar diretamente que para quaisquer  $f(z)$ ,  $h(z)$  e  $g(z)$  de  $L_2$  o produto interno possui as seguintes propriedades:

$$\langle f(z)h(z), g(z) \rangle = \langle f(z), h(z^{-1})g(z) \rangle \quad \text{e} \quad \langle f(z), g(z) \rangle = \langle g(z), f(z) \rangle, \quad (1.17)$$

e que a norma  $L_2$  também pode ser expressa como

$$\|f(z)\| = \langle f(z), f(z) \rangle^{1/2} = \left( \sum_{k=-\infty}^{\infty} f_k^2 \right)^{1/2}. \quad (1.18)$$

O produto interno pode ser usado para exprimir a correlação entre dois sinais  $v(n) = F(z)u(n)$  e  $w(n) = G(z)u(n)$  obtidos pela passagem de um mesmo sinal de entrada  $u(n)$  por dois sistemas diferentes:

$$\begin{aligned} E[v(n)w(n)] &= \langle F(z), G(z)\mathcal{S}_u(z) \rangle \\ &= \langle F(z)\mathcal{S}_u(z), G(z) \rangle, \end{aligned} \quad (1.19)$$

onde  $\mathcal{S}_u(z) = \mathcal{S}_u(z^{-1})$  é a transformada  $z$  da autocorrelação da entrada  $u(n)$ :

$$\mathcal{S}_u(z) = \sum_{k=-\infty}^{\infty} E[u(n)u(n+k)]z^k. \quad (1.20)$$

Uma outra operação de interesse é a derivada de um produto interno em relação a um determinado parâmetro  $w$ , que pode ser escrita como:

$$\begin{aligned}
\frac{\partial}{\partial w} \langle f(z), g(z) \rangle &= \frac{\partial}{\partial w} \sum_{k=-\infty}^{\infty} f_k g_k = \sum_{k=-\infty}^{\infty} \left( \frac{\partial}{\partial w} f_k \right) g_k + f_k \left( \frac{\partial}{\partial w} g_k \right) \\
&= \left\langle \frac{\partial}{\partial w} f(z), g(z) \right\rangle + \left\langle f(z), \frac{\partial}{\partial w} g(z) \right\rangle.
\end{aligned} \tag{1.21}$$

Por fim, a seguinte definição será utilizada no restante deste trabalho:

**Definição 1.1** O grau  $\deg[H(z)]$  de uma função  $H(z)$  qualquer é o menor número de retardos necessários para implementar  $H(z)$ .

### 1.5.2 Propriedades de autovalores e valores singulares

Indiquemos os autovalores de uma matriz  $\mathbf{H}$  por  $\{\lambda_i(\mathbf{H})\}$ . Em algumas situações será necessário caracterizar os autovalores máximo e mínimo de uma matriz simétrica positiva semi-definida. A seguinte propriedade será então útil:

**Propriedade 1.1** O maior e o menor autovalores de uma matriz  $\mathbf{H}$  simétrica positiva semi-definida são dados por

$$\max\{\lambda_i(\mathbf{H})\} = \max_{\mathbf{x}} \frac{\mathbf{x}^\top \mathbf{H} \mathbf{x}}{\mathbf{x}^\top \mathbf{x}} \geq 0 \tag{1.22}$$

e

$$\min\{\lambda_i(\mathbf{H})\} = \min_{\mathbf{x}} \frac{\mathbf{x}^\top \mathbf{H} \mathbf{x}}{\mathbf{x}^\top \mathbf{x}} \geq 0. \tag{1.23}$$

*Prova:* Uma vez que  $\mathbf{H}$  é simétrica,  $\mathbf{H}$  é também uma matriz normal, isto é,  $\mathbf{H}\mathbf{H}^\top = \mathbf{H}^\top\mathbf{H}$ . Sendo uma matriz normal, segue que sempre pode ser escrita como

$$\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{U}^\top, \tag{1.24}$$

onde  $\mathbf{D}$  é a matriz diagonal dos autovalores de  $\mathbf{H}$  e  $\mathbf{U}$  é ortogonal, isto é  $\mathbf{U}^\top\mathbf{U} = \mathbf{I}$  e as colunas de  $\mathbf{U}$  formam uma base ortonormal ([31], p. 175). Notar que, devido à normalidade de  $\mathbf{H}$ , isto é válido mesmo que existam autovalores repetidos.

Qualquer vetor  $\mathbf{x}$  pode agora ser escrito em função das colunas de  $\mathbf{U}$  como  $\mathbf{x} = x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + \dots + x_N\mathbf{u}_N$ , e, portanto, usando ainda a hipótese de que  $\mathbf{H}$  é positiva semi-definida,

podemos escrever, usando (1.24)

$$\frac{\mathbf{x}^\top \mathbf{H} \mathbf{x}}{\mathbf{x}^\top \mathbf{x}} = \frac{x_1^2 \lambda_1 + x_2^2 \lambda_2 + \dots + x_N^2 \lambda_N}{x_1^2 + x_2^2 + \dots + x_N^2} \geq 0, \quad (1.25)$$

onde  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$  são os autovalores de  $\mathbf{H}$ . Temos portanto que

$$0 \leq \frac{\mathbf{x}^\top \mathbf{H} \mathbf{x}}{\mathbf{x}^\top \mathbf{x}} \leq \frac{x_1^2 + x_2^2 + \dots + x_N^2}{x_1^2 + x_2^2 + \dots + x_N^2} \lambda_1 = \lambda_1 \quad (1.26)$$

e

$$\frac{\mathbf{x}^\top \mathbf{H} \mathbf{x}}{\mathbf{x}^\top \mathbf{x}} \geq \frac{x_1^2 + x_2^2 + \dots + x_N^2}{x_1^2 + x_2^2 + \dots + x_N^2} \lambda_N = \lambda_N \geq 0. \quad (1.27)$$

Como os valores máximo e mínimo são assumidos, respectivamente, para  $\mathbf{x} = x_1 \mathbf{u}_1$  e  $\mathbf{x} = x_N \mathbf{u}_N$ , a propriedade fica demonstrada. ■

Ao lidar com matrizes não necessariamente simétricas ou positivas semi-definidas, será mais conveniente às vezes obter propriedades relacionadas com seus valores singulares, ao invés de relacionadas com seus autovalores (os quadrados dos valores singulares de  $\mathbf{H}$  são os autovalores de  $\mathbf{H}^\top \mathbf{H}$ ). A propriedade apresentada a seguir estabelece uma relação entre estas grandezas.

**Propriedade 1.2** *Sejam  $\{\lambda_i(\mathbf{H})\}$  os autovalores (possivelmente complexos) de  $\mathbf{H}$  e sejam  $\{\sigma_i(\mathbf{H})\}$  seus valores singulares. As seguintes relações são válidas:*

$$\min\{|\lambda_i(\mathbf{H})|\} \geq \min\{\sigma_i(\mathbf{H})\}, \quad (1.28)$$

$$\max\{|\lambda_i(\mathbf{H})|\} \leq \max\{\sigma_i(\mathbf{H})\}. \quad (1.29)$$

*Prova:* Se  $\lambda_i$  é um autovalor de  $\mathbf{H}$  e  $\mathbf{u}_i$  o autovetor, de norma unitária, associado a  $\mathbf{u}_i$ , então  $\mathbf{H} \mathbf{u}_i = \lambda_i \mathbf{u}_i$  e, portanto,  $\mathbf{u}_i^\top \mathbf{H}^\top \mathbf{H} \mathbf{u}_i = |\lambda_i|^2$ . Separando  $\mathbf{u}_i$  em suas partes real e imaginária, e usando a propriedade 1.1 temos

$$|\lambda_i|^2 = \mathbf{u}_{i,R}^\top \mathbf{H}^\top \mathbf{H} \mathbf{u}_{i,R} + \mathbf{u}_{i,I}^\top \mathbf{H}^\top \mathbf{H} \mathbf{u}_{i,I} \quad (1.30)$$

$$\geq \min \mathbf{u}_{i,R}^\top \mathbf{H}^\top \mathbf{H} \mathbf{u}_{i,R} + \min \mathbf{u}_{i,I}^\top \mathbf{H}^\top \mathbf{H} \mathbf{u}_{i,I} \quad (1.31)$$

$$= \min\{\sigma_i^2(\mathbf{H})\} [\mathbf{u}_{i,R}^\top \mathbf{u}_{i,R} + \mathbf{u}_{i,I}^\top \mathbf{u}_{i,I}] \quad (1.32)$$

$$= \min\{\sigma_i^2(\mathbf{H})\}, \quad (1.33)$$



e, portanto  $\min\{|\lambda_i(\mathbf{H})|\} \geq \min\{\sigma_i(\mathbf{H})\}$ , e, procedendo analogamente,  $\max\{|\lambda_i(\mathbf{H})|\} \leq \max\{\sigma_i(\mathbf{H})\}$ . ■

### 1.5.3 Excitação persistente

Usaremos no trabalho a seguinte definição para uma excitação persistente:

**Definição 1.2** *Um sinal  $u(n)$  é denominado de uma excitação persistente de ordem  $N$  se sua densidade espectral de potência*

$$S_u(e^{j\omega}) = \sum_{k=-\infty}^{\infty} E[u(n)u(n-k)](e^{j\omega k}) \quad (1.34)$$

for não nula para ao menos  $N$  valores de  $\omega$ .

Para maiores detalhes sobre o conceito de excitação persistente ver, por exemplo, [17] ou [1].

### 1.5.4 Ortogonalidade a um vetor de regressão

Consideremos os polinômios  $C(z)$  e  $D(z)$ , com  $D(z)$  de fase mínima e

$$\max\{\deg[C(z)], \deg[D(z)]\} = M \quad (1.35)$$

e o sinal gerado como

$$y(n) = \frac{C(z)}{D(z)}u(n), \quad (1.36)$$

onde  $u(n)$  é uma excitação persistente de ordem  $2M + 1$  ou maior. Definimos aqui um vetor de regressão  $\mathbf{r}(n)$  associado a  $C(z)$  e  $D(z)$  como:

$$\mathbf{r}(n) = \begin{bmatrix} -y(n-1) \\ -y(n-2) \\ \vdots \\ -y(n-M) \\ u(n) \\ u(n-1) \\ \vdots \\ u(n-M) \end{bmatrix}. \quad (1.37)$$

Repartindo os elementos de um vetor  $\mathbf{e}$  como

$$\mathbf{e} = \left[ r_0 \ r_1 \ \cdots \ r_{M-1} \ q_0 \ q_1 \ \cdots \ q_M \right]^T, \quad (1.38)$$

o produto interno  $\mathbf{e}^T \mathbf{r}(n)$  pode ser escrito como

$$\mathbf{e}^T \mathbf{r}(n) = [Q(z) - zR(z) \frac{C(z)}{D(z)}] u(n) \quad (1.39)$$

onde  $R(z) = r_0 + r_1 z + \dots + r_{M-1} z^{M-1}$  e  $Q(z) = q_0 + q_1 z + \dots + q_M z^M$ . A seguinte propriedade é válida:

**Propriedade 1.3** Com  $\mathbf{r}(n)$  dado por (1.37) e (1.36), existe um vetor  $\mathbf{e}$  tal que  $\mathbf{e}^T \mathbf{r}(n) \equiv 0$  se, e somente se,  $C(z)$  e  $D(z)$  tem alguma raiz comum.

*Prova:* Em (1.39),  $Q(z) - zR(z)C(z)/D(z)$  tem, no máximo,  $2M$  zeros. Uma vez que  $u(n)$  é uma excitação persistente de ordem  $2M + 1$  ou maior, resulta que  $\mathbf{e}^T \mathbf{r}(n) \equiv 0$  se, e somente se,  $Q(z) - zR(z)C(z)/D(z) \equiv 0$ , ou

$$\frac{Q(z)}{zR(z)} = \frac{C(z)}{D(z)}. \quad (1.40)$$

Por hipótese,  $D(z)$  é de fase mínima e portanto não pode ter  $z = 0$  como raiz. Para valer a igualdade acima o termo  $z$  do denominador tem que ser cancelado então por  $Q(z) = zQ'(z)$

resultando em

$$\frac{Q'(z)}{R(z)} = \frac{C(z)}{D(z)}. \quad (1.41)$$

Agora, uma vez que  $\deg[Q'(z)] \leq M - 1$ ,  $\deg[R(z)] \leq M - 1$  e por hipótese  $\max\{\deg[C(z)], \deg[D(z)]\} = M$ , resulta que esta igualdade é válida se, e somente se,  $C(z)$  e  $D(z)$  tiverem alguma raiz comum, o que conclui a prova. ■

## Capítulo 2

# Características de algoritmos de adaptação para filtros IIR na forma direta

Neste capítulo são descritos algoritmos de adaptação comumente usados para filtros IIR adaptativos implementados na forma direta e seu comportamento é analisado.

No Item 2.1 é definida a notação referente à variação no tempo de um filtro adaptativo. No Item 2.2, inicialmente, descreve-se a forma genérica de algoritmos de passo constante e o método ODE (*ordinary differential equation*) de análise de convergência, no qual, para o caso de modelamento suficiente, a matriz de estados de um sistema de equações diferenciais assume um papel importante. Considerando o caso genérico em que esta matriz não é simétrica são obtidas então relações entre os seus valores singulares extremos, o passo de adaptação máximo e a velocidade de convergência local do algoritmo. Mostra-se também que a simetria desta matriz favorece a validade da aproximação da convergência local do algoritmo de adaptação pelo decaimento dos modos normais de um sistema linear livre associado ao mesmo.

Nos Itens 2.3 a 2.5 a forma genérica dos algoritmos de passo constante é particularizada para os seguintes algoritmos, comumente usados para filtragem IIR adaptativa: algoritmo do gradiente recursivo (RG), algoritmo baseado no método de identificação Steiglitz-McBride (SMM), regressão pseudo-linear (PLR) e o algoritmo SHARF. Em relação ao algoritmo SMM apresenta-se uma interpretação simples da passagem de sua forma *off-line* para a forma recursiva. A análise descrita no Item 2.2 também é particularizada para cada algoritmo,

o que conduz a condições específicas para a convergência ao valor correto dos parâmetros. Para o algoritmo RG mostra-se como a proximidade à circunferência unitária dos polos do sistema sendo modelado reduz a parcela que pode ser garantida do domínio de atração deste ponto.

No Item 2.6 apresenta-se uma análise do efeito da proximidade ao círculo unitário dos polos do sistema sendo modelado no condicionamento da matriz de estados da ODE. A motivação desta análise é dada, por um lado, pelo interesse de se usar filtros IIR adaptativos em situações onde o sistema sendo modelado tem polos próximos ao círculo unitário, discutido no Capítulo 1; e, por outro lado, pela relação mostrada no Item 2.2 entre o condicionamento da matriz de estados da ODE e a velocidade de convergência local do algoritmo de adaptação. O método de análise apresentado é baseado em uma transposição ao domínio da frequência de certas formas quadráticas cujos limites superior e inferior correspondem aos valores singulares extremos da matriz em questão. Esta transposição ao domínio da frequência é bastante conveniente para a verificação do efeito da posição dos polos do sistema sendo modelado. É conveniente também para a introdução de certas restrições ao mesmo, como por exemplo, norma  $L_2$  fixa e resposta em frequência passa-tudo. Alguns resultados da análise apresentada no Item 2.6 podem ser destacados:

- A única situação em que o espalhamento dos valores singulares da matriz de estados da ODE não tende a infinito quando um número qualquer de polos do sistema sendo modelado  $H(z)$  se aproxima da circunferência unitária, é quando  $H(z)$  é passa-tudo e emprega-se o algoritmo PLR.
- Para o algoritmo SHARF, o aparente ideal em termos do atendimento de uma condição de positividade necessária para convergência, não é, necessariamente, o ideal em termos da velocidade de convergência local do algoritmo.
- As relações obtidas entre as velocidades de crescimento ou decaimento dos valores singulares extremos da matriz de estados da ODE para os diferentes algoritmos permitem estabelecer relações entre os valores limites para estas mesmas grandezas, mesmo sem especificar completamente  $H(z)$ .

Por fim, no Item 2.7 são apresentadas várias simulações da execução dos algoritmos de adaptação em questão, onde são exemplificados os aspectos analisados do ponto de visto teórico nos itens precedentes.

## 2.1 Notação para filtros com parâmetros variantes no tempo

Para introduzirmos a dependência em relação ao tempo na expressão da saída do filtro adaptativo escrevemos

$$\hat{y}(n) = \hat{H}(z, n)u(n), \quad (2.1)$$

que deve ser entendido como representando uma equação de diferenças com coeficientes variantes no tempo:

$$\hat{y}(n) = \sum_{i=0}^M b_i(n)u(n-i) - \sum_{i=1}^M a_i(n)\hat{y}(n-i), \quad n = 0, 1, \dots \quad (2.2)$$

A expressão (2.1) corresponde então à operação real de um filtro IIR adaptativo, com a saída a cada instante  $n$  sendo calculada com coeficientes diferentes dos usados nos instantes anteriores.

É conveniente introduzir aqui uma diferenciação na terminologia: o termo "coeficiente" será aplicado aos elementos dos conjuntos  $\{a_i\}_1^M$  e  $\{b_i\}_0^M$  que definem a forma racional de  $H(z) = B(z)/A(z)$ , independentemente da estrutura de implementação. O termo "parâmetro" será aplicado às grandezas que são efetivamente adaptadas na estrutura de implementação de  $H(z)$ , e usaremos o vetor  $\mathbf{w}$  para designar o conjunto de parâmetros adaptados. Em uma estrutura de implementação direta, os parâmetros adaptados são os próprios coeficientes de  $H(z)$ .

Passando a usar então  $\hat{y}(n)$  para indicar a saída do filtro adaptativo cujos parâmetros estão variando no tempo, introduzimos a notação  $\hat{y}(n, \mathbf{w})$  para indicar o caso em que estes parâmetros estão fixos em  $\mathbf{w}$ :

$$\hat{y}(n, \mathbf{w}) = \hat{H}(z)u(n) \quad (2.3)$$

A mesma notação vale para o erro de saída: com a saída do sistema sendo modelado dada por

$$y(n) = H(z)u(n) + \zeta(n), \quad (2.4)$$

o erro de saída, no caso de parâmetros variantes no tempo, é dado por

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

$$= [H(z) - \hat{H}(z, n)]u(n) + \zeta(n), \quad (2.5)$$

e no caso de parâmetros fixados em  $\mathbf{w}$ , por

$$e(n, \mathbf{w}) = [H(z) - \hat{H}(z)]u(n) + \zeta(n). \quad (2.6)$$

## 2.2 Convergência local de algoritmos de passo constante

### 2.2.1 Formulação genérica de algoritmos de adaptação e equação diferencial associada

Vários algoritmos utilizados para filtros IIR adaptativos podem ser enquadrados na seguinte formulação genérica:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \boldsymbol{\psi}(n) \epsilon(n), \quad (2.7)$$

onde  $\mathbf{w}(n)$  é o vetor de parâmetros,  $\epsilon(n) = P(z)e(n)$  é o erro de saída filtrado,  $P(z)$  e o vetor  $\boldsymbol{\psi}(n)$  dependem do algoritmo específico e  $\mu$  é o passo de adaptação.

Se  $u(n)$ ,  $y(n)$ ,  $\boldsymbol{\psi}(n)$  e  $\epsilon(n)$  obedecerem as seguintes condições pouco restritivas ([1]):

1. as sequências  $u(n)$  e  $y(n)$  são conjuntamente estacionárias,
2. as sequências  $\boldsymbol{\psi}(n)$  e  $\epsilon(n)$  são geradas por filtros que são exponencialmente estáveis para  $\mathbf{w}(n) \in W_s$ ,

pode-se demonstrar ([13], e também [14] para a aplicação em filtros IIR adaptativos) que, para valores do passo de adaptação  $\mu$  suficientemente pequenos, o processo de adaptação descrito por (2.7) pode ser aproximado pela solução da seguinte equação diferencial ordinária (ODE):

$$\frac{d\mathbf{w}(t)}{dt} = E[\boldsymbol{\psi}(n, \mathbf{w}(t)) \epsilon(n, \mathbf{w}(t))], \quad \mathbf{w}(t) \in W_s, \quad (2.8)$$

com  $t = n\mu$ . O sub-conjunto do espaço de parâmetros  $W_s$  é tomado como o domínio de estabilidade dos filtros fixos que gerariam  $\boldsymbol{\psi}(n, \mathbf{w})$  e  $\epsilon(n, \mathbf{w})$ .

Na expressão acima  $\mathbf{w}(t)$  é um sinal determinístico, razão pela qual foi adotado um tipo diferente (sans-serif) do tipo de  $\mathbf{w}(n)$  (roman negrito). Por outro lado, símbolos em negrito sem o argumento de tempo, como  $\mathbf{w}$ , denotam sempre um vetor constante, não aleatório.

A notação  $\psi(n, \mathbf{w}(t))$  e  $\epsilon(n, \mathbf{w}(t))$  indica, como convencionado no Item 2.1, que  $\psi(n)$  e  $\epsilon(n)$  são calculados com o valor dos parâmetros fixado em  $\mathbf{w} = \mathbf{w}(t)$ . A operação de esperança, portanto, é efetuada apenas sobre o conjunto de realizações de  $u(n)$  e  $\epsilon(n)$ .

Em relação à natureza da aproximação de (2.7) realizada por (2.8), tem-se que para  $\mu$  suficientemente pequeno existem duas constantes positivas e finitas  $C$  e  $C'$  e uma função positiva  $\delta(\mu)$  tendendo a zero com  $\mu$  tais que ([13], [14] )

$$\Pr \left\{ \sup_n \|\mathbf{w}(n) - \mathbf{w}(n\mu)\| > C\delta(\mu) \right\} < C'\delta(\mu), \quad (2.9)$$

onde, genericamente,  $\Pr\{A\}$  indica a probabilidade do evento  $A$ . Segue que quando  $\mu$  tende a zero a média de  $\mathbf{w}(n)$  tende a  $\mathbf{w}(n\mu)$  e a variância de  $\mathbf{w}(n)$  tende a zero. Além disso, quando (2.8) tende a algum limite  $\mathbf{w}^*$ , (2.9) garante que para qualquer  $\varepsilon > 0$  pode-se achar um passo de adaptação  $\mu$  suficientemente pequeno tal que

$$E [\|\mathbf{w}(n) - \mathbf{w}^*\|] < \varepsilon, \quad n > n_c(\mu), \quad (2.10)$$

para um certo  $n_c(\mu)$  que cresce conforme  $\mu$  diminui.

Por fim, é importante notar que na formulação genérica feita acima, não há menção da existência, no algoritmo de adaptação, de um mecanismo para evitar que os parâmetros  $\mathbf{w}(n)$  saiam do domínio de estabilidade. Como será visto a seguir, a validade da aproximação do algoritmo pela equação diferencial (2.8) permite estabelecer condições suficientes para a convergência do mesmo, e, portanto, tal mecanismo não seria, a rigor, necessário. Porém, quando viável (para filtros de ordem superior a 2, por exemplo, a complexidade computacional de um controle de estabilidade pode ser proibitiva), a sua utilização é vantajosa pois permite a utilização de um passo de adaptação  $\mu$  maior. Na prática, portanto, a questão do controle de estabilidade é um dos aspectos importantes de filtros IIR adaptativos. Neste trabalho porém, a sua análise não será mais aprofundada.

### 2.2.2 Condição suficiente para convergência local

Consideremos que existe um ponto  $\mathbf{w} = \mathbf{w}^*$  do espaço de parâmetros onde  $\hat{H}(z) = H(z)$ , e, portanto, onde o erro de saída é dado apenas pelo erro de medida:

$$e(n, \mathbf{w}^*) = \zeta(n). \quad (2.11)$$



Em outros termos, considerando a mesma parametrização que a de  $\widehat{H}(z)$ ,  $\mathbf{w}^*$  é o vetor de parâmetros de  $H(z)$ . Em razão disto, o ponto  $\mathbf{w}^*$  do espaço de parâmetros será também referido, como "o valor correto dos parâmetros". As características locais da convergência de algoritmos expressos genericamente por (2.7), em torno deste ponto  $\mathbf{w}^*$ , são analisadas a seguir.

Do que foi visto no item anterior resulta que, para valores suficientemente pequenos do passo de adaptação  $\mu$ , a convergência do algoritmo pode ser caracterizada através da equação diferencial associada ao mesmo. Passamos a analisar, então, esta equação diferencial. Usando (2.11) temos, de início, que  $E[\boldsymbol{\psi}(n, \mathbf{w}^*) \boldsymbol{\epsilon}(n, \mathbf{w}^*)] = E[\boldsymbol{\psi}(n, \mathbf{w}^*) P(z)\zeta(n)]$ . Portanto se

$$E[\boldsymbol{\psi}(n, \mathbf{w}^*) P(z)\zeta(n)] = \mathbf{0} \quad (2.12)$$

então o ponto  $\mathbf{w}^*$  é um ponto estacionário da equação diferencial (2.8). Para obter-se uma condição para que  $\mathbf{w}^*$  seja, adicionalmente, um ponto estacionário *estável* da equação diferencial ordinária, escrevemos o erro de saída  $e(n, \mathbf{w})$  como função do erro em relação a  $\mathbf{w}^*$  dos parâmetros de uma implementação direta de  $\widehat{H}(z)$ : inicialmente, com

$$\widehat{H}(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1z + \dots + b_Mz^M}{1 + a_1z + \dots + a_Mz^M} \quad (2.13)$$

e

$$H(z) = \frac{C(z)}{D(z)} = \frac{c_0 + c_1z + \dots + c_Mz^M}{1 + d_1z + \dots + d_Mz^M}, \quad (2.14)$$

onde  $\deg[H(z)] = M$ , temos

$$y'(n) = y(n) - \zeta(n) = \frac{C(z)}{D(z)}u(n) \quad (2.15)$$

e portanto

$$y'(n)D(z) = C(z)u(n). \quad (2.16)$$

Resulta das expressões acima a seguinte identidade [14] que nos será útil:

$$\begin{aligned} [C(z) - B(z)] \frac{u(n)}{A(z)} - [D(z) - A(z)] \frac{y'(n)}{A(z)} &= \frac{C(z)}{A(z)}u(n) - \widehat{y}(n) - \frac{D(z)}{A(z)}y'(n) + y'(n) \\ &= y'(n) - \widehat{y}(n) \\ &= e(n, \mathbf{w}) - \zeta(n). \end{aligned} \quad (2.17)$$

Portanto, o erro de saída filtrado pode ser escrito em forma matricial como

$$\epsilon(n, \mathbf{w}) = P(z)e(n, \mathbf{w}) = \boldsymbol{\phi}_1^\top(n, \mathbf{w})\tilde{\mathbf{w}} + P(z)\zeta(n), \quad (2.18)$$

onde

$$\boldsymbol{\phi}_1(n, \mathbf{w}) = \begin{bmatrix} -y'(n-1)P(z)/A(z) \\ -y'(n-2)P(z)/A(z) \\ \vdots \\ -y'(n-M)P(z)/A(z) \\ u(n)P(z)/A(z) \\ u(n-1)P(z)/A(z) \\ \vdots \\ u(n-M)P(z)/A(z) \end{bmatrix} \quad (2.19)$$

e

$$\tilde{\mathbf{w}} = \begin{bmatrix} d_1 - a_1 \\ d_2 - a_2 \\ \vdots \\ d_M - a_M \\ c_0 - b_0 \\ c_1 - b_1 \\ \vdots \\ c_M - b_M \end{bmatrix}. \quad (2.20)$$

Da mesma maneira, uma vez que

$$\hat{y}(n)A(z) = B(z)u(n), \quad (2.21)$$

podemos escrever [14]

$$\begin{aligned} [C(z) - B(z)]\frac{u(n)}{D(z)} - [D(z) - A(z)]\frac{\hat{y}(n)}{D(z)} &= -\frac{B(z)}{D(z)}u(n) + y'(n) + \frac{A(z)}{D(z)}\hat{y}(n) - \hat{y}(n) \\ &= y'(n) - \hat{y}(n) \\ &= e(n, \mathbf{w}) - \zeta(n), \end{aligned} \quad (2.22)$$

e, portanto, o erro de saída filtrado também pode ser escrito como

$$\epsilon(n, \mathbf{w}) = P(z)e(n, \mathbf{w}) = \boldsymbol{\phi}_2^\top(n, \mathbf{w})\tilde{\mathbf{w}} + P(z)\zeta(n), \quad (2.23)$$

onde

$$\boldsymbol{\phi}_2(n, \mathbf{w}) = \begin{bmatrix} -\hat{y}(n-1)P(z)/D(z) \\ -\hat{y}(n-2)P(z)/D(z) \\ \vdots \\ -\hat{y}(n-M)P(z)/D(z) \\ u(n)P(z)/D(z) \\ u(n-1)P(z)/D(z) \\ \vdots \\ u(n-M)P(z)/D(z) \end{bmatrix}. \quad (2.24)$$

Definindo agora  $\tilde{\mathbf{w}}(t) = \mathbf{w}^* - \mathbf{w}(t)$  e notando que  $d\tilde{\mathbf{w}}(t)/dt = -d\mathbf{w}(t)/dt$ , (2.18) pode ser introduzido em (2.8) para obter

$$\begin{aligned} \frac{d\tilde{\mathbf{w}}(t)}{dt} &= -E [\boldsymbol{\psi}(n, \mathbf{w}(t))\boldsymbol{\phi}_1^\top(n, \mathbf{w}(t))\tilde{\mathbf{w}}(t)] - E [\boldsymbol{\psi}(n, \mathbf{w}(t))P(z)\zeta(n)] \\ &= -E [\boldsymbol{\psi}(n, \mathbf{w}(t))\boldsymbol{\phi}_1^\top(n, \mathbf{w}(t))] \tilde{\mathbf{w}}(t) - E [\boldsymbol{\psi}(n, \mathbf{w}(t))P(z)\zeta(n)] \\ &= -\mathbf{R}(\mathbf{w}(t))\tilde{\mathbf{w}}(t) - E [\boldsymbol{\psi}(n, \mathbf{w}(t))P(z)\zeta(n)], \end{aligned} \quad (2.25)$$

onde

$$\mathbf{R}(\mathbf{w}(t)) = E [\boldsymbol{\psi}(n, \mathbf{w}(t))\boldsymbol{\phi}_1^\top(n, \mathbf{w}(t))]. \quad (2.26)$$

Uma vez que estamos assumindo que (2.12) é válida, esta equação diferencial é ainda simplificada resultando em

$$\frac{d\tilde{\mathbf{w}}(t)}{dt} = -\mathbf{R}(\mathbf{w}(t))\tilde{\mathbf{w}}(t). \quad (2.27)$$

A solução de (2.27), devido a (2.18) e (2.23), é igual à solução de

$$\frac{d\tilde{\mathbf{w}}(t)}{dt} = -\mathbf{R}'(\mathbf{w}(t))\tilde{\mathbf{w}}(t), \quad (2.28)$$

onde

$$\mathbf{R}'(\mathbf{w}(t)) = E [\boldsymbol{\psi}(n, \mathbf{w}(t))\boldsymbol{\phi}_2^\top(n, \mathbf{w}(t))]. \quad (2.29)$$

Denominamos então a matriz  $\mathbf{R}$  (ou a matriz  $\mathbf{R}'$ ) de matriz de estados do método ODE.

Devido à dependência de  $\mathbf{R}$  em relação a  $\mathbf{w}(t)$ , (2.27) é uma equação diferencial não-linear. Uma maneira de estudar-se a estabilidade de seus pontos estacionários é pelo método de Lyapunov. Para determinar em que condições o ponto estacionário  $\tilde{\mathbf{w}}(t) \equiv \mathbf{0}$  (o que corresponde a  $\mathbf{w}(t) \equiv \mathbf{w}^*$ ) é estável, uma função de Lyapunov conveniente [19] é

$$L(\tilde{\mathbf{w}}(t)) = \frac{1}{2} \|\tilde{\mathbf{w}}(t)\|^2 = \frac{1}{2} \sum_{i=0}^{2M} \tilde{w}_i^2(t). \quad (2.30)$$

Se  $dL(\tilde{\mathbf{w}}(t))/dt \leq 0$  em alguma vizinhança da origem (com a igualdade só podendo valer na origem), então a origem é um ponto estacionário estável [32]. Usando (2.30) e (2.27), temos

$$\begin{aligned} \frac{dL(\tilde{\mathbf{w}}(t))}{dt} &= \sum_{i=0}^{2M} \tilde{w}_i(t) \frac{d\tilde{w}_i(t)}{dt} \\ &= -\tilde{\mathbf{w}}(t)^\top (t) \mathbf{R}(\mathbf{w}(t)) \tilde{\mathbf{w}}(t). \end{aligned} \quad (2.31)$$

Portanto, uma condição suficiente para que a origem seja um ponto estacionário estável da equação diferencial (2.27) é que a matriz  $\mathbf{R}(\mathbf{w})$  seja positiva definida em alguma vizinhança de  $\mathbf{w}^*$ .

A análise acima pode ser sintetizada na seguinte propriedade:

**Propriedade 2.1** *Suponha um algoritmo de adaptação da forma (2.7), a saída do sistema  $H(z)$  sendo modelado dada por (2.4) e que em  $\mathbf{w} = \mathbf{w}^*$  (2.11) seja válido. Sob as condições 1) e 2) do item 2.2.1 e para valores suficientemente pequenos de  $\mu$ , se as seguintes condições forem verdadeiras: 1)  $\psi(n)$  e  $\zeta(n)$  obedecem à expressão (2.12) e 2) a matriz de estados da ODE  $\mathbf{R}(\mathbf{w})$  é positiva definida em uma vizinhança de  $\mathbf{w}^*$  ou a matriz de estados da ODE  $\mathbf{R}'(\mathbf{w})$  é positiva definida em uma vizinhança de  $\mathbf{w}^*$ ; então  $\mathbf{w}^*$  é um ponto de convergência local do algoritmo, no sentido dado por (2.9).*

Notar que a estabilidade de (2.8) não depende de  $\mu$ , mas que a convergência de (2.7) segundo (2.9) depende de  $\mu$  ser suficientemente pequeno.

A propriedade acima estipula uma condição suficiente para a convergência do algoritmo. Uma condição necessária para a convergência é obtida como segue. Assumindo que a origem seja um ponto estacionário estável da equação diferencial (2.27) podemos aproximar, local-

mente, a solução desta equação pela solução da equação diferencial linear

$$\frac{d\Theta(t)}{dt} = -\mathbf{R}(\mathbf{w}^*)\Theta(t). \quad (2.32)$$

Notar que basta considerarmos  $\mathbf{R}(\mathbf{w}^*)$  pois de (2.19) e (2.24) resulta que

$$\mathbf{R}(\mathbf{w}^*) = \mathbf{R}'(\mathbf{w}^*). \quad (2.33)$$

Da teoria de sistemas lineares temos que a origem é um ponto estacionário estável de (2.32) se, e somente se, os autovalores de  $\mathbf{R}(\mathbf{w}^*)$  tiverem parte real negativa. Se a origem é um ponto estacionário estável da equação diferencial (2.27), necessariamente é um ponto estacionário estável da equação diferencial (2.32), uma vez que a solução desta aproxima localmente a de (2.27). Portanto, necessariamente, a parte real dos autovalores de  $\mathbf{R}(\mathbf{w}^*)$  é positiva. Disto resulta a propriedade:

**Propriedade 2.2** *Nas condições da propriedade 2.1, uma condição necessária para que  $\mathbf{w}^*$  seja um ponto de convergência do algoritmo é que a parte real dos autovalores de  $\mathbf{R}(\mathbf{w}^*)$  seja positiva.*

**Observação:** Se  $\mathbf{R}(\mathbf{w}^*)$  é simétrica, então tem autovalores reais e será positiva definida se os seus autovalores forem positivos. Supondo que  $\mathbf{R}(\mathbf{w})$  é contínua em  $\mathbf{w}^*$ , haverá então uma vizinhança de  $\mathbf{w}^*$  na qual  $\mathbf{R}(\mathbf{w})$  é positiva definida. Portanto, se (2.12) é verdadeira, a positividade de  $\mathbf{R}(\mathbf{w}^*)$  passa a ser, neste caso, uma condição necessária e suficiente para a convergência. Se  $\mathbf{R}(\mathbf{w}^*)$  não for simétrica isto não pode, a princípio, ser afirmado. Como será visto no Item 2.5, porém, para os algoritmos específicos que serão considerados, mesmo com  $\mathbf{R}(\mathbf{w}^*)$  não-simétrica resulta que sua positividade é necessária para que  $\mathbf{w}^*$  seja um ponto de convergência do algoritmo para qualquer sinal  $u(n)$  persistente de ordem  $2M + 1$ .

### Domínio de atração

Voltando agora à equação diferencial (2.27), o fato de a mesma ter  $\mathbf{w}(t) \equiv \mathbf{w}^*$  como ponto estacionário estável significa que existe alguma vizinhança  $W$  de  $\mathbf{w}^*$  tal que se  $\mathbf{w}(0) \in W$  então  $\lim_{t \rightarrow \infty} \mathbf{w}(t) = \mathbf{w}^*$ . Denomina-se de o *domínio de atração*  $W_d$  de  $\mathbf{w}^*$  a união de todas as vizinhanças de  $\mathbf{w}^*$  para a qual o limite anterior é válido. Da discussão anterior relativa ao método ODE segue que, para valores suficientemente pequenos de  $\mu$ ,  $W_d$  também é o

domínio de atração para a adaptação dos parâmetros dada por (2.7), isto é se  $\mathbf{w}(0) \in W_d$  então  $\mathbf{w}(n)$  tende a  $\mathbf{w}^*$  conforme (2.10).

No âmbito da análise das características de filtros adaptativos pode ser de interesse determinar o domínio de atração de  $\mathbf{w}^*$  para ter-se uma idéia de quão boa precisa ser a estimativa inicial  $\mathbf{w}(0)$  de  $\mathbf{w}^*$  para ocorrer a convergência desejada. A vizinhança  $W_p$  de  $\mathbf{w}^*$  na qual a matriz  $\mathbf{R}$  ou a matriz  $\mathbf{R}'$  são positivas pode ser empregada para esta caracterização. Notar, de início, que não é possível afirmar que  $W_p \subseteq W_d$ , pois não é possível afirmar que se  $\mathbf{w}(0) \in W_p$  então  $\mathbf{w}(t) \in W_p$ , para qualquer  $t$ . Consideremos agora uma bola de raio  $\rho$ :  $B_\rho = \{\mathbf{w} \mid \|\tilde{\mathbf{w}}\| \leq \rho\}$  e o maior raio  $r$  tal que a bola  $B_r$  esteja contida em  $W_p$ , isto é

$$r = \max\{\rho \mid B_\rho \subseteq W_p\}. \quad (2.34)$$

Temos então que  $B_r$  está contida no domínio de atração  $W_d$  de  $\mathbf{w}^*$ , pois agora a própria função de Lyapunov escolhida (expressão (2.30)) garante que se  $\mathbf{w}(0) \in B_r$  então  $\mathbf{w}(t) \in W_p$ , para qualquer  $t$  (uma prova formal de um resultado que implica em que  $B_r \subseteq W_d$  é dada em [32], p. 196). A seguinte propriedade, portanto, pode ser enunciada:

**Propriedade 2.3** *Nas condições da propriedade 2.1, o domínio de atração  $W_d$  de  $\mathbf{w}^*$  é dado, ao menos, pela maior bola  $B_r$  contida no conjunto  $W_p$  onde  $\mathbf{R}$  ou  $\mathbf{R}'$  são positivas definidas.*

### 2.2.3 Limites para o passo de adaptação

Em aplicações práticas de filtros adaptativos, deseja-se que o tempo decorrido do início da adaptação até o instante em que uma vizinhança satisfatória do ponto  $\mathbf{w}^*$  é atingida seja o menor possível. O parâmetro sob controle do usuário que determina diretamente esta velocidade de convergência (juntamente com outros fatores que não estão sob controle do usuário, como o condicionamento de  $\mathbf{R}(\mathbf{w}^*)$ ) é o passo de adaptação  $\mu$ . Notar que apesar da equação diferencial (2.8) usada no método ODE não depender de  $\mu$ , para retornar ao algoritmo de adaptação deve-se empregar a relação  $t = n\mu$ . Ou seja, fixando as demais variáveis, se é atendida a condição de  $\mu$  ser "suficientemente pequeno" estipulada na Propriedade 2.1, a velocidade de convergência é diretamente proporcional a  $\mu$ . Isto conduz, portanto, a se querer usar o maior valor possível para  $\mu$ . É de interesse prático, então, poder relacionar este maior valor possível de  $\mu$  com características de  $H(z)$  que possam eventualmente ser

estimadas *a priori* (como por exemplo, a proximidade dos seus polos ao círculo unitário), pois assim encurta-se a etapa de tentativa e erro para a escolha de  $\mu$ .

Para fins práticos, porém, o método de análise ODE, não indica diretamente quão "suficientemente pequeno"  $\mu$  deve ser. Limites superiores para  $\mu$  podem ser obtidos a partir da aproximação local dada por (2.32), como visto a seguir.

### Utilização da aproximação local

Inicialmente, recapitulemos o caminho percorrido para chegar na expressão (2.32): para valores suficientemente pequenos de  $\mu$ , os parâmetros determinados pelo algoritmo de adaptação (2.7) são aproximados pela solução da equação diferencial (2.8), segundo (2.9); reformulando (2.8) em termos do erro de parâmetros foi obtida a equação diferencial (2.27); finalmente, uma linearização em torno de  $\mathbf{w}^*$  resultou em (2.32).

Se, pelos motivos que ficarão claros a seguir, quiser-se retornar agora de (2.32) a uma expressão de tempo discreto, podemos simplesmente usar a aproximação de uma equação diferencial linear por uma equação de diferenças. Por exemplo, para o caso de primeira ordem:

$$\begin{aligned} \frac{df(t)}{dt} = af(t) &\Rightarrow \frac{f((n+1)\Delta t) - f(n\Delta t)}{\Delta t} \simeq af(n\Delta t) \\ &\Rightarrow f((n+1)\Delta t) - f(n\Delta t) \simeq \Delta t af(n\Delta t), \quad \Delta t \simeq 0, \end{aligned} \quad (2.35)$$

e, portanto, a sequência  $g(n)$  obtida recursivamente como

$$g(n+1) - g(n) = \mu ag(n) \Rightarrow g(n+1) = (1 - \mu a)g(n), \quad (2.36)$$

para  $\mu = \Delta t$  suficientemente pequeno aproxima  $f(t)$  com  $t = n\Delta t = n\mu$ . A equação diferencial (2.32), para  $\mu$  suficientemente pequeno, é aproximada, portanto, por

$$\begin{aligned} \boldsymbol{\theta}(n+1) &= \boldsymbol{\theta}(n) - \mu \mathbf{R}(\mathbf{w}^*) \boldsymbol{\theta}(n) \\ &= [\mathbf{I} - \mu \mathbf{R}(\mathbf{w}^*)] \boldsymbol{\theta}(n). \end{aligned} \quad (2.37)$$

Reunindo (2.27), (2.32) e (2.37), consideremos então que para  $\mu$  suficientemente pequeno (o que podemos denotar por  $\mu < \mu_1$  para algum  $\mu_1$ ), tenhamos, localmente em torno da

origem,

$$\|\tilde{\mathbf{w}}(n\mu) - \boldsymbol{\theta}(n)\| < C_1\delta_1(\mu), \quad \mu < \mu_1, \quad (2.38)$$

onde  $C_1$  é uma constante positiva limitada e  $\delta_1(\mu)$  é positiva e tende a zero com  $\mu$ . Notar que, neste caso, se (2.27) é estável, a condição  $\mu < \mu_1$  equivale simplesmente a impor que (2.37) seja estável. Usando (2.38) podemos escrever agora

$$\begin{aligned} \|\tilde{\mathbf{w}}(n) - \boldsymbol{\theta}(n)\| &= \|\tilde{\mathbf{w}}(n) - \tilde{\mathbf{w}}(n\mu) + \tilde{\mathbf{w}}(n\mu) - \boldsymbol{\theta}(n)\| \\ &\leq \|\tilde{\mathbf{w}}(n) - \tilde{\mathbf{w}}(n\mu)\| + \|\tilde{\mathbf{w}}(n\mu) - \boldsymbol{\theta}(n)\| \\ &< \|\tilde{\mathbf{w}}(n) - \tilde{\mathbf{w}}(n\mu)\| + C_1\delta_1(\mu), \quad \mu < \mu_1, \end{aligned} \quad (2.39)$$

e, portanto,

$$\sup_n \|\tilde{\mathbf{w}}(n) - \boldsymbol{\theta}(n)\| < \sup_n \|\tilde{\mathbf{w}}(n) - \tilde{\mathbf{w}}(n\mu)\| + C_1\delta_1(\mu), \quad \mu < \mu_1. \quad (2.40)$$

Por outro lado, com

$$\begin{aligned} \|\mathbf{w}(n) - \mathbf{w}(n\mu)\| &= \|\mathbf{w}^* - \mathbf{w}(n) - \mathbf{w}^* + \mathbf{w}(n\mu)\| \\ &= \|\tilde{\mathbf{w}}(n) - \tilde{\mathbf{w}}(n\mu)\|, \end{aligned} \quad (2.41)$$

(2.9) pode ser reescrita em termos do erro de parâmetros como

$$\Pr \left\{ \sup_n \|\tilde{\mathbf{w}}(n) - \tilde{\mathbf{w}}(n\mu)\| > C\delta(\mu) \right\} < C'\delta(\mu), \quad \mu < \mu_2, \quad (2.42)$$

para algum  $\mu_2$ , onde não necessariamente  $\mu_2$  é igual a  $\mu_1$  em (2.40). Reunindo esta expressão e (2.40) temos então

$$\begin{aligned} C'\delta(\mu) &> \Pr \left\{ \sup_n \|\tilde{\mathbf{w}}(n) - \tilde{\mathbf{w}}(n\mu)\| > C\delta(\mu) \right\} \\ &> \Pr \left\{ \sup_n \|\tilde{\mathbf{w}}(n) - \boldsymbol{\theta}(n)\| - C_1\delta_1(\mu) > C\delta(\mu) \right\} \end{aligned}$$



$$= \Pr \left\{ \sup_n \|\tilde{\mathbf{w}}(n) - \boldsymbol{\theta}(n)\| > C\delta(\mu) + C_1\delta_1(\mu) \right\}, \quad (2.43)$$

$$\mu < \min(\mu_1, \mu_2) \leq \mu_1. \quad (2.44)$$

Portanto, para  $\mu$  suficientemente pequeno, a adaptação dos parâmetros dada por (2.7) é aproximada localmente por (2.37), segundo (2.43), ou seja, no mesmo sentido em que (2.7) é aproximada por (2.8). Consequentemente, se para um certo  $\mu < \min(\mu_1, \mu_2)$  o vetor  $\mathbf{w}(n)$  em (2.7) convergir para  $\mathbf{w}^*$  (no sentido sendo utilizado aqui) então, necessariamente (2.37) é estável, ou seja  $\mu < \mu_1$ . E, chegando ao aspecto que motivou esta discussão, uma condição  $\mu < \mu_1$  que seja necessária para a estabilidade de (2.37) será necessária para a convergência de (2.7). Isto pode ser formalizado na seguinte propriedade:

**Propriedade 2.4** *Se a equação diferencial linear (2.27) é estável, uma condição relativa a  $\mu$  que seja necessária para a estabilidade da equação de diferenças linear (2.37) é uma condição necessária para a convergência local (no sentido dado por (2.9)) do algoritmo de adaptação dado por (2.7).*

No sentido oposto, porém, o mesmo não pode ser afirmado: uma condição relativa a  $\mu$  que seja necessária para a convergência de (2.7) pode não ser, devido às aproximações realizadas, uma condição necessária para a estabilidade de (2.37). Para obter uma condição relativa a  $\mu$  necessária e suficiente para a convergência do algoritmo, seria preciso levar em conta a não-linearidade em (2.27) ou diretamente em (2.7), que dependem do algoritmo de adaptação particular sendo considerado. Por exemplo, podem-se obter tais condições, usando o conceito de hiper-estabilidade, para o algoritmo denominado HARF [28], do qual o algoritmo SHARF (que é um dos algoritmos considerados neste trabalho) é uma simplificação (a condição obtida para o algoritmo HARF, porém, não se mantém para o algoritmo SHARF). Para os outros algoritmos considerados neste trabalho, condições para convergência resultantes de um tratamento exato da não-linearidade envolvida não são conhecidas.

A seguir, a partir de (2.37) são obtidos limites superiores para  $\mu$ , necessários para a sua estabilidade.

## Limites superiores para $\mu$

Inicialmente nota-se que se  $\mathbf{R}(\mathbf{w}^*)$  for positiva definida e simétrica, a equação de diferenças (2.37) tem a mesma forma que a equação obtida na análise da convergência de filtros FIR adaptativos [12]. Os resultados aí obtidos podem então ser aproveitados, localmente, para o caso IIR: deve-se ter  $0 < \mu < 2/\lambda_{\max}$ , onde  $\lambda_{\max}$  é o maior autovalor de  $\mathbf{R}(\mathbf{w}^*)$ ; e se o espalhamento dos autovalores  $\lambda_{\max}/\lambda_{\min}$  (onde  $\lambda_{\min}$  é o menor autovalor de  $\mathbf{R}(\mathbf{w}^*)$ ) for elevado, a velocidade de convergência local, partindo de um ponto inicial qualquer, será, em geral, baixa.

Como será visto no itens a seguir, nem todos os algoritmos de adaptação resultam em uma matriz  $\mathbf{R}(\mathbf{w}^*)$  simétrica. Devido a isso, a análise feita para filtros FIR é generalizada a seguir, obtendo-se condições relacionadas aos valores singulares de  $\mathbf{R}(\mathbf{w}^*)$ , à parte real e ao valor absoluto dos seus autovalores.

No que segue, para maior simplicidade de notação, abandonamos momentaneamente o argumento de  $\mathbf{R}(\mathbf{w}^*)$ . Da teoria de sistemas lineares de tempo discreto temos que a equação de diferenças (2.37) é estável se, e somente se, os autovalores de  $(\mathbf{I} - \mu\mathbf{R})$  tiverem valor absoluto menor do que 1. Estes, por sua vez, são relacionados com os autovalores de  $\mathbf{R}$ : para os autovetores  $\mathbf{u}_i$  de  $(\mathbf{I} - \mu\mathbf{R})$  vale a expressão  $(\mathbf{I} - \mu\mathbf{R})\mathbf{u}_i = \lambda'_i \mathbf{u}_i$ , que pode ser re-arranjada como

$$\begin{aligned} \mathbf{R}\mathbf{u}_i &= \frac{1 - \lambda'_i}{\mu} \mathbf{u}_i \\ &= \lambda_i \mathbf{u}_i, \end{aligned} \tag{2.45}$$

ou seja,  $\mathbf{u}_i$  também é um autovetor de  $\mathbf{R}$ , associado ao autovalor  $\lambda_i = (1 - \lambda'_i)/\mu$ . Para que todos os autovetores  $\lambda'_i = 1 - \mu\lambda_i$  obedeçam a  $|\lambda'_i| < 1$  é necessário, portanto, que todos os autovetores  $\lambda_i$  de  $\mathbf{R}$  obedeçam a  $|1 - \mu\lambda_i| < 1$ , o que equivale a  $1 - \mu(\lambda_i + \lambda_i^*) + \mu^2 |\lambda_i|^2 < 1$ . Ou seja, assumindo que (2.32) é estável e portanto  $\text{Re}[\lambda_i] > 0$ , a condição a ser atendida pelo passo de adaptação  $\mu$ , necessária e suficiente para que (2.37) seja estável, e, portanto, conforme a Propriedade 2.4, necessária para a convergência de (2.7), é

$$0 < \mu < \frac{2 \text{Re}[\lambda_i]}{|\lambda_i|^2}, \quad i = 1, 2, \dots, 2M + 1. \tag{2.46}$$

Esta expressão pode ser útil para obter um valor máximo de  $\mu$  para um dado sistema  $H(z)$ .

Porém, ao analisar o comportamento do valor máximo de  $\mu$  com características mais genéricas de  $H(z)$ , como feito no Item 2.6, é de interesse utilizar os valores singulares de  $\mathbf{R}$  ao invés de seus autovalores.

Para obter uma relação com os valores singulares de  $\mathbf{R}$ , consideremos inicialmente um critério mais rigoroso de convergência: deseja-se uma condição necessária para que a norma da aproximação do erro dos coeficientes decresça a cada iteração. Isto é, deseja-se uma condição para que

$$\|\boldsymbol{\theta}(n+1)\| < \|\boldsymbol{\theta}(n)\|, \quad (2.47)$$

critério análogo ao determinado pela função de Lyapunov (2.30) escolhida para analisar a estabilidade da equação diferencial (2.27). Reunindo (2.37) e (2.47), e abandonando o argumento  $n$  temos que para qualquer  $\boldsymbol{\theta}$  deve valer

$$\boldsymbol{\theta}^\top \boldsymbol{\theta} - \mu 2\boldsymbol{\theta}^\top \mathbf{R}\boldsymbol{\theta} + \mu^2 \boldsymbol{\theta}^\top \mathbf{R}^\top \mathbf{R}\boldsymbol{\theta} < \boldsymbol{\theta}^\top \boldsymbol{\theta}, \quad (2.48)$$

o que leva a

$$\mu(\mu \boldsymbol{\theta}^\top \mathbf{R}^\top \mathbf{R}\boldsymbol{\theta}^\top - 2\boldsymbol{\theta}^\top \mathbf{R}\boldsymbol{\theta}^\top) < 0. \quad (2.49)$$

Impondo  $\mu > 0$ , temos então

$$0 < \mu < \frac{2\boldsymbol{\theta}^\top \mathbf{R}\boldsymbol{\theta}^\top}{\boldsymbol{\theta}^\top \mathbf{R}^\top \mathbf{R}\boldsymbol{\theta}^\top}. \quad (2.50)$$

A primeira condição necessária para que esta expressão seja válida é que  $\mathbf{R}(\mathbf{w}^*)$  seja positiva definida. Mas, como visto no item anterior, a análise não local da convergência pelo método da equação diferencial associada já conduz à necessidade desta condição. Podemos considerá-la atendida, portanto. Temos então, usando a desigualdade de Schwarz,  $\boldsymbol{\theta}^\top \mathbf{R}\boldsymbol{\theta}^\top = |\boldsymbol{\theta}^\top \mathbf{R}\boldsymbol{\theta}^\top| \leq \|\boldsymbol{\theta}\| \|\mathbf{R}\boldsymbol{\theta}^\top\|$ , e

$$0 < \mu < \frac{2\boldsymbol{\theta}^\top \mathbf{R}\boldsymbol{\theta}^\top}{\boldsymbol{\theta}^\top \mathbf{R}^\top \mathbf{R}\boldsymbol{\theta}^\top} \leq \frac{2\|\boldsymbol{\theta}\| \|\mathbf{R}\boldsymbol{\theta}^\top\|}{\boldsymbol{\theta}^\top \mathbf{R}^\top \mathbf{R}\boldsymbol{\theta}^\top} = \frac{2\|\boldsymbol{\theta}\|}{\sqrt{\boldsymbol{\theta}^\top \mathbf{R}^\top \mathbf{R}\boldsymbol{\theta}^\top}}. \quad (2.51)$$

Uma vez que esta expressão tem que ser válida para qualquer  $\boldsymbol{\theta}$ , podemos escrever

$$0 < \mu < \min \left\{ \frac{2\|\boldsymbol{\theta}\|}{\sqrt{\boldsymbol{\theta}^\top \mathbf{R}^\top \mathbf{R}\boldsymbol{\theta}^\top}} \right\}. \quad (2.52)$$

Usando a Propriedade 1.1, temos finalmente que para (2.47) ser válida, é necessário que

$$0 < \mu < \frac{2}{\sigma_{\max}} \quad (2.53)$$

onde  $\sigma_{\max} = \sqrt{\max\{\lambda_i(\mathbf{R}^\top(\mathbf{w}^*)\mathbf{R}(\mathbf{w}^*))\}}$  é o maior valor singular de  $\mathbf{R}(\mathbf{w}^*)$ . Temos portanto que quando o valor singular máximo de  $\mathbf{R}(\mathbf{w}^*)$  tende a infinito, para que (2.47) seja válida  $\mu$  deve tender a zero. Notar que quando  $\mathbf{R}(\mathbf{w}^*)$  é simétrica e positiva definida  $\sigma_{\max} = \lambda_{\max}$  e esta condição recai na condição obtida para filtros FIR, mencionada anteriormente.

Para obter uma condição envolvendo  $\sigma_{\max}$  que seja necessária para a estabilidade de (2.37), não impondo, porém, a condição (2.47), poderíamos, lembrando que  $\text{Re}[\lambda_i] > 0$ , escrever (2.46) como

$$0 < \mu < \frac{2 \text{Re}[\lambda_i]}{|\lambda_i|^2} \leq \frac{2|\lambda_i|}{|\lambda_i|^2} = \frac{2}{|\lambda_i|}, \quad i = 1, 2, \dots, 2M + 1. \quad (2.54)$$

Se uma relação  $\max\{|\lambda_i|\} \geq k(M)\sigma_{\max}$ ,  $0 < k(M) \leq 1$ , fosse disponível, onde  $k(M)$  depende da ordem  $M$  de  $H(z)$ , uma condição necessária para a convergência do algoritmo seria  $\mu < 2/k(M)\sigma_{\max}$ . Porém, apesar de uma condição do tipo  $\max\{|\lambda_i|\} \geq k(M)\sigma_{\max}$  parecer ser verdadeira quando  $\mathbf{R}(\mathbf{w}^*)$  é positiva definida, não foi possível demonstrá-la.

## 2.2.4 Velocidade de convergência local

Passamos agora a analisar a questão da velocidade de convergência local dos algoritmos de adaptação de passo constante. Consideremos, inicialmente, a questão da velocidade de convergência de (2.37). Assumindo que  $\mathbf{R}(\mathbf{w}^*)$  é diagonalizável, podemos escrever esta expressão como

$$\boldsymbol{\theta}(n+1) = [\mathbf{I} - \mu \mathbf{P} \boldsymbol{\Lambda} \mathbf{P}^{-1}] \boldsymbol{\theta}(n), \quad (2.55)$$

onde as colunas de  $\mathbf{P}$  são os autovetores de norma unitária de  $\mathbf{R}(\mathbf{w}^*)$  e a matriz diagonal  $\boldsymbol{\Lambda}$  contém os seus autovalores. Definindo  $\boldsymbol{\varphi}(n) = \mathbf{P}^{-1}\boldsymbol{\theta}(n)$  e  $\varphi_i$  sendo os elementos de  $\boldsymbol{\varphi}$ , a fórmula recursiva dos modos normais de (2.37) é então

$$\varphi_i(n+1) = (1 - \mu \lambda_i) \varphi_i(n). \quad (2.56)$$

A velocidade de convergência de (2.37) pode ser caracterizada em termos do decaimento, por iteração, da energia dos elementos de  $\boldsymbol{\varphi}$ , que, usando (2.56) é dado por

$$\frac{|\varphi_i(n+1)|^2}{|\varphi_i(n)|^2} = |(1-\mu\lambda_i)|^2. \quad (2.57)$$

Desta maneira, quando algum  $|(1-\mu\lambda_i)|^2$  se aproxima de 1 considera-se que a velocidade de convergência local se aproxima de zero.

### Utilização dos modos normais: efeito da simetria de $\mathbf{R}(\mathbf{w}^*)$

Pela expressão (2.43) podemos ver que, diminuindo  $\mu$ , pode-se fazer, localmente, o erro de parâmetros  $\tilde{\mathbf{w}}(n)$  ser tão próximo quanto se queira de  $\boldsymbol{\theta}(n)$ , e, portanto, para um valor suficientemente pequeno de  $\mu$ , o comportamento local de (2.7) pode ser descrito por (2.56), com  $\tilde{\mathbf{w}}(n) \simeq \mathbf{P}\boldsymbol{\varphi}(n)$ . Podemos ver isto de uma outra maneira, que permite destacar diferenças entre os algoritmos específicos: re-escrevendo inicialmente (2.7) em termos do erro de parâmetros  $\tilde{\mathbf{w}}(n) = \mathbf{w}^* - \mathbf{w}(n)$ , temos

$$\tilde{\mathbf{w}}(n+1) = [\mathbf{I} - \mu\boldsymbol{\psi}(n)\boldsymbol{\phi}_1^\top(n)] \tilde{\mathbf{w}}(n), \quad (2.58)$$

onde foi usado (2.18) e foi assumido, para maior simplicidade, ruído de medida nulo. Definindo agora

$$\begin{aligned} \mathbf{R}_d(n) &= \boldsymbol{\psi}(n)\boldsymbol{\phi}_1^\top(n) - E[\boldsymbol{\psi}(n, \mathbf{w}^*)\boldsymbol{\phi}_1^\top(n, \mathbf{w}^*)] \\ &= \boldsymbol{\psi}(n)\boldsymbol{\phi}_1^\top(n) - \mathbf{R}(\mathbf{w}^*) \end{aligned} \quad (2.59)$$

(ver (2.26)), temos

$$\tilde{\mathbf{w}}(n+1) = [\mathbf{I} - \mu\mathbf{R}(\mathbf{w}^*)] \tilde{\mathbf{w}}(n) - \mu\mathbf{R}_d(n)\tilde{\mathbf{w}}(n). \quad (2.60)$$

E, como acima, escrevendo  $\mathbf{R}(\mathbf{w}^*) = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}$ , e definindo  $\mathbf{v}(n) = \mathbf{P}^{-1}\tilde{\mathbf{w}}(n)$ , podemos escrever para os elementos  $v_i$  de  $\mathbf{v}$

$$v_i(n+1) = (1 - \mu\lambda_i)v_i(n) - \mu\mathbf{q}_i^\top \mathbf{R}_d(n)\tilde{\mathbf{w}}(n), \quad (2.61)$$

onde  $\mathbf{q}_i^\top$  é  $i$ -ésima linha de  $\mathbf{P}^{-1}$ . Comparando a expressão acima com a expressão (2.56) podemos ver que passou-se de um sistema linear livre para um sistema linear forçado, e que, para  $\mu$  suficientemente pequeno, a solução da expressão acima pode ser aproximada localmente pela de (2.56). É importante notar, porém, que este valor suficientemente pequeno de  $\mu$  pode ser menor do que o valor de  $\mu$  para o qual a convergência local de (2.7) é garantida pela Propriedade 2.1. Isto é, para um dado valor de  $\mu$ , o algoritmo pode convergir localmente mas não conforme o sistema livre dado por (2.56).

A expressão (2.61) é útil para obter uma relação qualitativa entre o tipo do algoritmo em questão e o quão pequeno precisa ser  $\mu$  para poder-se descrever o comportamento local deste algoritmo por (2.56). Para tanto, pode-se notar inicialmente que se  $\mathbf{R}(\mathbf{w}^*)$  é simétrica e tem autovalores distintos, então os seus autovetores são ortogonais e portanto  $\mathbf{P}^{-1} = \mathbf{P}^\top$ . As linhas  $\mathbf{q}_i^\top$  de  $\mathbf{P}^{-1}$  em (2.61) são então os próprios autovetores transpostos de  $\mathbf{R}(\mathbf{w}^*)$  e  $\|\mathbf{q}_i\| = 1$ . Por outro lado, se  $\mathbf{R}(\mathbf{w}^*)$  não é simétrica então os seus autovetores não são necessariamente ortogonais. Consideremos o que ocorre quando estes autovetores  $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{2M+1}$  são quase linearmente dependentes obedecendo a

$$\beta_1 \mathbf{p}_1 + \beta_2 \mathbf{p}_2 + \dots + \beta_{2M+1} \mathbf{p}_{2M+1} \simeq \mathbf{0}, \quad (2.62)$$

para algum conjunto  $\{\beta_i\}$  de números reais não todos nulos, com  $\beta_i = \beta_j$  se  $\mathbf{p}_i = \mathbf{p}_j^*$ . Se fizermos

$$\mathbf{z} = \left[ \beta_1 \quad \beta_1 \quad \dots \quad \beta_{2M+1} \right]^\top, \quad (2.63)$$

então teremos  $\mathbf{P}\mathbf{z} \simeq \mathbf{0}$ , e, portanto o menor autovalor de  $\mathbf{P}^\top \mathbf{P}$ , conforme a Propriedade 1.1, obedece a

$$\min\{\lambda_i(\mathbf{P}^\top \mathbf{P})\} = \min_{\mathbf{x}} \frac{\mathbf{x}^\top \mathbf{P}^\top \mathbf{P} \mathbf{x}}{\mathbf{x}^\top \mathbf{x}} \leq \frac{\mathbf{z}^\top \mathbf{P}^\top \mathbf{P} \mathbf{z}}{\mathbf{z}^\top \mathbf{z}} \simeq 0. \quad (2.64)$$

O maior autovalor de  $\mathbf{P}^{-1} \mathbf{P}^{-\top}$  vale  $1/\min\{\lambda_i(\mathbf{P}^\top \mathbf{P})\}$  e é, portanto, grande. Segue que o traço de  $\mathbf{P}^{-1} \mathbf{P}^{-\top}$ , que é igual à soma dos autovalores de  $\mathbf{P}^{-1} \mathbf{P}$ , também é grande. Uma vez que o traço de  $\mathbf{P}^{-1} \mathbf{P}^{-\top}$  também é igual a soma das normas das linhas de  $\mathbf{P}^{-1}$ , segue que alguma linha  $\mathbf{q}_i^\top$  de  $\mathbf{P}^{-1}$  tem norma grande. Portanto, em (2.61) a norma do termo  $\mathbf{q}_i^\top \mathbf{R}_d(n) \tilde{\mathbf{w}}(n)$  pode ser bem maior do que no caso de  $\mathbf{R}(\mathbf{w}^*)$  simétrica, quando  $\mathbf{q}_i^\top$  tem norma unitária. Consequentemente, para algoritmos em que  $\mathbf{R}(\mathbf{w}^*)$  não é simétrica, o valor máximo de  $\mu$  para valer a aproximação de (2.61) por (2.56) pode ser bem menor do que

quando  $\mathbf{R}(\mathbf{w}^*)$  é simétrica. Este fato será constatado nos exemplos apresentados no Item 2.7.

### Efeito do espalhamento dos valores singulares

Assumindo que o passo de adaptação  $\mu$  seja suficientemente pequeno para poder-se aproximar a solução de (2.61) pela de (2.56), verifica-se agora como o espalhamento dos valores singulares de  $\mathbf{R}(\mathbf{w}^*)$  afeta o decaimento dos modos normais de (2.37). Inicialmente, desenvolvemos (2.57) como

$$\frac{|\varphi_i(n+1)|^2}{|\varphi_i(n)|^2} = |(1-\mu\lambda_i)|^2 = 1 - 2\mu \operatorname{Re}[\lambda_i] + \mu^2 |\lambda_i|^2. \quad (2.65)$$

Consideremos agora que um limite superior para  $\mu$  mais restritivo do que (2.46) é obedecido:

$$0 < \mu < \frac{\operatorname{Re}[\lambda_i]}{|\lambda_i|^2}, \quad i = 1, 2, \dots, 2M + 1. \quad (2.66)$$

Usando em (2.65): o limite acima, a Propriedade 1.2, e o fato de que  $\operatorname{Re}[\lambda_i] > 0$  (pois  $\mathbf{w}^*$  é um ponto de convergência local), temos

$$\frac{|\varphi_i(n+1)|^2}{|\varphi_i(n)|^2} < 1 - \mu^2 |\lambda_i|^2 \leq 1 - \mu^2 \sigma_{\min}^2 \leq 1, \quad (2.67)$$

onde  $\sigma_{\min}$  é o menor valor singular de  $\mathbf{R}(\mathbf{w}^*)$ . Ou seja, a condição (2.66) determina que a energia em cada modo normal é sempre decrescente.

Escrevendo agora  $\mu$  como

$$\mu = \frac{\alpha}{\sigma_{\max}}, \quad (2.68)$$

para um certo  $\alpha > 0$ , temos então

$$0 \leq \frac{|\varphi_i(n+1)|^2}{|\varphi_i(n)|^2} < 1 - \left( \alpha \frac{\sigma_{\min}}{\sigma_{\max}} \right)^2, \quad i = 1, 2, \dots, 2M + 1. \quad (2.69)$$

Para um determinado  $\alpha$ , portanto, à medida em que aumenta o espalhamento dos valores singulares de  $\mathbf{R}(\mathbf{w}^*)$ , o limite superior para o decaimento dos modos normais de (2.37) se aproxima de 1. Assim, uma condição suficiente para que a velocidade de convergência local de (2.37) não se aproxime de zero é que  $\alpha$  e  $\sigma_{\min}/\sigma_{\max}$  (o inverso do espalhamento dos valores singulares de  $\mathbf{R}(\mathbf{w}^*)$ ) não se aproximem de zero. No Item 2.6 é analisado, para

alguns algoritmos de passo constante comumente usados, como o espalhamento dos valores singulares de  $\mathbf{R}(\mathbf{w}^*)$  se comporta quando  $H(z)$  tem polos próximos à circunferência unitária. Em relação a  $\alpha$ , como já discutido acima, para obter condições mais restritivas do que (2.46) e (2.53) é necessário ir além da aproximação local dada por (2.37). Apesar de tratar-se de um problema de interesse a ser investigado, esta questão não será abordada teoricamente neste trabalho, mas apenas em termos de resultados obtidos da prática.

Notar, por fim, que sem especificar mais a forma de  $\mathbf{R}(\mathbf{w}^*)$ , não pode-se afirmar que um alto espalhamento dos seus valores singulares é suficiente para que a velocidade de convergência se aproxime de zero. Isto pode ser visto notando que é possível construir uma matriz a partir de autovetores quase linearmente dependentes (o que, analogamente ao discutido para a expressão (2.64) resulta em que para seu valor singular mínimo vale  $\sigma_{\min} \simeq 0$ ) e autovalores quaisquer. Nos casos observados na prática, porém, como os apresentados no Item 2.7, um alto espalhamento dos valores singulares de  $\mathbf{R}(\mathbf{w}^*)$  sempre resultou em uma baixa velocidade de convergência.

### Decaimento do modo normal mais lento

No caso em que é usado um valor de  $\mu$  muito menor do que o limite superior dado por (2.46) (o que, na prática, é necessário na maior parte dos casos observados), e denotando por  $\varphi_m$  o componente (ou par de componentes conjugados) de  $\varphi$  associado ao autovalor  $\lambda_m$  de  $\mathbf{R}(\mathbf{w}^*)$  de menor parte real, aproximamos (2.65) com  $i = m$  como

$$\frac{|\varphi_m(n+1)|^2}{|\varphi_m(n)|^2} \simeq 1 - 2\mu \operatorname{Re}[\lambda_m], \quad \mu \ll \frac{2 \operatorname{Re}[\lambda_m]}{|\lambda_m|^2}. \quad (2.70)$$

Passando (2.70) para a escala logarítmica e usando a aproximação  $\ln(1+x) \simeq x$ ,  $x \ll 1$ , temos

$$10 \log \frac{|\varphi_m(n+1)|^2}{|\varphi_m(n)|^2} \simeq -20 \log(e) \mu \operatorname{Re}[\lambda_m], \quad \mu \ll \frac{2 \operatorname{Re}[\lambda_m]}{|\lambda_m|^2}. \quad (2.71)$$

Esta expressão fornece então o decaimento em dB/iteração da energia do modo normal mais lento do sistema associado ao algoritmo de adaptação em (2.37). Se  $\mu$  é suficientemente pequeno para valer a aproximação de (2.61) por (2.56), então, para um  $\varphi(0)$  genérico, este modo normal mais lento dominará os demais após um certo número de iterações e o erro de



parâmetros será dado, no caso de  $\lambda_m$  ser real por

$$\tilde{\mathbf{w}}(n) = \mathbf{P}\mathbf{v}(n) \simeq \mathbf{P}\varphi(n) = \varphi_m(n)\mathbf{p}_m, \quad (2.72)$$

onde  $\mathbf{p}_m$  é o autovetor de  $\mathbf{R}(\mathbf{w}^*)$  associado a  $\lambda_m$ , caso em que

$$\|\tilde{\mathbf{w}}(n)\| = \|\varphi_m(n)\mathbf{p}_m\| = |\varphi_m(n)|. \quad (2.73)$$

No caso de  $\lambda_m$  ser complexo temos

$$\tilde{\mathbf{w}}(n) \simeq \varphi_m(n)\mathbf{p}_m + \varphi_m^*(n)\mathbf{p}_m^*, \quad (2.74)$$

de onde resulta que cada elemento  $\tilde{w}_i$  de  $\tilde{\mathbf{w}}(n)$  é dado por

$$\tilde{w}_i = |\varphi_m(n)| |p_{m,i}| \cos\{\arg[\varphi_m(n)] + \arg[p_{m,i}]\}, \quad (2.75)$$

e, portanto,

$$\begin{aligned} \|\tilde{\mathbf{w}}(n)\| &= |\varphi_m(n)| \left\{ \sum_{i=1}^{2M+1} |p_{m,i}|^2 \cos^2\{\arg[\varphi_m(n)] + \arg[p_{m,i}]\} \right\}^{1/2} \\ &\leq |\varphi_m(n)| \left\{ \sum_{i=1}^{2M+1} |p_{m,i}|^2 \right\}^{1/2} \\ &= |\varphi_m(n)| \|\mathbf{p}_m\| \\ &= |\varphi_m(n)|. \end{aligned} \quad (2.76)$$

Portanto, a norma  $\|\tilde{\mathbf{w}}(n)\|$  do erro de parâmetros, ou o limite superior para o envelope desta norma no caso complexo, são dados por  $|\varphi_m(n)|$ . Usando aqui, para maior simplicidade,  $\|\cdot\|$  para denotar tanto a norma quanto o envelope da norma temos então, usando (2.71):

$$10 \log \frac{\|\tilde{\mathbf{w}}(n+1)\|^2}{\|\tilde{\mathbf{w}}(n)\|^2} \simeq -20 \log(e)\mu \operatorname{Re}[\lambda_m], \quad \mu \ll \frac{2 \operatorname{Re}[\lambda_m]}{|\lambda_m|^2}. \quad (2.77)$$

O erro quadrático médio de aproximação relaciona-se localmente com o erro de parâmet-

ros, conforme visto adiante no Item 3.4.2, por

$$E[\varepsilon^2(n)] = \frac{1}{2} \tilde{\mathbf{w}}^\top(n) \mathbf{H} \tilde{\mathbf{w}}(n), \quad (2.78)$$

onde  $\mathbf{H}$  é a matriz Hessiana do erro. Temos então, no caso real,

$$E[\varepsilon^2(n)] = |\varphi_m(n)|^2 \mathbf{p}_m^\top \mathbf{H} \mathbf{p}_m. \quad (2.79)$$

No caso complexo, um limite superior do envelope do erro, é dado por  $|\varphi_m(n)|^2 |\mathbf{p}_m^\top| \mathbf{H} |\mathbf{p}_m|$ . Valendo então a aproximação de (2.61) por (2.56), o decaimento local do erro quadrático de aproximação (ou do limite superior de seu envelope) é igual ao do erro de parâmetros:

$$\frac{E[\varepsilon^2(n+1)]}{E[\varepsilon^2(n)]} = \frac{\|\tilde{\mathbf{w}}(n+1)\|^2}{\|\tilde{\mathbf{w}}(n)\|^2} = \frac{|\varphi_m(n+1)|^2}{|\varphi_m(n)|^2}. \quad (2.80)$$

Para empregar uma terminologia mais usual, este decaimento local do erro quadrático de aproximação será referido, daqui por diante, por "velocidade de convergência local". Esta grandeza, além de poder ser determinante no tempo de convergência do algoritmo quando o ponto inicial da adaptação estiver próximo do valor correto dos coeficientes, pode também ser determinante para a capacidade de rastreamento do mesmo, isto é, a capacidade do algoritmo de adaptação em, uma vez tendo convergido, acompanhar variações temporais do sistema sendo modelado.

No Item 2.7 a velocidade de convergência local (em dB/iteração) é comparada para alguns casos práticos com o valor obtido de (2.71).

## 2.3 Algoritmo do gradiente recursivo

Neste item apresenta-se inicialmente a obtenção do algoritmo do gradiente recursivo (RG) a partir de um algoritmo ideal de gradiente, para o caso da estrutura direta de implementação. A análise genérica da convergência para o caso de modelamento suficiente, feita no Item 2.2, é então aplicada ao algoritmo RG. Disto resulta que se a entrada é uma excitação persistente o valor correto dos parâmetros é um ponto de convergência local do algoritmo. Esta conclusão segue da validade local de um determinado critério de positividade. Além disso, a região de validade deste mesmo critério permite determinar uma região que está contida no domínio

de atração do ponto de convergência. Verifica-se que quando os polos do sistema sendo modelado estão próximos do círculo unitário e do eixo real o tamanho desta região tende a diminuir.

### 2.3.1 Obtenção do algoritmo

Consideremos inicialmente um algoritmo de gradiente ideal que utilizasse, para a adaptação, o gradiente do erro quadrático médio de aproximação  $E[\varepsilon^2(n, \mathbf{w})]$ , onde  $\varepsilon(n)$  definido em (1.6) é dado aqui por  $\varepsilon(n, \mathbf{w}) = [H(z) - \hat{H}(z)]u(n)$  (como estabelecido no Item 2.1, a inclusão de  $\mathbf{w}$  no argumento de  $\varepsilon$  indica que os parâmetros de  $\hat{H}(z)$  são considerados fixos). Isto é, para cada parâmetro  $w_i$  do filtro a adaptação seria dada por

$$w_i(n+1) = w_i(n) - \frac{\mu}{2} \frac{\partial E[\varepsilon^2(n, \mathbf{w}(n))]}{\partial w_i}, \quad i = 1, 2, \dots, 2M+1. \quad (2.81)$$

Para obter as derivadas na expressão acima escreve-se inicialmente, usando (1.19),

$$E[\varepsilon^2(n, \mathbf{w})] = \left\langle H(z) - \hat{H}(z), \mathcal{S}_u(z)[H(z) - \hat{H}(z)] \right\rangle, \quad (2.82)$$

onde  $\mathcal{S}_u(z)$  é a Transformada  $z$  da autocorrelação da entrada  $u(n)$ :

$$\mathcal{S}_u(z) = \sum_{k=-\infty}^{\infty} E[u(n)u(n+k)]z^k. \quad (2.83)$$

Usando (1.21) e notando que  $\mathcal{S}_u(z) = \mathcal{S}_u(z^{-1})$  temos então

$$\begin{aligned} \frac{\partial E[\varepsilon^2(n, \mathbf{w})]}{\partial w_i} &= \frac{\partial}{\partial w_i} \left\langle H(z) - \hat{H}(z), \mathcal{S}_u(z)[H(z) - \hat{H}(z)] \right\rangle \\ &= -2 \left\langle \frac{\partial \hat{H}(z)}{\partial w_i}, \mathcal{S}_u(z)[H(z) - \hat{H}(z)] \right\rangle. \end{aligned} \quad (2.84)$$

Para colocar esta expressão em uma forma que conduza a uma aproximação prática, consideremos a expressão do erro de saída

$$e(n, \mathbf{w}) = [H(z) - \hat{H}(z)]u(n) + \zeta(n), \quad (2.85)$$

e também sua derivada em relação a  $w_i$

$$\nabla_{w,i}^e(n, \mathbf{w}) = \frac{\partial e(n, \mathbf{w})}{\partial w_i} = -\frac{\partial \hat{H}(z)}{\partial w_i} u(n). \quad (2.86)$$

Notar que  $\nabla_{w,i}^e(n, \mathbf{w})$  é a saída no instante  $n$  de um filtro invariante no tempo com função de transferência  $-\frac{\partial}{\partial w_i} \hat{H}(z)$  e entrada  $u(n)$ . Usando (2.85) e (2.86) e aplicando uma operação de média temos então

$$\begin{aligned} E[\nabla_{w,i}^e(n, \mathbf{w}) e(n, \mathbf{w})] &= -E\left[\frac{\partial \hat{H}(z)}{\partial w_i} u(n) [H(z) - \hat{H}(z)] u(n)\right] - E\left[\frac{\partial \hat{H}(z)}{\partial w_i} u(n) \zeta(n)\right] \\ &= -\left\langle \frac{\partial \hat{H}(z)}{\partial w_i}, \mathcal{S}_u(z) [H(z) - \hat{H}(z)] \right\rangle - E\left[\frac{\partial \hat{H}(z)}{\partial w_i} u(n) \zeta(n)\right]. \end{aligned} \quad (2.87)$$

Considerando agora  $u(n)$  e  $\zeta(n)$  estatisticamente independentes e de média nula, temos que

$$E\left[\frac{\partial \hat{H}(z)}{\partial w_i} u(n) \zeta(n)\right] = 0. \quad (2.88)$$

Portanto, usando (2.84) e (2.87), temos

$$E[\nabla_{w,i}^e(n, \mathbf{w}) e(n, \mathbf{w})] = \frac{1}{2} \frac{\partial E[\varepsilon^2(n, \mathbf{w})]}{\partial w_i}. \quad (2.89)$$

Agrupando elementos em vetores, teríamos então que a adaptação dos parâmetros por um algoritmo de gradiente seria dada, idealmente, por

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu E[\nabla_{\mathbf{w}}^e(n, \mathbf{w}(n)) e(n, \mathbf{w}(n))]. \quad (2.90)$$

Como primeiro passo para uma aproximação prática de (2.90), abandona-se a operação de média obtendo

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \nabla_{\mathbf{w}}^e(n, \mathbf{w}(n)) e(n, \mathbf{w}(n)). \quad (2.91)$$

Esta forma ainda não seria viável em termos práticos, uma vez que o cálculo de  $\nabla_{\mathbf{w}}^e(n, \mathbf{w}(n))$  e  $e(n, \mathbf{w}(n))$  envolveria, a cada  $\mathbf{w}(n)$ , uma nova recursão desde  $n=0$ . Emprega-se então uma

aproximação, assumindo que  $\mathbf{w}(n)$  varia lentamente no tempo e, portanto,

$$e(n, \mathbf{w}(n)) \simeq [H(z) - \hat{H}(z, n)]u(n) + \zeta(n) = e(n), \quad (2.92)$$

e que é disponível um filtro variante no tempo  $D_{w,i}(z, n)$  tal que:

$$\nabla_{w,i}^e(n, \mathbf{w}(n)) \simeq D_{w,i}(z, n)u(n) \triangleq \nabla_{w,i}^e(n). \quad (2.93)$$

Chegamos então na seguinte expressão para o algoritmo do gradiente recursivo (RG):

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \nabla_{\mathbf{w}}^e(n) e(n). \quad (2.94)$$

Esta expressão é genérica no que diz respeito à implementação, faltando definir as expressões exatas dos elementos  $D_{w,i}(z, n)u(n)$  que compõe  $\nabla_{\mathbf{w}}^e(n)$ . Para a estrutura de implementação direta, com a introdução de mais uma aproximação baseada na variação lenta de  $\mathbf{w}(n)$ , estes elementos tem uma expressão simples, como visto a seguir. Para outras estruturas de implementação estas expressões podem ser bastante complexas, sendo necessário recorrer a mais aproximações. Para estruturas do tipo lattice há uma extensa análise deste aspecto em [1].

### 2.3.2 Aplicação à estrutura direta de implementação

No caso da estrutura direta os parâmetros  $w_i$  são os próprios coeficientes do numerador e do denominador de  $\hat{H}(z)$ , que, para manter a coerência com (2.20), são agrupados como:

$$\mathbf{w} = \left[ w_1 \quad w_2 \quad \cdots \quad w_{2M+1} \right]^T \quad (2.95)$$

$$= \left[ a_1 \quad a_2 \quad \cdots \quad a_M \quad b_0 \quad b_1 \quad \cdots \quad b_M \right]^T \quad (2.96)$$

As derivadas relativas aos coeficientes  $\{a_i\}$  são dadas por

$$\begin{aligned} \frac{\partial \hat{H}(z)}{\partial a_i} &= \frac{\partial B(z)}{\partial a_i A(z)} \\ &= -\frac{B(z)}{A^2(z)} \frac{\partial}{\partial a_i} A(z) \end{aligned}$$

$$= -\frac{z^i}{A(z)}\widehat{H}(z), \quad i = 1, 2, \dots, M, \quad (2.97)$$

e as derivadas relativas aos coeficientes  $\{b_i\}$ , por

$$\begin{aligned} \frac{\partial \widehat{H}(z)}{\partial b_i} &= \frac{\partial B(z)}{\partial b_i A(z)} \\ &= \frac{z^i}{A(z)}, \quad i = 0, 1, \dots, M. \end{aligned} \quad (2.98)$$

Usando (2.93), teríamos então

$$\nabla_w^e(n) = \begin{bmatrix} \frac{z}{A(z, n)}\widehat{H}(z, n)u(n) \\ \frac{z^2}{A(z, n)}\widehat{H}(z, n)u(n) \\ \vdots \\ \frac{z^M}{A(z, n)}\widehat{H}(z, n)u(n) \\ -\frac{1}{A(z, n)}u(n) \\ -\frac{z}{A(z, n)}u(n) \\ \vdots \\ -\frac{z^M}{A(z, n)}u(n) \end{bmatrix} = \begin{bmatrix} \frac{z}{A(z, n)}\widehat{y}(n) \\ \frac{z^2}{A(z, n)}\widehat{y}(n) \\ \vdots \\ \frac{z^M}{A(z, n)}\widehat{y}(n) \\ -\frac{1}{A(z, n)}u(n) \\ -\frac{z}{A(z, n)}u(n) \\ \vdots \\ -\frac{z^M}{A(z, n)}u(n) \end{bmatrix}. \quad (2.99)$$

Se, a partir da expressão acima, forem escritas explicitamente as equações de diferenças para cada elemento de  $\nabla_w^e(n)$ , pode-se ver que é necessário, a rigor, um filtro variante no tempo para cada um dos  $2M + 1$  elementos. Para parâmetros variando lentamente no tempo, porém, pode-se ver também que  $\nabla_{w,i}^e(n) \simeq \nabla_{w,i-1}^e(n-1)$ , para  $i = 2, 3, \dots, M$  ou  $i = M + 2, M + 3, \dots, 2M + 1$ , e, portanto, pode-se usar apenas um filtro variante no tempo para obter as aproximações das derivadas relativas aos coeficientes do numerador e outro

para as derivadas relativas aos coeficientes do denominador, resultando em

$$\nabla_w^e(n) \simeq \begin{bmatrix} \nabla_a(n-1) \\ \nabla_a(n-2) \\ \vdots \\ \nabla_a(n-M) \\ -\nabla_b(n) \\ -\nabla_b(n-1) \\ \vdots \\ -\nabla_b(n-M) \end{bmatrix} = -\boldsymbol{\psi}(n), \quad (2.100)$$

onde

$$\nabla_a(n) = \frac{\hat{y}(n)}{A(z, n)} \quad \text{e} \quad \nabla_b(n) = \frac{u(n)}{A(z, n)}. \quad (2.101)$$

As aproximações em (2.93), (2.92) e (2.100) serão tão melhores quanto mais lenta for a variação dos parâmetros em relação à memória de  $\hat{H}(z, n)$ . Ou seja, se o sistema  $H(z)$  tiver polos próximos à circunferência unitária será necessário um valor pequeno de  $\mu$  se quisermos garantir uma adaptação dada de fato pelo gradiente do erro quadrático.

Chegamos então a uma expressão de adaptação na forma de (2.7), isto é  $\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \boldsymbol{\psi}(n) \epsilon(n)$ , onde  $\epsilon(n) = e(n)$  e  $\boldsymbol{\psi}(n)$  é dado por (2.100). Explicitando a adaptação dos coeficientes do numerador e do denominador temos:

$$a_i(n+1) = a_i(n) - \mu \nabla_a(n-i) e(n), \quad i = 1, 2, \dots, M,$$

$$b_i(n+1) = b_i(n) + \mu \nabla_b(n-i) e(n), \quad i = 0, 1, \dots, M. \quad (2.102)$$

### 2.3.3 Condição suficiente para convergência

Vejam agora uma condição suficiente para a convergência local do algoritmo RG aos parâmetros  $\mathbf{w}^*$  do sistema sendo modelado  $H(z)$ , onde estamos considerando que  $H(z)$  é racional e  $\deg[H(z)] = M$ .

Se as sequências de entrada  $\{u(n)\}$  e do ruído de medida  $\{\zeta(n)\}$  forem estatisticamente independentes, então os elementos de  $\boldsymbol{\psi}(n, \mathbf{w})$  dados por (2.100) (fixando os parâmetros em  $\mathbf{w}$ ) também são estatisticamente independentes da sequência  $\{\zeta(n)\}$ . Supondo  $\zeta(n)$  de

média nula temos então:

$$\begin{aligned} E[\boldsymbol{\psi}(n, \mathbf{w}^*) \zeta(n)] &= E[\boldsymbol{\psi}(n, \mathbf{w}^*)] E[\zeta(n)] \\ &= \mathbf{0}, \end{aligned}$$

e a primeira condição da Propriedade 2.1 é atendida.

Para verificarmos a segunda condição da propriedade 2.1, consideremos a matriz  $\mathbf{R}'$  dada por (2.29). Queremos então que, para qualquer vetor  $\mathbf{e}$ ,

$$\mathbf{e}^\top \mathbf{R}'(\mathbf{w}) \mathbf{e} = E[\mathbf{e}^\top \boldsymbol{\psi}(n, \mathbf{w}) \boldsymbol{\phi}_2^\top(n, \mathbf{w}) \mathbf{e}] > 0, \quad (2.103)$$

onde  $\boldsymbol{\phi}_2^\top(n, \mathbf{w})$  é dado por (2.24) com  $P(z) \equiv 1$ . Escrevendo, de maneira semelhante à usada no Item 1.5.4,

$$\mathbf{e} = \begin{bmatrix} r_0 & r_1 & \cdots & r_{M-1} & q_0 & q_1 & \cdots & q_M \end{bmatrix}^\top, \quad (2.104)$$

$R(z) = r_0 + r_1 z + \dots + r_{M-1} z^{M-1}$ ,  $Q(z) = q_0 + q_1 z + \dots + q_M z^M$ , e

$$v(n) = [Q(z) - zR(z) \frac{B(z)}{A(z)}] u(n), \quad (2.105)$$

e usando (1.19), podemos escrever

$$\begin{aligned} \mathbf{e}^\top \mathbf{R}'(\mathbf{w}) \mathbf{e} &= E[\mathbf{e}^\top \boldsymbol{\psi}(n, \mathbf{w}) \boldsymbol{\phi}_2^\top(n, \mathbf{w}) \mathbf{e}] \\ &= E \left[ \frac{v(n)}{A(z)} \frac{v(n)}{D(z)} \right] \\ &= \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{A(e^{j\omega})} \frac{1}{D(e^{-j\omega})} \mathcal{S}_v(e^{j\omega}) d\omega. \end{aligned} \quad (2.106)$$

Escrevendo agora

$$\frac{1}{2\pi} \int_0^{2\pi} \frac{1}{A(e^{j\omega})} \frac{1}{D(e^{-j\omega})} \mathcal{S}_v(e^{j\omega}) d\omega = \frac{1}{2\pi} \int_0^{2\pi} \frac{A(e^{-j\omega})}{D(e^{-j\omega})} \left| \frac{1}{A(e^{j\omega})} \right|^2 \mathcal{S}_v(e^{j\omega}) d\omega$$



$$= \frac{1}{2\pi} \int_0^{2\pi} \operatorname{Re} \left[ \frac{A(e^{j\omega})}{D(e^{j\omega})} \right] \left| \frac{1}{A(e^{j\omega})} \right|^2 \mathcal{S}_v(e^{j\omega}) d\omega, \quad (2.107)$$

segue a propriedade:

**Propriedade 2.5** *Uma condição suficiente para que, no algoritmo RG,  $\mathbf{R}'(\mathbf{w})$  seja positiva definida é que*

$$\operatorname{Re} \left[ \frac{A(e^{j\omega})}{D(e^{j\omega})} \right] > 0, \quad -\pi < \omega \leq \pi, \quad (2.108)$$

e também que

$$\mathcal{S}_v(e^{j\omega}) > 0. \quad (2.109)$$

Notar que se existirem  $\omega_1$  e  $\omega_2$  tais que

$$\operatorname{Re} \left[ \frac{A(e^{j\omega})}{D(e^{j\omega})} \right] < 0, \quad \omega_1 < |\omega| < \omega_2, \quad (2.110)$$

poderá existir  $\mathcal{S}_u(e^{j\omega})$  tal que  $\mathbf{R}'(\mathbf{w})$  não seja positiva definida. Neste sentido, a condição (2.108) não é conservadora em relação à condição  $\mathbf{R}'(\mathbf{w}) > \mathbf{0}$ .

Com a propriedade acima podemos então verificar o atendimento da segunda condição de (2.1). Uma vez que, quando  $\mathbf{w} = \mathbf{w}^*$  temos  $A(z) = D(z)$ , sempre existirá uma vizinhança de  $\mathbf{w}^*$  em que a condição (2.108) é obedecida. Consideremos agora a condição (2.109), que só não será atendida se  $v(n) \equiv 0$ . Notar que em (2.105)  $v(n)$  tem a mesma forma que  $e^{\top} \mathbf{r}(n)$  em (1.39). Pela Propriedade 1.3, portanto, se  $u(n)$  é uma excitação persistente de ordem  $2M+1$  ou maior e  $B(z)$  e  $A(z)$  não tiverem alguma raiz comum então  $v(n)$  não é identicamente nulo. Mas, como foi assumido que  $\deg[H(z)] = M$ , sempre existirá uma vizinhança de  $\mathbf{w}^*$  na qual  $B(z)$  e  $A(z)$  não tem uma raiz comum.

Em função da análise precedente, concluímos então que a seguinte propriedade é válida:

**Propriedade 2.6** *Para o algoritmo RG, dado por (2.102), se as sequências de entrada  $\{u(n)\}$  e do ruído de medida  $\{\zeta(n)\}$  forem estatisticamente independentes,  $u(n)$  for uma excitação persistente de ordem  $2M+1$  ou maior,  $\zeta(n)$  tiver média nula e o passo de adaptação  $\mu$  for suficientemente pequeno, então o vetor de parâmetros  $\mathbf{w}^*$  do sistema sendo modelado  $H(z)$ , onde  $\deg[H(z)] = M$ , é um ponto de convergência local.*

### 2.3.4 Sobremodelamento

No caso em que, diferentemente do que foi assumido acima, tivéssemos  $\deg[H(z)] < M$ , o filtro adaptativo teria mais parâmetros do que os necessários para modelar  $H(z)$ , situação que recebe a denominação de *sobremodelamento*. Neste caso, conforme pode ser visto acima, no ponto  $\mathbf{w} = \mathbf{w}^*$ , isto é, quando  $B(z) = C(z)$  e  $A(z) = D(z)$ , é possível tornar  $v(n)$  identicamente nulo. Conseqüentemente, a matriz  $\mathbf{R}(\mathbf{w}^*) = \mathbf{R}'(\mathbf{w}^*)$  é singular, não existindo, conforme exigido pela Propriedade 2.1, uma vizinhança de  $\mathbf{w}^*$  onde  $\mathbf{R}'(\mathbf{w})$  seja positiva definida. Porém, esta situação não implica, necessariamente, em problemas de convergência para o algoritmo de adaptação. Existem, neste caso, infinitos pontos no espaço de parâmetros onde  $\hat{H}(z) = H(z)$  e o algoritmo pode convergir de maneira rápida para algum deles. Uma situação mais problemática ocorre quando há um "quase-sobremodelamento", isto é, quando  $H(z)$  tem algum polo muito próximo de um zero. Neste caso,  $\mathbf{R}(\mathbf{w}^*)$  não é singular mas é mal condicionada, e, mesmo que, como na situação de sobremodelamento, o algoritmo convirja rapidamente para as proximidades de  $\mathbf{w}^*$ , a partir de um determinado momento a convergência será lenta, com o erro de saída permanecendo um longo tempo em um valor possivelmente insatisfatório.

Uma análise mais detalhada do condicionamento de  $\mathbf{R}(\mathbf{w}^*)$  nestas situações de "quase-sobremodelamento" não será realizada neste trabalho. Trata-se de uma questão importante, mas preferiu-se abordar o aspecto do efeito da proximidade à circunferência unitária dos polos de  $H(z)$  no condicionamento de  $\mathbf{R}(\mathbf{w}^*)$  (analisado no Item 2.6), que é mais crucial na determinação do interesse prático por filtros IIR adaptativos, conforme discutido no Capítulo 1.

Por fim, estas considerações a respeito de sobremodelamento se aplicam também aos outros algoritmos considerados adiante.

### 2.3.5 Influência da posição dos polos sobre a região de positividade e o domínio de atração

Vejamos agora como a análise anterior permite caracterizar parcialmente o domínio de atração de  $\mathbf{w}^*$ . Notar inicialmente que, reunindo (2.100), (2.89) e valendo a aproximação da adaptação dos parâmetros pela equação diferencial (2.8), resulta que o algoritmo RG também poderá convergir para um eventual ponto de mínimo local da superfície de erro. Na

situação de modelamento suficiente sendo considerada aqui, isto é,  $\deg[H(z)] = M$ , conforme será visto no Item 3.2, uma condição necessária para que existam pontos de mínimo local é que a entrada  $u(n)$  não seja branca.

O fato de poder convergir para mínimos locais e não para o mínimo global  $\mathbf{w}^*$  da superfície de erro é uma deficiência importante do algoritmo RG. É de interesse então analisar condições que garantam a convergência ao mínimo global mesmo na existência de mínimos locais. Para tanto, temos, inicialmente, que para garantir o atendimento de (2.109) é necessário que  $B(z)$  e  $A(z)$  não tenham raízes comuns. Como visto no item anterior, se isto não for atendido, não é possível, com a função de Lyapunov escolhida, garantir a convergência a  $\mathbf{w}^*$  da equação diferencial (2.8). Ao passar para o algoritmo de adaptação, porém, esta condição perde sua importância: uma pequena variação dos parâmetros, resultado da própria natureza estocástica do algoritmo, é suficiente para que  $B(z)$  e  $A(z)$  voltem a não ter raízes comuns e que portanto a convergência em direção a  $\mathbf{w}^*$  prossiga [1]. Com base nisto e nas Propriedades 2.3 e 2.1, a seguinte propriedade é então válida:

**Propriedade 2.7** *Seja  $B_r$  a bola de maior raio contida no conjunto  $W_{pos}$  para o qual (2.108) é válida. Para o algoritmo RG, então, nas condições da propriedade 2.6,  $B_r$  está contido no domínio de atração de  $\mathbf{w}^*$ .*

Em aplicações práticas de filtros IIR adaptativos, frequentemente dispõe-se de alguma estimativa *a priori* sobre o sistema sendo modelado. Se esta estimativa for tal que o valor inicial  $\mathbf{w}(0)$  da adaptação esteja contido em  $B_r$ , então, para o algoritmo RG, a convergência a  $\mathbf{w}^*$  estará garantida, a despeito da existência de eventuais mínimos locais. Em razão disto é de interesse verificar como  $B_r$  varia com a posição dos polos de  $H(z)$ . Este aspecto é abordado a seguir.

Para maior facilidade de representação gráfica, os polinômios serão considerados aqui em potências de  $z^{-1}$ . Se  $D(z) = 1 + d_1z + \dots + d_Mz^M$  tem zeros fora do círculo unitário, então

$$\begin{aligned} D(z^{-1}) &= 1 + d_1z^{-1} + \dots + d_Mz^{-M} \\ &= \frac{z^M + d_1z^{M-1} + \dots + d_M}{z^M}, \end{aligned} \quad (2.111)$$

tem zeros dentro do círculo unitário e  $M$  polos em  $z = 0$ .

Consideremos quatro polinômios  $D(z^{-1})$  de segunda ordem, com raízes dadas por  $0,7\angle \pm 60^\circ$ ,  $0,9\angle \pm 60^\circ$ ,  $0,7\angle \pm 15^\circ$ ,  $0,9\angle \pm 15^\circ$ . Na Figura 2-1 está indicada, para cada um destes

casos, a região das raízes conjugadas de  $A(z^{-1})$  para a qual a condição (2.108) é obedecida e a região das raízes conjugadas de  $A(z^{-1})$  para a qual  $\mathbf{w} \in B_r$ . Podemos observar que se os polos do sistema sendo modelado estão próximos ao círculo unitário é necessário uma estimativa inicial de  $D(z)$  muito boa para poder-se garantir a convergência ao valor correto dos parâmetros.

## 2.4 Algoritmo Steiglitz-McBride

Neste item apresenta-se, inicialmente, uma interpretação simples da relação entre a versão recursiva do método Steiglitz-McBride (SMM) e sua formulação *off-line*, considerando a estrutura direta de implementação. Aplica-se então ao algoritmo SMM a análise genérica da convergência para o caso de modelamento suficiente, feita no Item 2.2. Disto resulta que se o ruído de medida é branco e a entrada é uma excitação persistente, o valor correto dos parâmetros é um ponto de convergência global do algoritmo. Se, por um lado, a restrição ao ruído de medida pode representar uma limitação importante, a garantia de convergência local pode representar uma grande vantagem.

### 2.4.1 Obtenção do algoritmo recursivo

No método Steiglitz-McBride de identificação *off-line* de sistemas [9] tem-se uma sequência de problemas de otimização lineares. Neste método, conforme representado na figura 2-2, na iteração de índice  $i$ , determinam-se as funções  $A_{i+1}(z)$  e  $B_{i+1}(z)$  que minimizam  $E[e_i^2(n)]$ , com  $e_i(n)$  dado por

$$e_i(n) = \frac{A_{i+1}(z)}{A_i(z)} y(n) - \frac{B_{i+1}(z)}{A_i(z)} u(n), \quad (2.112)$$

onde  $A_i(z)$  é a função obtida da minimização de  $E[e_{i-1}^2(n)]$ . Se a sequência de funções  $A_i(z)$  convergir para um ponto estacionário  $A_\infty(z)$  o erro resultante será

$$e_\infty(n) = y(n) - \frac{B_\infty(z)}{A_\infty(z)} u(n), \quad (2.113)$$

o que corresponde ao erro de saída  $e(n) = [H(z) - \hat{H}(z)]u(n) + \zeta(n)$ , com  $\hat{H}(z) = B_\infty(z)/A_\infty(z)$ . Coloca-se então um problema de convergência: naturalmente, deseja-se que exista um ponto estacionário da iteração e que, além disso, o mesmo corresponda ao  $\hat{H}(z)$  que minimiza o erro quadrático médio de saída  $E[e^2(n)]$ . A convergência do algoritmo Steiglitz-McBride *off-line*

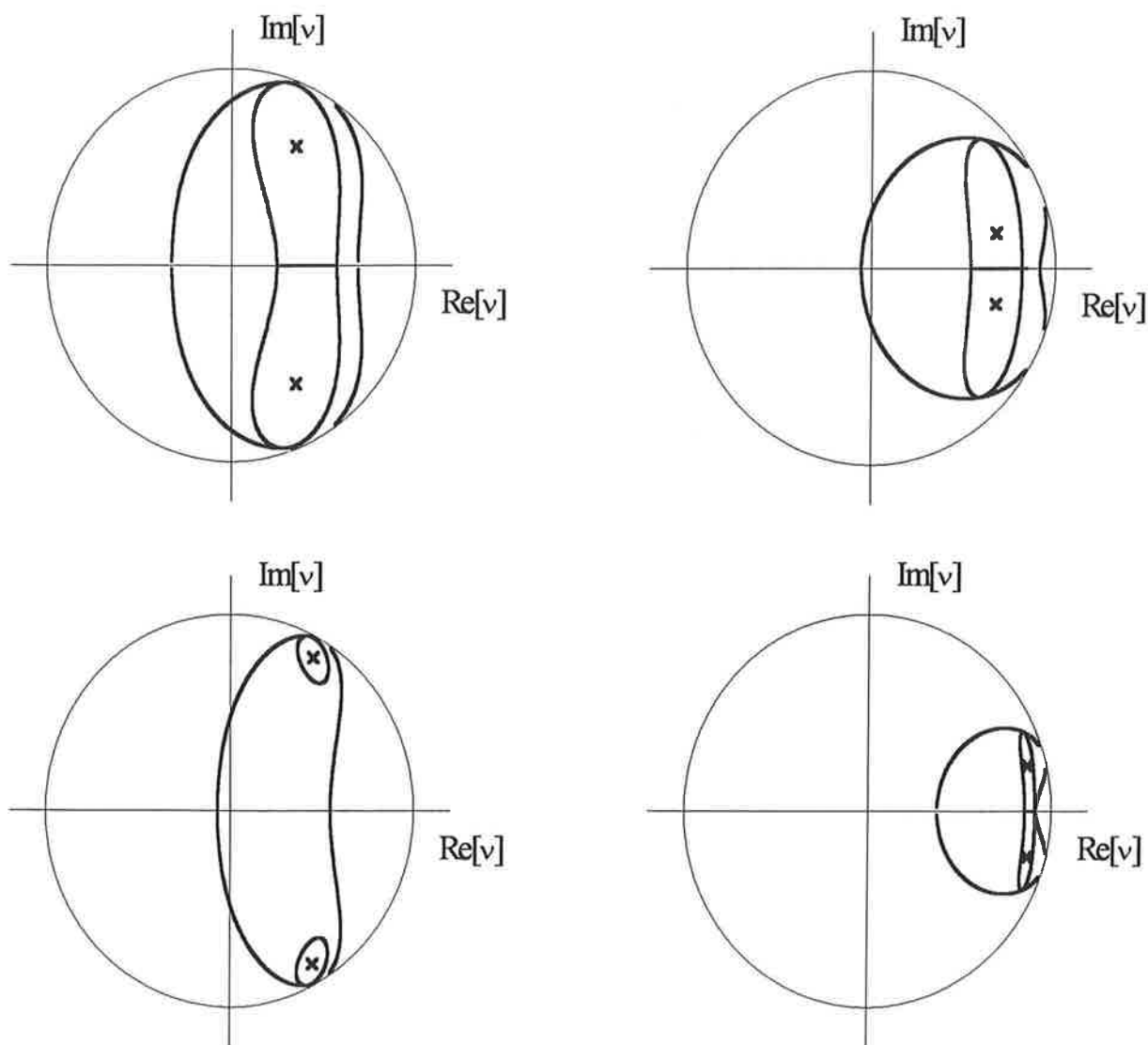


Figura 2-1: Algoritmo RG, sistemas de segunda ordem (polos de  $H(z^{-1})$  indicados com  $\times$ ): região das raízes de  $A(z^{-1})$  na qual a condição de positividade é obedecida e a imagem da maior bola no espaço de parâmetros contida nesta região; a circunferência representada é a circunferência de raio unitário.

é analisada em [10] e [1]. Nos limitaremos aqui a apresentar uma análise de convergência para a versão recursiva deste algoritmo, obtida a seguir.

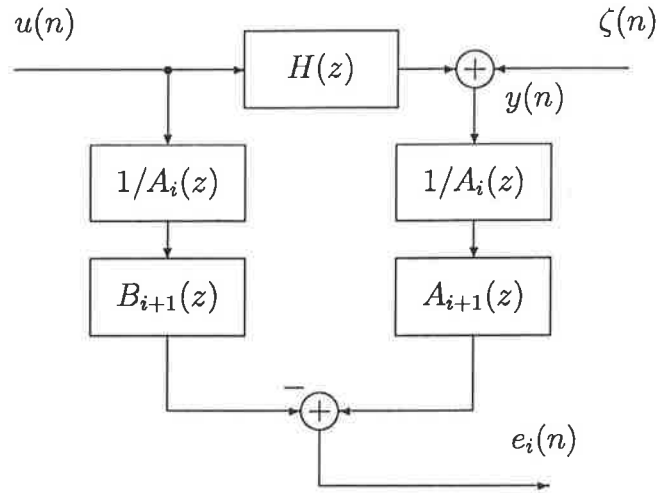


Figura 2-2: método Steiglitz-McBride de identificação de sistemas

Para obtermos uma versão recursiva do algoritmo *off-line*, consideremos inicialmente que, ao invés de, a cada iteração  $i$ , escolher  $A_{i+1}(z)$  e  $B_{i+1}(z)$  tal que  $E[e_i^2(n)]$  seja minimizado, escolhamos  $A_{i+1}(z)$  e  $B_{i+1}(z)$  a partir de  $A_i(z)$  e  $B_i(z)$  e do gradiente de  $e_i^2(n)$  no instante  $n = i$ . A princípio, não há nada que impeça a convergência à solução desejada deste procedimento, e podemos, de fato, como visto a seguir, interpretar a passagem da versão *off-line* do algoritmo para a versão recursiva como uma mudança no método de escolha de  $A_{i+1}(z)$  e  $B_{i+1}(z)$ . Teríamos então, supondo a forma direta de implementação, e usando (2.112),

$$\begin{aligned} a_{j,i+1} &= a_{j,i} - \mu \left[ \frac{\partial}{\partial a_{j,i+1}} \frac{A_{i+1}(z)}{A_i(z)} \right] y(n) e_i(n) \\ &= a_{j,i} - \mu \frac{z^j}{A_i(z)} y(n) e_i(n), \quad j = 1, 2, \dots, M, \quad n = i, \end{aligned} \quad (2.114)$$

e

$$b_{j,i+1} = b_{j,i} + \mu \left[ \frac{\partial}{\partial b_{j,i+1}} \frac{B_{i+1}(z)}{A_i(z)} \right] u(n) e_i(n)$$

$$= b_{j,i} + \mu \frac{z^j}{A_i(z)} u(n) e_i(n), \quad j = 0, 1, \dots, M, \quad n = i, \quad (2.115)$$

onde para maior clareza na discussão a seguir, os símbolos  $n$  e  $i$  são usados apesar de  $n = i$ . Notar que, nestas expressões, o cálculo de  $e_i(n)$ ,  $[z^j/A_i(z)]y(n)$  e  $[z^j/A_i(z)]u(n)$ ,  $n = i$ , exige, a cada iteração  $i$ , novas recursões de  $n = 0$  até  $n = i$ . No entanto, supondo que  $A_i(z)$  está variando lentamente a cada iteração, pode-se dispensar esta recursão desde  $n = 0$ , e escrever

$$\frac{z^j}{A_i(z)} y(n) \simeq \nabla_{a,i}(n-j), \quad (2.116)$$

onde  $\nabla_{a,i}(n) = y(n)/A_i(z) = y(n) - \sum_{l=1}^M a_{l,i} \nabla_{a,i}(n-l)$ , e, da mesma maneira,

$$\frac{z^j}{A_i(z)} u(n) \simeq \nabla_{b,i}(n-j), \quad (2.117)$$

onde  $\nabla_{b,i}(n) = u(n)/A_i(z) = u(n) - \sum_{l=1}^M a_{l,i} \nabla_{b,i}(n-l)$ . Para o sinal de erro também podemos fazer a aproximação

$$\begin{aligned} e_i(n) &= \frac{A_{i+1}(z)}{A_i(z)} y(n) - \frac{B_{i+1}(z)}{A_i(z)} u(n) \\ &\simeq y(n) - \frac{B_i(z)}{A_i(z)} u(n). \end{aligned} \quad (2.118)$$

Com estas aproximações e substituindo agora  $i$  por  $n$ , as expressões (2.114) e (2.115) podem ser re-escritas como

$$\begin{aligned} a_j(n+1) &= a_j(n) - \mu \nabla_a(n-j) e(n), \quad j = 1, 2, \dots, M, \\ b_j(n+1) &= b_j(n) + \mu \nabla_b(n-j) e(n), \quad j = 0, 1, \dots, M, \end{aligned} \quad (2.119)$$

onde agora, usando a notação para filtros variantes no tempo definida no Item 2.1, temos

$$\nabla_a(n) = \frac{y(n)}{A(z, n)} \quad \text{e} \quad \nabla_b(n) = \frac{u(n)}{A(z, n)}. \quad (2.120)$$

As expressões acima são as expressões finais do algoritmo Steiglitz-McBride para filtros IIR

adaptativos implementados na forma direta. Notar que a única diferença em relação ao algoritmo RG (expressões (2.102)) é a utilização de  $y(n)$  no lugar de  $\hat{y}(n)$ . Isto porém terá consequências importantes no comportamento do algoritmo, como visto a seguir.

## 2.4.2 Condição para convergência

Consideremos novamente as condições para que os parâmetros  $\mathbf{w}^*$  do sistema  $H(z)$ , onde  $\deg[H(z)] = M$ , sejam um ponto de convergência de um algoritmo de passo constante, expressas na Propriedade 2.1. Para o algoritmo SMM colocado na forma genérica (2.7) com  $P(z) \equiv 1$ , a primeira condição da Propriedade 2.1 é

$$E[\boldsymbol{\psi}(n, \mathbf{w}^*) \zeta(n)] = 0. \quad (2.121)$$

onde, pelas expressões (2.119)

$$\boldsymbol{\psi}(n, \mathbf{w}^*) = \begin{bmatrix} -y(n-1)/A(z) \\ -y(n-2)/A(z) \\ \vdots \\ -y(n-M)/A(z) \\ u(n)/A(z) \\ u(n-1)/A(z) \\ \vdots \\ u(n-M)/A(z) \end{bmatrix}. \quad (2.122)$$

Separando (2.121) quanto aos dois tipos de elemento de  $\boldsymbol{\psi}(n, \mathbf{w}^*)$ , temos:

$$E \left[ \frac{u(n-i)}{A(z)} \zeta(n) \right] = 0, \quad i = 1, 2, \dots, M, \quad (2.123)$$

e, usando  $y(n) = H(z)u(n) + \zeta(n) = y'(n) + \zeta(n)$ ,

$$\begin{aligned} E \left[ \frac{y(n-i)}{A(z)} \zeta(n) \right] &= E \left[ \frac{y'(n-i)}{A(z)} \zeta(n) \right] + E \left[ \frac{\zeta(n-i)}{A(z)} \zeta(n) \right] \\ &= 0, \quad i = 1, 2, \dots, M. \end{aligned} \quad (2.124)$$



Se as sequências de entrada  $\{u(n)\}$  e do ruído de medida  $\{\zeta(n)\}$  forem estatisticamente independentes e  $\zeta(n)$  for branco e de média nula, as duas equações acima são atendidas. Porém, se o ruído de medida não for branco, o valor correto dos parâmetros  $\mathbf{w}^*$  poderá não ser um ponto de convergência do algoritmo SMM. Esta é uma restrição importante, que pode impedir a utilização deste algoritmo em diversas aplicações práticas. Há, no entanto, uma questão de compromisso: o algoritmo RG, por exemplo, não coloca esta restrição ao ruído de medida; por outro lado, se a restrição for obedecida, o algoritmo SMM apresenta melhores características de convergência globais do que o algoritmo RG, como será visto a seguir.

Consideremos agora a segunda condição da Propriedade 2.1, analisando a forma da matriz  $\mathbf{R}(\mathbf{w})$  dada por (2.26). Os elementos de  $\mathbf{R}(\mathbf{w})$  tem a forma

$$\begin{aligned} E \left[ \frac{y(n-i)}{A(z)} \frac{y'(n-j)}{A(z)} \right] &= E \left[ \frac{y'(n-i)}{A(z)} \frac{y'(n-j)}{A(z)} \right] + \underbrace{E \left[ \frac{\zeta(n-i)}{A(z)} \frac{y'(n-j)}{A(z)} \right]}_0 \\ &= E \left[ \frac{y'(n-i)}{A(z)} \frac{y'(n-j)}{A(z)} \right], \end{aligned} \quad (2.125)$$

e

$$\begin{aligned} E \left[ \frac{y(n-i)}{A(z)} \frac{u(n-j)}{A(z)} \right] &= E \left[ \frac{y'(n-i)}{A(z)} \frac{u(n-j)}{A(z)} \right] + \underbrace{E \left[ \frac{\zeta(n-i)}{A(z)} \frac{u(n-j)}{A(z)} \right]}_0 \\ &= E \left[ \frac{y'(n-i)}{A(z)} \frac{u(n-j)}{A(z)} \right], \end{aligned} \quad (2.126)$$

onde os termos nulos resultam do pressuposto de que  $\{u(n)\}$  e  $\{\zeta(n)\}$  são estatisticamente independentes e  $\zeta(n)$  tem média nula. Temos portanto que o ruído de medida não afeta a matriz  $\mathbf{R}(\mathbf{w})$  que pode ser escrita neste caso como

$$\mathbf{R}(\mathbf{w}) = E[\boldsymbol{\phi}_1(n, \mathbf{w})\boldsymbol{\phi}_1^\top(n, \mathbf{w})], \quad (2.127)$$

com  $\boldsymbol{\phi}_1(n, \mathbf{w})$  dado por (2.19) com  $P(z) \equiv 1$ . Podemos escrever, portanto,

$$\mathbf{e}^\top \mathbf{R}(\mathbf{w}) \mathbf{e} = E[\mathbf{e}^\top \boldsymbol{\phi}_1(n, \mathbf{w})\boldsymbol{\phi}_1^\top(n, \mathbf{w}) \mathbf{e}] \quad (2.128)$$

$$= E \left[ \frac{v(n)}{A(z)} \frac{v(n)}{A(z)} \right], \quad (2.129)$$

onde, analogamente a (2.105),

$$v(n) = [Q(z) - zR(z)H(z)]u(n). \quad (2.130)$$

Se a entrada  $u(n)$  for uma excitação persistente de ordem  $2M + 1$  ou maior e  $\deg[H(z)] = M$ , então, pela Propriedade 1.3,  $E[v^2(n)] > 0$  e  $\mathbf{R}(\mathbf{w})$  é positiva definida. Notar que neste caso, para valer a positividade de  $\mathbf{R}(\mathbf{w})$ , não há nenhuma restrição aos valores de  $\mathbf{w}$ , como a dada pela expressão (2.108) para o caso do algoritmo RG. O domínio de atração de  $\mathbf{w}^*$  consiste portanto em todo o domínio de estabilidade, ou seja, há convergência global para  $\mathbf{w}^*$ . Esta análise permite enunciar a seguinte propriedade:

**Propriedade 2.8** *Para o algoritmo SMM, dado por (2.119), se as sequências de entrada  $\{u(n)\}$  e do ruído de medida  $\{\zeta(n)\}$  forem estatisticamente independentes,  $u(n)$  for uma excitação persistente de ordem  $2M + 1$  ou maior,  $\zeta(n)$  for branco e de média nula e o passo de adaptação  $\mu$  for suficientemente pequeno, então o vetor de parâmetros  $\mathbf{w}^*$  do sistema sendo modelado  $H(z)$ , onde  $\deg[H(z)] = M$ , é um ponto de convergência global.*

A propriedade de convergência global do algoritmo SMM pode representar uma grande vantagem em relação ao algoritmo RG, uma vez que a existência de mínimos locais da superfície de erro passa a ser irrelevante.

## 2.5 Algoritmos PLR e SHARF

### 2.5.1 Expressões de adaptação e condição para convergência do algoritmo PLR

Consideremos o vetor usado como uma aproximação do gradiente no algoritmo RG, dado por (2.102) e (2.101). Uma aproximação adicional do gradiente pode ser feita eliminando a filtragem por  $1/A(z, n)$ , obtendo então as expressões de adaptação do algoritmo denominado *regressão pseudo-linear* (PLR):

$$a_i(n+1) = a_i(n) - \mu \hat{y}(n-i)e(n), \quad i = 1, 2, \dots, M,$$

$$b_i(n+1) = b_i(n) + \mu u(n-i)e(n), \quad i = 0, 1, \dots, M. \quad (2.131)$$

Para o algoritmo PLR, o vetor  $\psi(n)$  na forma genérica (2.7) assume, portanto, a forma

$$\psi(n) = \begin{bmatrix} -\hat{y}(n-1) \\ -\hat{y}(n-2) \\ \vdots \\ -\hat{y}(n-M) \\ u(n) \\ u(n-1) \\ \vdots \\ u(n-M) \end{bmatrix}. \quad (2.132)$$

De imediato pode-se dizer que a eliminação da filtragem por  $1/A(z, n)$  produzirá um erro de  $\psi(n)$  em relação ao gradiente que será, de modo geral, tanto pior quanto mais próximos estiverem do círculo unitário os polos de  $1/A(z, n)$ . Uma formulação mais precisa deste aspecto resulta da análise de convergência a seguir.

Consideremos a convergência do algoritmo PLR aos parâmetros  $\mathbf{w}^*$  do sistema  $H(z)$ , onde  $\deg[H(z)] = M$ . Temos, inicialmente, que, da mesma maneira que para o algoritmo RG, a primeira condição da Propriedade 2.1 é satisfeita se as sequências de entrada  $\{u(n)\}$  e do ruído de medida  $\{\zeta(n)\}$  forem estatisticamente independentes e  $\zeta(n)$  tiver média nula. Quanto à segunda condição, podemos escrever

$$\begin{aligned} \mathbf{e}^\top \mathbf{R}'(\mathbf{w}) \mathbf{e} &= E[\mathbf{e}^\top \psi(n, \mathbf{w}) \phi_2^\top(n, \mathbf{w}) \mathbf{e}] \\ &= E \left[ v(n) \frac{v(n)}{D(z)} \right] \\ &= \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{D(e^{j\omega})} \mathcal{S}_v(e^{j\omega}) d\omega \\ &= \frac{1}{2\pi} \int_0^{2\pi} \operatorname{Re} \left[ \frac{1}{D(e^{j\omega})} \right] \mathcal{S}_v(e^{j\omega}) d\omega \end{aligned} \quad (2.133)$$

onde, analogamente a (2.105),

$$v(n) = [Q(z) - zR(z)\frac{B(z)}{A(z)}]u(n). \quad (2.134)$$

Por uma argumentação análoga à utilizada no caso do algoritmo RG temos então que uma condição suficiente para que  $\mathbf{R}'(\mathbf{w})$  seja positiva definida é que  $u(n)$  seja uma excitação persistente de ordem  $2M + 1$  ou maior, que  $B(z)/A(z)$  não tenham raízes comuns e que

$$\operatorname{Re} \left[ \frac{1}{D(e^{j\omega})} \right] > 0, \quad -\pi \leq \omega < \pi. \quad (2.135)$$

Notar que, assim como para o algoritmo SMM, esta expressão não coloca nenhuma restrição aos valores de  $\mathbf{w}$ . Por outro lado, notando que em  $\mathbf{w}^*$  temos  $\hat{y}(n) = y(n)$ , os elementos da diagonal principal de  $\mathbf{R}'(\mathbf{w}^*)$ , são dados por

$$\begin{aligned} E \left[ y(n-i) \frac{y(n-i)}{D(z)} \right] &= E \left[ H(z)u(n-i) \frac{H(z)u(n-i)}{D(z)} \right] \\ &= \frac{1}{2\pi} \int_0^{2\pi} \operatorname{Re} \left[ \frac{1}{D(e^{j\omega})} \right] |H(e^{j\omega})|^2 \mathcal{S}_u(e^{j\omega}) d\omega, \end{aligned} \quad (2.136)$$

e

$$E \left[ u(n-i) \frac{u(n-i)}{D(z)} \right] = \frac{1}{2\pi} \int_0^{2\pi} \operatorname{Re} \left[ \frac{1}{D(e^{j\omega})} \right] \mathcal{S}_u(e^{j\omega}) d\omega, \quad (2.137)$$

e, portanto, existirão densidades espectrais  $\mathcal{S}_u(e^{j\omega})$  e  $|H(e^{j\omega})|^2$  para as quais, se (2.135) não for atendida, o traço de  $\mathbf{R}'(\mathbf{w}^*)$  será negativo, ou seja,  $\mathbf{R}'(\mathbf{w}^*)$  terá auto-valores com parte real negativa, e, conseqüentemente, pela Propriedade 2.2,  $\mathbf{w}^*$  não será um ponto de convergência do algoritmo. Portanto, para que  $\mathbf{w}^*$  seja um ponto de convergência para quaisquer  $\mathcal{S}_u(e^{j\omega})$  e  $|H(e^{j\omega})|^2$  a condição de positividade (2.135) não só é suficiente como é necessária.

Segue desta análise e da discussão do efeito de raízes comuns a  $B(z)$  e  $A(z)$  feita para o algoritmo RG, a seguinte propriedade:

**Propriedade 2.9** *Para o algoritmo PLR, dado por (2.131), se as sequências de entrada  $\{u(n)\}$  e do ruído de medida  $\{\zeta(n)\}$  forem estatisticamente independentes,  $\zeta(n)$  tiver média nula e o passo de adaptação  $\mu$  for suficientemente pequeno, então (2.135) é uma condição*

necessária e suficiente para que o vetor de parâmetros  $\mathbf{w}^*$  de um sistema  $H(z)$  qualquer com  $\deg[H(z)] = M$  seja um ponto de convergência global para qualquer entrada  $u(n)$  que seja uma excitação persistente de ordem  $2M + 1$  ou maior.

No enunciado acima, o termo *convergência global*, assim como também usado para o algoritmo SMM, corresponde a dizer que o domínio de atração de  $\mathbf{w}^*$  é todo o domínio de estabilidade.

Em aplicações onde a condição (2.135) pudesse ser garantida, o algoritmo PLR seria preferível ao algoritmo RG, por  $\mathbf{w}^*$  ser um ponto de convergência global e não apenas local, e preferível também ao algoritmo SMM, por não colocar restrições ao ruído de medida. Além disso, como será visto nos Itens 2.6 e 2.7, o comportamento da velocidade de convergência do algoritmo PLR quando os polos de  $H(z)$  tendem ao círculo unitário é mais favorável do que o dos algoritmos RG e SMM.

Em contrapartida a estas vantagens, porém, há o fato de que a condição (2.135) representa uma restrição importante ao sistema sendo modelado  $H(z)$ , como visto a seguir.

### 2.5.2 Influência da posição dos polos sobre a região de positividade

Na Figura 2-3 está indicada a região na qual devem estar os polos conjugados de um sistema  $H(z^{-1})$  de segunda ordem para atender a (2.135). Na Figura 2-4 está indicada a região na qual deve estar um par de polos conjugados de um sistema de quarta ordem que tem outro par em  $0,7 \angle \pm 30^\circ$  ou  $0,9 \angle \pm 30^\circ$  para atender a (2.135). Como pode ser visto, sistemas com polos próximos ao círculo unitário tendem a não satisfazer à condição (2.135).

### 2.5.3 Expressões de adaptação e condição para convergência do algoritmo SHARF

Para tentar atenuar a restrição que o algoritmo PLR coloca ao sistema sendo modelado, vista acima, podemos usar um filtro fixo  $P(z)$  para filtrar o erro de saída  $e(n)$ , obtendo as expressões de adaptação do algoritmo SHARF:

$$\begin{aligned} a_i(n+1) &= a_i(n) - \mu \hat{y}(n-i) \epsilon(n), \quad i = 1, 2, \dots, M, \\ b_i(n+1) &= b_i(n) + \mu u(n-i) \epsilon(n), \quad i = 0, 1, \dots, M, \\ \epsilon(n) &= P(z)e(n). \end{aligned} \tag{2.138}$$

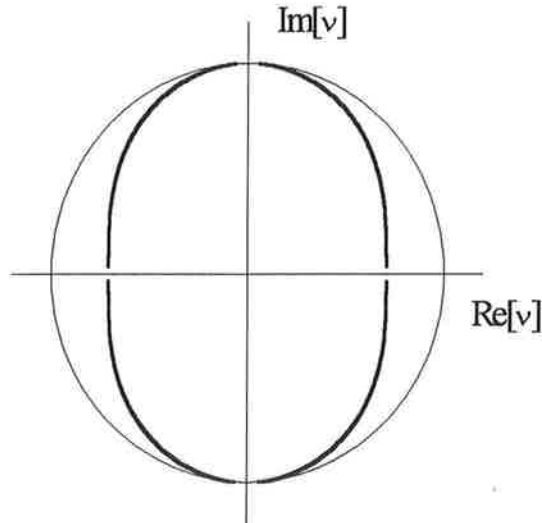


Figura 2-3: Algoritmo PLR, sistema de segunda ordem: região dos polos de  $H(z^{-1})$  onde a condição de positividade é obedecida; a circunferência representada é a circunferência de raio unitário.

O vetor  $\psi(n)$  na forma genérica (2.7) tem a mesma forma que no algoritmo PLR, sendo que agora temos que

$$\begin{aligned}
 \mathbf{e}^T \mathbf{R}'(\mathbf{w}) \mathbf{e} &= E[\mathbf{e}^T \psi(n, \mathbf{w}) \phi_2(n, \mathbf{w}) \mathbf{e}] \\
 &= E \left[ v(n) \frac{P(z)}{D(z)} v(n) \right] \\
 &= \frac{1}{2\pi} \int_0^{2\pi} \frac{P(e^{j\omega})}{D(e^{j\omega})} \mathcal{S}_v(e^{j\omega}) d\omega \\
 &= \frac{1}{2\pi} \int_0^{2\pi} \operatorname{Re} \left[ \frac{P(e^{j\omega})}{D(e^{j\omega})} \right] \mathcal{S}_v(e^{j\omega}) d\omega. \tag{2.139}
 \end{aligned}$$

Estipulando, portanto, a condição

$$\operatorname{Re} \left[ \frac{P(e^{j\omega})}{D(e^{j\omega})} \right] > 0, \quad -\pi \leq \omega < \pi, \tag{2.140}$$

e seguindo a mesma argumentação que para o algoritmo PLR, obtemos a propriedade:

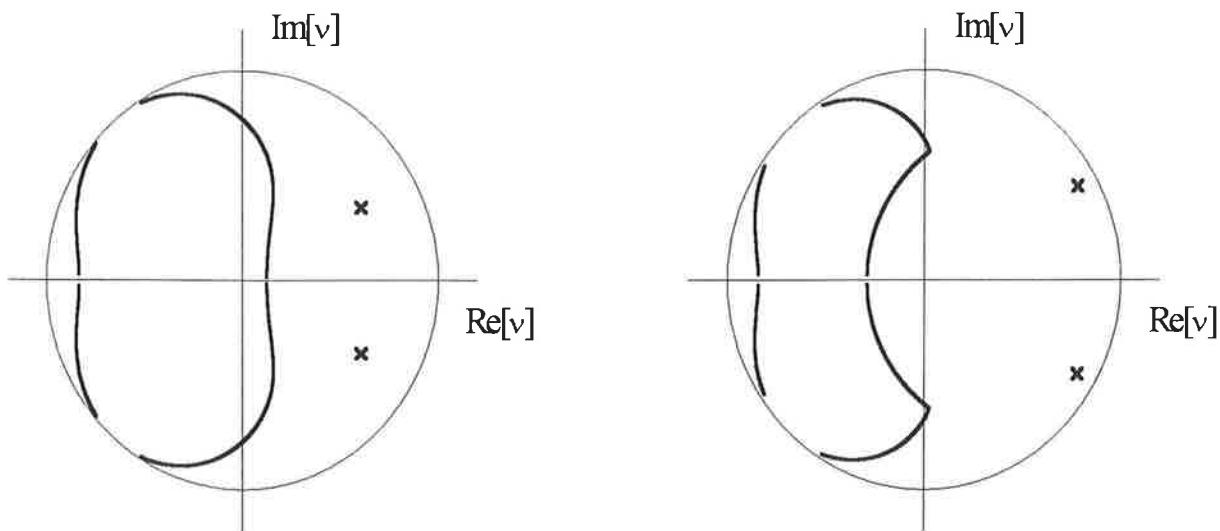


Figura 2-4: Algoritmo PLR, sistemas de quarta ordem (um par de polos de  $H(z^{-1})$  indicado com  $\times$ ): região do outro par de polos na qual a condição de positividade é obedecida; a circunferência representada é a circunferência de raio unitário.

**Propriedade 2.10** *Para o algoritmo SHARF, dado por (2.138), se as sequências de entrada  $\{u(n)\}$  e do ruído de medida  $\{\zeta(n)\}$  forem estatisticamente independentes,  $\zeta(n)$  tiver média nula e o passo de adaptação  $\mu$  for suficientemente pequeno, então (2.140) é uma condição necessária e suficiente para que o vetor de parâmetros  $\mathbf{w}^*$  de um sistema  $H(z)$  qualquer com  $\deg[H(z)] = M$  seja um ponto de convergência global para qualquer entrada  $u(n)$  que seja uma excitação persistente de ordem  $2M + 1$  ou maior.*

Notar que a condição de positividade (2.140) tem a mesma forma que a condição (2.108) referente ao algoritmo RG, sendo que a validade de ambas está vinculada à existência de uma boa estimativa inicial dos polos de  $H(z)$ . O efeito da posição destes polos para o algoritmo SHARF é portanto o mesmo exemplificado na Figuras 2-1. As implicações da validade das duas condições, porém, não são as mesmas. No caso do algoritmo RG, a condição depende dos parâmetros  $\mathbf{w}$  do filtro adaptativo, e é sempre atendida localmente. A sua validade em outras regiões do domínio de estabilidade permite, como visto, caracterizar parcialmente o comportamento global do algoritmo. No caso do algoritmo SHARF, a condição não depende dos parâmetros  $\mathbf{w}$  do filtro adaptativo. Porém, se for atendida, então o algoritmo converge globalmente.

## 2.6 Efeito da posição dos polos de $H(z)$ no condicionamento da matriz de estados da ODE

Neste item analisa-se o efeito da posição dos polos do sistema sendo modelado  $H(z)$  no condicionamento, isto é, no espalhamento dos valores singulares da matriz de estados dos algoritmos de adaptação de passo constante apresentados nos itens anteriores. Como discutido no Capítulo 1, o maior interesse na aplicação de filtros IIR adaptativos ocorre para casos em que  $H(z)$  tem polos pouco amortecidos, isto é, próximos à circunferência unitária. Sistemas com esta característica são considerados neste item, e é analisado, para cada algoritmo, se o espalhamento dos valores singulares da matriz de estados, no ponto correspondente aos parâmetros de  $H(z)$ , tende a infinito quando o raio de algum polo de  $H(z)$  tende a 1. A importância do espalhamento dos valores singulares desta matriz, como visto no Item 2.2.4, é que um alto espalhamento dos mesmos pode levar a uma baixa velocidade de convergência local do algoritmo.

Conforme (2.26),  $\mathbf{R}(\mathbf{w}^*)$  é a matriz de estados da equação diferencial associada ao algoritmo sendo analisado, calculada no vetor de parâmetros  $\mathbf{w}^*$  de  $H(z)$ . Dada uma função  $H(z)$ ,  $\mathbf{R}(\mathbf{w}^*)$  é obtida diretamente das expressões particulares de cada algoritmo e os valores singulares de  $\mathbf{R}(\mathbf{w}^*)$  são computados. Os diferentes algoritmos apresentados podem então ser comparados quanto ao critério do espalhamento destes valores singulares, para a função  $H(z)$  específica que foi considerada.

Por estarmos tratando de filtros adaptativos, porém, é de interesse analisar a dependência do espalhamento dos valores singulares de  $\mathbf{R}(\mathbf{w}^*)$  em relação a características mais genéricas de  $H(z)$ , como por exemplo, a proximidade à circunferência unitária de seus polos. Uma análise deste tipo não fornece, em geral, uma comparação direta entre o espalhamento para cada algoritmo, como no caso do cálculo para um determinado  $H(z)$ , mas, por outro lado, permite uma caracterização mais ampla dos diferentes algoritmos.

### 2.6.1 Expressões para a análise dos valores singulares

Os valores singulares de uma matriz qualquer  $\mathbf{A}$  são definidos como a raiz quadrada dos autovalores de  $\mathbf{A}^T \mathbf{A}$  [31]. Uma vez que  $\mathbf{A}^T \mathbf{A}$  é simétrica e positiva semi-definida, segue diretamente da Propriedade 1.1, que os valores singulares máximo e mínimo de  $\mathbf{R}(\mathbf{w}^*)$  são



dados por:

$$\sigma_{\max} = \sqrt{\max_{\|\mathbf{e}\|=1} \mathbf{e}^T \mathbf{R}^T(\mathbf{w}^*) \mathbf{R}(\mathbf{w}^*) \mathbf{e}}, \quad \sigma_{\min} = \sqrt{\min_{\|\mathbf{e}\|=1} \mathbf{e}^T \mathbf{R}^T(\mathbf{w}^*) \mathbf{R}(\mathbf{w}^*) \mathbf{e}}. \quad (2.141)$$

Se  $\mathbf{A}$  é uma matriz simétrica e positiva semi-definida qualquer, seus valores singulares são seus próprios autovalores, o que pode ser visto por:

$$\mathbf{A}\mathbf{u} = \lambda\mathbf{u} \Rightarrow \mathbf{A}^T \mathbf{A}\mathbf{u} = \lambda\mathbf{A}^T \mathbf{u} = \lambda\mathbf{A}\mathbf{u} = \lambda^2\mathbf{u}. \quad (2.142)$$

Portanto, se  $\mathbf{R}(\mathbf{w}^*)$  é simétrica e positiva semi-definida, segue diretamente da Propriedade 1.1 que seus valores singulares máximo e mínimo são dados por

$$\sigma_{\max} = \max_{\|\mathbf{e}\|=1} \mathbf{e}^T \mathbf{R}(\mathbf{w}^*) \mathbf{e}, \quad \sigma_{\min} = \min_{\|\mathbf{e}\|=1} \mathbf{e}^T \mathbf{R}(\mathbf{w}^*) \mathbf{e}. \quad (2.143)$$

No que segue, são obtidas, para os algoritmos de adaptação considerados nos itens anteriores, expressões para a forma quadrática  $\mathbf{e}^T \mathbf{R}(\mathbf{w}^*) \mathbf{e}$  utilizando produtos internos de funções racionais em  $z$ . Com estas expressões, o efeito da posição dos polos de  $H(z)$  no condicionamento de  $\mathbf{R}(\mathbf{w}^*)$  será então analisado.

Consideremos os algoritmos RG e SMM, analisados nos Itens 2.3 e 2.4. De (2.33), (2.106) e (2.129) temos que, para os dois algoritmos,  $\mathbf{R}(\mathbf{w}^*)$  é simétrica e, escrevendo, como em (1.38),  $\mathbf{e} = [r_0 \ r_1 \ \dots \ r_{M-1} \ q_0 \ q_1 \ \dots \ q_M]^T$ ,  $R(z) = r_0 + r_1 z + \dots + r_{M-1} z^{M-1}$  e  $Q(z) = q_0 + q_1 z + \dots + q_M z^M$ , e notando que em  $\mathbf{w}^*$  vale  $A(z) = D(z)$ , temos

$$\begin{aligned} \mathbf{e}^T \mathbf{R}(\mathbf{w}^*) \mathbf{e} &= E \left[ \frac{v(n)}{D(z)} \frac{v(n)}{D(z)} \right] \\ &= \left\langle \frac{F(z)}{D(z)}, \frac{F(z)}{D(z)} \mathcal{S}_u(z) \right\rangle, \end{aligned} \quad (2.144)$$

onde

$$v(n) = F(z)u(n), \quad (2.145)$$

e

$$F(z) = Q(z) - zR(z)H(z). \quad (2.146)$$

Notar, pela definição de  $\mathbf{e}$ , que a restrição  $\|\mathbf{e}\| = 1$  em (2.141) e (2.143) significa que ao variar  $Q(z)$  e  $R(z)$  em (2.146) devemos manter  $\|Q(z)\|^2 + \|R(z)\|^2 = 1$ .

Consideremos agora o caso dos algoritmos PLR e SHARF, analisados no Item 2.5. No algoritmo SHARF,  $P(z)$  é uma estimativa (fixa) de  $D(z)$ . Podemos considerar então os dois casos extremos: 1) o caso ideal (em termos do atendimento da condição de positividade (2.140)),  $P(z) = D(z)$  e 2)  $P(z) \equiv 1$ , que recai no algoritmo PLR. Para o caso  $P(z) = D(z)$ , de (2.33) e (2.139) temos

$$\begin{aligned} \mathbf{e}^\top \mathbf{R}(\mathbf{w}^*) \mathbf{e} &= E \left[ v(n) \frac{P(z)}{D(z)} v(n) \right] \\ &= E [v(n)v(n)] \\ &= \langle F(z), F(z) \mathcal{S}_u(z) \rangle, \end{aligned} \tag{2.147}$$

com  $F(z)$  dado ainda por (2.146).

No caso em que  $P(z) \equiv 1$ , temos de (2.33) e (2.133) que  $\mathbf{R}(\mathbf{w}^*)$  não é simétrica. Para poder usar a Propriedade 1.1 escrevemos então

$$\mathbf{e}^\top \mathbf{R}^\top(\mathbf{w}^*) \mathbf{R}(\mathbf{w}^*) \mathbf{e} = \mathbf{e}_R^\top \mathbf{e}_R,$$

onde

$$\mathbf{e}_R = E[\boldsymbol{\psi}(n, \mathbf{w}^*) \boldsymbol{\phi}_2^\top(n, \mathbf{w}^*) \mathbf{e}]$$

$$= \begin{bmatrix} - \left\langle \frac{F(z)}{D(z)}, \frac{zC(z)}{D(z)} \mathcal{S}_u(z) \right\rangle \\ - \left\langle \frac{F(z)}{D(z)}, \frac{z^2 C(z)}{D(z)} \mathcal{S}_u(z) \right\rangle \\ \vdots \\ - \left\langle \frac{F(z)}{D(z)}, \frac{z^M C(z)}{D(z)} \mathcal{S}_u(z) \right\rangle \\ \left\langle \frac{F(z)}{D(z)}, \mathcal{S}_u(z) \right\rangle \\ \left\langle \frac{F(z)}{D(z)}, z \mathcal{S}_u(z) \right\rangle \\ \vdots \\ \left\langle \frac{F(z)}{D(z)}, z^{M-1} \mathcal{S}_u(z) \right\rangle \end{bmatrix},$$

com  $F(z)$  dado por (2.146). Resulta, portanto,

$$\mathbf{e}^\top \mathbf{R}^\top(\mathbf{w}^*) \mathbf{R}(\mathbf{w}^*) \mathbf{e} = \sum_{i=1}^M \left\langle \frac{F(z)}{D(z)}, \frac{z^i C(z)}{D(z)} \mathcal{S}_u(z) \right\rangle^2 + \sum_{i=0}^M \left\langle \frac{F(z)}{D(z)}, z^i \mathcal{S}_u(z) \right\rangle^2. \quad (2.148)$$

## 2.6.2 Efeito do sinal de entrada

Como pode ser visto nas expressões (2.141), (2.143), (2.144), (2.147) e (2.148), para todos os algoritmos considerados as características espectrais do sinal de entrada  $u(n)$  afetam os valores singulares de  $\mathbf{R}(\mathbf{w}^*)$ . É importante notar que esta influência pode ser tanto no sentido de reduzir quanto no de aumentar o espalhamento dos valores singulares, dependendo da relação entre  $\mathcal{S}_u(z)$  e os polinômios  $C(z)$  e  $D(z)$  que compõe  $H(z)$ . Se considerarmos  $u(n)$  branco, o que tem a vantagem de simplificar a análise, não estaremos portanto nos restringindo a uma situação particularmente favorável ou desfavorável no que diz respeito ao espalhamento dos valores singulares de  $\mathbf{R}(\mathbf{w}^*)$ .

Considerando então  $u(n)$  branco e de potência unitária, temos para o algoritmos RG e SMM,

$$\mathbf{e}^\top \mathbf{R}(\mathbf{w}^*) \mathbf{e} = \left\langle \frac{F(z)}{D(z)}, \frac{F(z)}{D(z)} \right\rangle, \quad (2.149)$$

para o algoritmo SHARF com, idealmente,  $P(z) = D(z)$ ,

$$\mathbf{e}^\top \mathbf{R}(\mathbf{w}^*) \mathbf{e} = \langle F(z), F(z) \rangle, \quad (2.150)$$

e para o algoritmo PLR,

$$\mathbf{e}^\top \mathbf{R}^\top(\mathbf{w}^*) \mathbf{R}(\mathbf{w}^*) \mathbf{e} = \sum_{i=1}^M \left\langle \frac{F(z)}{D(z)}, \frac{z^i C(z)}{D(z)} \right\rangle^2 + \sum_{i=0}^M \left\langle \frac{F(z)}{D(z)}, z^i \right\rangle^2. \quad (2.151)$$

### 2.6.3 Restrições a $H(z)$

Com as expressões (2.149), (2.150) e (2.151), pode ser verificado qual o comportamento dos valores singulares de  $\mathbf{R}(\mathbf{w}^*)$  quando algum polo de  $H(z)$  se aproxima da circunferência unitária. Notar que esta variação da posição dos polos de  $H(z)$  não visa representar uma eventual variação no tempo do sistema sendo modelado, mas sim representar diferentes funções de transferência  $H(z)$  que este sistema (que, em uma aplicação prática de filtros adaptativos, é desconhecido para o usuário) pode assumir. Neste sentido, ao variar os polos de  $H(z)$ , diferentes restrições aos seus demais parâmetros podem ser assumidas, relacionadas com os mecanismos físicos envolvidos na aplicação prática em questão. Serão considerados aqui três casos "ideais":

1.  $C(z)$  é fixo para diferentes posições das raízes de  $D(z)$ , e, em nenhum caso cancela alguma destas raízes;
2. Os zeros de  $H(z)$  são fixos para diferentes posições das raízes de  $D(z)$ , em nenhum caso cancelando alguma destas raízes, porém  $\|H(z)\|$  é fixo para diferentes posições das raízes de  $D(z)$ .
3.  $H(z)$  é passa-tudo.

No caso da Restrição 1 temos que se algum polo de  $H(z)$  se aproxima da circunferência unitária sua norma  $L_2$  tende a infinito, o que não ocorre no caso da Restrição 2. Parece razoável que, em algumas aplicações práticas, a despeito da posição dos polos do sistema, a norma  $L_2$  do mesmo possa não exibir grandes variações. Os casos 1 e 2 constituem, neste sentido, dois possíveis extremos entre os quais situações reais estariam localizadas. Por sua vez, a Restrição 3 pode representar, de fato, uma característica do sistema sendo modelado.

O interesse em considerar este caso, no entanto, advém também da possibilidade, pouco investigada até o momento, de que a utilização de blocos passa-tudo em filtros adaptativos possa ter vantagens em relação a outras estruturas.

A seguir, para cada uma das restrições acima, analisa-se como variam os valores singulares de  $\mathbf{R}(\mathbf{w}^*)$  para cada algoritmo de adaptação, quando  $H(z)$  tem polos que se aproximam da circunferência unitária. O principal objetivo da análise a seguir é determinar os casos em que o espalhamento dos valores singulares de  $\mathbf{R}(\mathbf{w}^*)$  tende a infinito. Para tanto é suficiente obter, para o valor singular máximo, um limite inferior que tende a infinito quando algum polo de  $H(z)$  se aproxima da circunferência unitária e, para o valor singular mínimo, um limite superior finito; ou equivalentemente, obter, para o valor singular máximo, um limite inferior finito, e mostrar que o valor singular mínimo tende a zero quando algum polo de  $H(z)$  se aproxima da circunferência unitária.

Além deste aspecto, a análise a seguir também estabelece relações entre as velocidades com que os valores singulares extremos de  $\mathbf{R}(\mathbf{w}^*)$  tendem a infinito ou a zero quando algum polo de  $H(z)$  se aproxima da circunferência unitária.

## 2.6.4 Valores singulares extremos de $\mathbf{R}(\mathbf{w}^*)$ considerando a Restrição 1

### Valor singular máximo de $\mathbf{R}(\mathbf{w}^*)$

Fazendo  $R(z) \equiv 1$  e  $Q(z) \equiv 0$  em (2.146), ou seja,  $\mathbf{e} = [1 \ 0 \ \dots \ 0]^T$ , temos  $F(z) = -zH(z) = -zC(z)/D(z)$ , e, portanto, para os algoritmos RG e SMM, de (2.143) e (2.149),

$$\sigma_{\max, RG} = \max_{\|\mathbf{e}\|=1} \left\langle \frac{F(z)}{D(z)}, \frac{F(z)}{D(z)} \right\rangle \geq \left\langle \frac{C(z)}{D^2(z)}, \frac{C(z)}{D^2(z)} \right\rangle = \left\langle \frac{H(z)}{D(z)}, \frac{H(z)}{D(z)} \right\rangle \triangleq \sigma_1, \quad (2.152)$$

para o algoritmo SHARF, de (2.150),

$$\sigma_{\max, SHARF} = \max_{\|\mathbf{e}\|=1} \langle F(z), F(z) \rangle \geq \left\langle \frac{C(z)}{D(z)}, \frac{C(z)}{D(z)} \right\rangle = \langle H(z), H(z) \rangle \triangleq \sigma_3, \quad (2.153)$$

e para o algoritmo PLR, de (2.151), notando que a segunda somatória à direita nesta expressão é sempre limitada,

$$\begin{aligned}
\sigma_{\max, PLR} &= \left[ \max_{\|e\|=1} \sum_{i=1}^M \left\langle \frac{F(z)}{D(z)}, \frac{z^i C(z)}{D(z)} \right\rangle^2 + \sum_{i=0}^M \left\langle \frac{F(z)}{D(z)}, z^i \right\rangle^2 \right]^{1/2} \\
&\geq \max_{\|e\|=1} \left| \left\langle \frac{F(z)}{D(z)}, \frac{zC(z)}{D(z)} \right\rangle \right| \\
&\geq \left| \left\langle \frac{C(z)}{D^2(z)}, \frac{C(z)}{D(z)} \right\rangle \right| \triangleq \sigma_2.
\end{aligned} \tag{2.154}$$

Os limites inferiores para o valor singular máximo de  $\mathbf{R}(\mathbf{w}^*)$  para cada algoritmo, dados pelas expressões acima, são também aproximações razoáveis dos mesmos quando algum polo de  $H(z)$  se aproxima da circunferência unitária, pois, quando isto acontece, de modo a maximizar os produtos internos em (2.149), (2.150) e (2.151), o resíduo deste mesmo polo em  $F(z)$  não pode ser baixo, o que, lembrando que deve-se manter  $\|Q(z)\|^2 + \|R(z)\|^2 = 1$ , conduz a valores de  $\|R(z)\|$  próximos a 1.

As expressões (2.152) e (2.153) são as normas  $L_2$  de  $C(z)/D^2(z)$  e  $C(z)/D(z)$ , respectivamente. Portanto, uma vez que  $C(z)$  é fixo e não cancela nenhuma raiz de  $D(z)$ , o valor singular máximo de  $\mathbf{R}(\mathbf{w}^*)$  para os algoritmos RG, SMM e SHARF tende a infinito quando algum polo de  $H(z)$  se aproxima da circunferência unitária.

Para o caso do algoritmo PLR a análise não é tão imediata, pois a expressão à direita em (2.154) não é a norma  $L_2$  de uma função. Para analisar este caso, pode-se notar, de início, que se for assumido que  $\deg[C(z)] < \deg[D(z)]$  não há perda de generalidade em relação ao aspecto que deseja-se investigar. Isto resulta do fato de que uma função  $H(z) = C'(z)/D(z)$  sempre pode ser escrita como  $C(z)/D(z) + k(z)$  onde  $\deg[C(z)] < \deg[D(z)]$  e  $k(z)$  é um polinômio com norma finita, e, conseqüentemente, o produto interno com a forma de (2.154) pode ser escrito como

$$\left\langle \frac{C'(z)}{D(z)}, \frac{C'(z)}{D^2(z)} \right\rangle = \left\langle \frac{C(z)}{D(z)}, \frac{C(z)}{D^2(z)} \right\rangle + \left\langle \frac{k(z)}{D(z)}, \frac{C(z)}{D(z)} \right\rangle + \left\langle k(z), \frac{C(z)}{D^2(z)} \right\rangle + \left\langle k(z), \frac{g_0}{D(z)} \right\rangle. \tag{2.155}$$

Os dois últimos termos da expressão acima não tendem a infinito se algum polo de  $H(z)$  se aproximar da circunferência unitária, mas para os dois primeiros termos isto pode ocorrer, sendo que, como será visto, o primeiro termo da expressão acima poderá determinar um

crescimento mais rápido do produto interno do que o segundo termo. No Item A.1 do Apêndice A demonstra-se que se o raio  $r$  de algum polo de  $H(z)$  se aproxima de 1, o termo  $|\langle C(z)/D^2(z), C(z)/D(z) \rangle|$  é bem aproximado pelo produto de um termo  $|\gamma|$ , que tende a infinito quando  $r$  tende a 1, e  $|\langle C(z)/D(z), C(z)/D(z) \rangle|$  (que também tende a infinito quando  $r$  tende a 1). Para indicar esta propriedade escreve-se então

$$\left| \left\langle \frac{C(z)}{D^2(z)}, \frac{C(z)}{D(z)} \right\rangle \right| \xrightarrow{r \rightarrow 1} |\gamma| \left\langle \frac{C(z)}{D(z)}, \frac{C(z)}{D(z)} \right\rangle, \quad |\gamma| \xrightarrow{r \rightarrow 1} \infty. \quad (2.156)$$

Portanto, no caso Restrição 1, o valor singular máximo de  $\mathbf{R}(\mathbf{w}^*)$  para o algoritmo PLR também tende a infinito quando algum polo de  $H(z)$  se aproxima da circunferência unitária.

### Velocidade de crescimento de $\sigma_{\max}$

Consideremos agora que quando  $H(z)$  tem algum polo que se aproxima da circunferência unitária o valor singular máximo de  $\mathbf{R}(\mathbf{w}^*)$ , para todos os algoritmos considerados, seja bem aproximado por uma lei do tipo

$$\log[\sigma_{\max}] \xrightarrow{r \rightarrow 1} l_0 - v \log(r - 1), \quad (2.157)$$

ou equivalentemente,

$$\sigma_{\max} \xrightarrow{r \rightarrow 1} \frac{\sigma_0}{(r - 1)^v}, \quad (2.158)$$

onde  $\sigma_0$  e  $v$  dependem do algoritmo específico, sendo que  $v$  corresponde à velocidade com que  $\sigma_{\max}$  tende a infinito quando  $r$  tende a 1. Consideremos que uma lei deste tipo, e com o mesmo valor de  $v$ , vale também para os limites inferiores de  $\sigma_{\max}$ , dados pelas expressões (2.152), (2.153) e (2.154). Ou seja, respectivamente para os algoritmos RG/SMM, PLR e SHARF,

$$\sigma_1 \xrightarrow{r \rightarrow 1} \frac{\sigma_{1,0}}{(r - 1)^{v_1}}, \quad \sigma_2 \xrightarrow{r \rightarrow 1} \frac{\sigma_{2,0}}{(r - 1)^{v_2}}, \quad \sigma_3 \xrightarrow{r \rightarrow 1} \frac{\sigma_{3,0}}{(r - 1)^{v_3}}. \quad (2.159)$$

A validade desta suposição será verificada em alguns exemplos numéricos apresentados posteriormente.

Para os algoritmos considerados, podemos então comparar a velocidade  $v$  com que os limites inferiores para  $\sigma_{\max}$  tendem a infinito quando  $r$  tende a 1. Para comparar os algoritmos

PLR e SHARF, com as expressões (2.153), (2.154) e (2.156) temos que

$$0 \leq \lim_{r \rightarrow 1} \frac{\sigma_3}{\sigma_2} = \lim_{r \rightarrow 1} |\gamma|^{-1} = 0. \quad (2.160)$$

Segue que se  $\sigma_3$  e  $\sigma_2$  tem a forma dada por (2.159), a velocidade  $v_2$  é maior do que a velocidade  $v_3$ .

Para comparar agora os algoritmos RG/SMM e PLR, com  $\sigma_1$  e  $\sigma_2$  definidos, respectivamente, em (2.152) e (2.154), demonstra-se no Item A.2 do Apêndice A que

$$\lim_{r \rightarrow 1} \frac{\sigma_2}{\sigma_1} \leq \lim_{r \rightarrow 1} \frac{1}{\left| \frac{1}{D(1/p_j)} \right|} = 0, \quad (2.161)$$

onde  $p_j$  é algum dos polos de raio  $r$  de  $H(z)$ . Portanto, se  $\sigma_2$  e  $\sigma_1$  tem a forma dada por (2.159), a velocidade  $v_1$  é maior do que a velocidade  $v_2$ .

Chegamos, portanto, na seguinte relação de velocidades para os algoritmos considerados:

$$\boxed{v_1 \geq v_2 \geq v_3}. \quad (2.162)$$

A comparação entre os algoritmos realizada acima, como mencionado na introdução deste item, não é feita diretamente em termos do valor singular máximo de  $\mathbf{R}(\mathbf{w}^*)$  em cada caso. É importante notar, porém, que, uma vez que  $\sigma_1$ ,  $\sigma_2$  e  $\sigma_3$  tendem para infinito quando  $r$  tende a 1, a relação acima permite afirmar que, (2.159) sendo válido, existe um valor de  $r$  acima do qual  $\sigma_1 \geq \sigma_2$ ,  $\sigma_3$  e outro valor de  $r$  acima do qual  $\sigma_2 \geq \sigma_3$ .

### Valor singular mínimo de $\mathbf{R}(\mathbf{w}^*)$

Podemos passar a analisar agora o valor singular mínimo de  $\mathbf{R}(\mathbf{w}^*)$  para cada algoritmo.

Consideremos, inicialmente, que  $\deg[H(z)] = M$  e  $H(z)$  tem  $m = M$  polos com raio  $r$ .

De (2.143) e (2.149) temos, para os algoritmos RG e SMM,

$$\sigma_{\min, RG} = \min_{\|\mathbf{e}\|=1} \left\langle \frac{F(z)}{D(z)}, \frac{F(z)}{D(z)} \right\rangle = \min_{\|\mathbf{e}\|=1} \left\| \frac{F(z)}{D(z)} \right\|. \quad (2.163)$$



Com  $F(z)$  dado por (2.146) temos

$$\frac{F(z)}{D(z)} = \frac{Q(z)}{D(z)} - zR(z)\frac{C(z)}{D^2(z)}. \quad (2.164)$$

A menos que  $R(z) \equiv 0$  temos que  $\|R(z)C(z)/D^2(z)\|$  tende a infinito quando  $r$  tende a 1, pois  $\deg[R(z)] < M$  e, portanto,  $R(z)$  não pode cancelar todas as raízes de  $D^2(z)$ . Se  $R(z) \not\equiv 0$ , então para que  $\|F(z)/D(z)\|$  não tenda também a infinito é necessário que  $Q(z)D(z) = zR(z)C(z)$ , o que exigiria que  $R(z)$  tivesse as raízes de  $D(z)$ . Isto, porém, também não é possível, novamente devido ao fato de que  $\deg[R(z)] < M$ . Segue que, para obter  $\sigma_{\min, RG}$  deve-se fazer  $R(z) \equiv 0$ . Quanto a  $Q(z)$ , deve-se ter  $Q(z) = D(z)/\|D(z)\|$  (o que é possível pois  $\deg[Q(z)] = M$ ) do contrário  $\|Q(z)/D(z)\|$  tenderia a infinito. Temos portanto que

$$\lim_{r \rightarrow 1} \sigma_{\min, RG} = k_1, \quad m = M, \quad (2.165)$$

onde

$$k_1 = \lim_{r \rightarrow 1} 1/\|D(z)\|^2 \quad (2.166)$$

Para o algoritmo SHARF, de (2.143) e (2.150) temos,

$$\sigma_{\min, SHARF} = \min_{\|e\|=1} \langle F(z), F(z) \rangle. \quad (2.167)$$

Com  $F(z) = Q(z) - zR(z)H(z)$  pode-se ver que, uma vez que  $\deg[R(z)] < M$ , devemos ter  $R(z) \equiv 0$ , e portanto  $\|Q(z)\| = 1$ , do contrário  $\|F(z)\|$  tenderá a infinito quando  $r$  tender a 1. Resulta então,

$$\lim_{r \rightarrow 1} \sigma_{\min, SHARF} = 1, \quad m = M. \quad (2.168)$$

Se, por outro lado,  $\deg[H(z)] = M$  e  $H(z)$  tem  $m < M$  polos com raio  $r$ , escrevendo  $D(z) = D_1(z)D_2(z)$ , onde  $D_1(z)$  reúne as  $m$  raízes de raio  $r$  de  $D(z)$ , podemos fazer  $R(z) = \kappa D_1(z)$  o que resulta em

$$F(z) = Q(z) - z\kappa \frac{C(z)}{D_2(z)}, \quad (2.169)$$

sendo que agora, mesmo com  $\kappa \neq 0$ ,  $\|F(z)\|$  não tende a infinito se  $r$  tender a 1. Temos então

$$\sigma_{\min, SHARF} \leq 1, \quad m < M. \quad (2.170)$$

Nesta mesma situação, para os algoritmos RG e SMM, para que  $\|F(z)/D(z)\|$  não tenda a infinito, as raízes de  $D_1(z)$  devem ser raízes de  $Q(z)D_2(z) - z\kappa C(z)$ , o que é sempre possível uma vez que  $D_1(z)$  e  $D_2(z)$  não tem raízes comuns. Portanto,

$$\lim_{r \rightarrow 1} \sigma_{\min, RG} \leq k_1, \quad m < M. \quad (2.171)$$

Para o algoritmo PLR, independentemente do valor de  $m$ , fazendo em (2.146)

$$Q(z) = D(z)/\|D(z)\|, \quad R(z) \equiv 0, \quad (2.172)$$

e usando (2.151) temos

$$\sigma_{\min, PLR} \leq \sqrt{k_1} \left[ \sum_{i=1}^M \left\langle 1, \frac{z^i C(z)}{D(z)} \right\rangle^2 + \sum_{i=0}^M \langle 1, z^i \rangle^2 \right]^{1/2} = \sqrt{k_1}, \quad (2.173)$$

onde o termo entre colchetes é igual a 1 devido à causalidade de  $H(z)$ .

## 2.6.5 Valores singulares extremos de $R(w^*)$ considerando a Restrição 2

Notando que se  $\|H(z)\| = 1$  então  $H(z)$  sempre pode ser escrito como

$$H(z) = \frac{C(z)}{D(z)} = \frac{1}{\sqrt{\left\langle \frac{C'(z)}{D(z)}, \frac{C'(z)}{D(z)} \right\rangle}} \frac{C'(z)}{D(z)}, \quad (2.174)$$

onde  $C'(z)$  tem as raízes de  $C(z)$  mas é fixo, temos que a expressão (2.152) para o algoritmo RG pode ser escrita, quando  $\|H(z)\| = 1$ , como

$$\sigma_{\max, RG} \geq \sigma_4 \triangleq \left\langle \frac{C(z)}{D^2(z)}, \frac{C(z)}{D^2(z)} \right\rangle = \frac{\left\langle \frac{C'(z)}{D^2(z)}, \frac{C'(z)}{D^2(z)} \right\rangle}{\left\langle \frac{C'(z)}{D(z)}, \frac{C'(z)}{D(z)} \right\rangle}. \quad (2.175)$$

Comparando agora esta expressão com as expressões (2.152), (2.153) e (2.159) temos que se algum polo de  $H(z)$  se aproxima da circunferência unitária então  $\sigma_4$  tende a infinito com velocidade  $v_1 - v_3$ .

Para o algoritmo SHARF segue diretamente de (2.153) que  $\sigma_{\max,SHARF} \geq 1$ , e, após algumas manipulações simples que  $\sigma_{\max,SHARF} \leq (1 + \sqrt{2})^2$ . Ou seja, neste caso,  $\sigma_{\max,SHARF}$  não tende a infinito quando algum polo de  $H(z)$  se aproxima da circunferência unitária.

Para o algoritmo PLR, temos de (2.154) e (2.156), que

$$\sigma_{\max,PLR} \geq \sigma_5 \triangleq \left| \left\langle \frac{C(z)}{D^2(z)}, \frac{C(z)}{D(z)} \right\rangle \right| \xrightarrow{r \rightarrow 1} |\gamma| \left\langle \frac{C(z)}{D(z)}, \frac{C(z)}{D(z)} \right\rangle = |\gamma|, \quad (2.176)$$

e, portanto, se algum polo de  $H(z)$  se aproxima da circunferência unitária  $\sigma_5$  tende a infinito com velocidade  $v_2 - v_3$ . Notar que, em (A.7), escrevendo  $H(z)$  como em (2.174), percebe-se que todos os resíduos  $\alpha_i$  tendem a zero quando  $r$  tende a 1, e portanto a Condição 1 para a validade de (A.9) é atendida.

Consideremos agora os valores singulares mínimos, inicialmente com  $m = M$ .

Para os algoritmos RG e SMM temos a mesma situação que para a Restrição 1 (ver análise anterior): a menos que  $R(z) \equiv 0$ , ainda temos que  $\|R(z)C(z)/D^2(z)\|$  tende a infinito quando  $r$  tende a 1, uma vez que  $C(z)$  mantém apenas  $\|C(z)/D(z)\| = 1$ , e não  $\|C(z)/(D_1(z)D(z))\| = 1$ , onde as raízes de  $D_1(z)$  são as raízes de  $D(z)$  não canceladas por  $R(z)$ . Portanto vale ainda  $\lim_{r \rightarrow 1} \sigma_{\min,RG} = k_1$ ,  $m = M$ .

Para o algoritmo SHARF, como no caso da Restrição 1, se  $\|Q(z)\| = 1$  então  $\|F(z)\| = 1$ , onde, como já visto,  $F(z) = Q(z) - zR(z)H(z)$ . Porém, temos que agora não é necessário que  $R(z) \equiv 0$  e  $\|Q(z)\| = 1$  para que  $\|F(z)\|$  não tenda a infinito quando  $r$  tende a 1. Temos então que  $\sigma_{\min,SHARF} \leq 1$ ,  $m = M$ . Neste caso, é de interesse saber, adicionalmente, se pode-se garantir que  $\sigma_{\min,SHARF}$  não tende a zero, pois, como  $\sigma_{\max,SHARF}$  não tende a infinito, isto significaria que o espalhamento dos valores singulares não tenderia a infinito. Pode-se verificar que se  $H(z)$  só tem polos simples então, quando os polos de  $H(z)$  se aproximam da circunferência unitária,  $\sigma_{\min,SHARF}$  tende a um valor não nulo. Este valor é tanto menor quanto mais próximos forem os polos de  $H(z)$ , se anulando no caso limite em que  $H(z)$  tem polos múltiplos. Temos então

$$0 < \sigma_{\min,SHARF} \leq 1, \quad m = M, \quad H(z) \text{ só tem polos simples}, \quad (2.177)$$

$$\lim_{r \rightarrow 1} \sigma_{\min,SHARF} = 0, \quad m = M, \quad H(z) \text{ tem algum polo múltiplo}. \quad (2.178)$$

Para o algoritmo PLR vale a mesma análise feita para a Restrição 1.

Consideremos agora o caso  $m < M$ . A partir da análise feita no caso da Restrição 1 para os algoritmos RG/SMM e SHARF (ver (2.169)), temos agora que, com  $\kappa = 1$ ,  $\|F(z)\|$  tende a zero quando algum polo de  $H(z)$  se aproxima da circunferência unitária, pois  $\|C(z)\|$  tende a zero (ver (2.174)) e  $D_2(z)$  não tem nenhuma raiz que se aproxima da circunferência unitária. Neste caso, então, para todos os algoritmos, os valores singulares mínimos tendem a zero. Para os algoritmos RG e SHARF, em particular, a velocidade com que  $\sigma_{\min}$  tende a zero é a velocidade com que  $\|C(z)\|$  tende a zero, a qual, de (2.174), (2.153) e (2.159) é igual a  $v_3$ .

### 2.6.6 Valores singulares extremos de $R(w^*)$ considerando a Restrição 3

Com  $H(z)$  passa-tudo temos  $H(z)H(z^{-1}) \equiv 1$ , o que resulta, para o algoritmo RG, em

$$\sigma_{\max, RG} \geq \sigma_6 \triangleq \left\langle \frac{C(z)}{D^2(z)}, \frac{C(z)}{D^2(z)} \right\rangle = \left\langle \frac{1}{D(z)}, \frac{C(z^{-1})}{D(z^{-1})} \frac{C(z)}{D^2(z)} \right\rangle = \left\langle \frac{1}{D(z)}, \frac{1}{D(z)} \right\rangle, \quad (2.179)$$

e, portanto,  $\sigma_6$  tende para infinito com velocidade  $v_3$  quando algum polo de  $H(z)$  se aproxima da circunferência unitária.

Para o algoritmo SHARF vale a mesma análise relativa ao valor singular máximo feita para a Restrição 2. Para o algoritmo PLR podemos considerar diretamente a expressão (2.154). Nesta expressão, a segunda somatória tem sempre um valor finito. Para os termos da primeira somatória temos, usando  $H(z)H(z^{-1}) \equiv 1$ ,

$$\left\langle \frac{1}{D(z)} [Q(z) - zR(z)H(z)], z^i H(z) \right\rangle = \left\langle \frac{Q(z)}{D(z)}, z^i H(z) \right\rangle - \left\langle z \frac{R(z)}{D(z)}, z^i \right\rangle. \quad (2.180)$$

O segundo termo da expressão acima tem sempre um valor finito, ao passo que para o primeiro termo, escrevendo  $H(z) = z^M D(z^{-1})/D(z)$  temos

$$\begin{aligned} \left\langle \frac{Q(z)}{D(z)}, z^i H(z) \right\rangle &= \left\langle \frac{Q(z)}{D(z)} D(z), \frac{z^{i+M}}{D(z)} \right\rangle \\ &= \left\langle Q(z), \frac{z^{i+M}}{D(z)} \right\rangle \\ &= 0, \end{aligned} \quad (2.181)$$

pois  $\deg Q(z) \leq M$  e  $i > 0$ . Portanto, neste caso  $\sigma_{\max, PLR}$  não tende a infinito se algum polo de  $H(z)$  se aproxima da circunferência unitária.

Em relação aos valores singulares mínimos, para o algoritmo RG vale a mesma análise feita para a Restrição 1. Temos então para  $m = M$ ,  $\lim_{r \rightarrow 1} \sigma_{\min, RG} = k_1$  e para  $m < M$ ,  $\sigma_{\min, RG} \leq k_1$ .

Para o algoritmo SHARF, porém, o mesmo não ocorre: com  $F(z) = Q(z) - zR(z)H(z)$ , se  $m = M$  então  $H(z)$  tende a 1 quando seus polos se aproximam da circunferência unitária e portanto  $\|F(z)\|$  tende a zero se  $Q(z) = -zR(z)$ . Se  $m < M$  então  $H(z)$  tende a uma função  $H_1(z)$  com grau menor do que  $M$ , e pode-se escolher  $R(z)$  de modo a cancelar os polos de  $H_1(z)$  e  $Q(z)$  de modo a cancelar o numerador de  $zH_1(z)$  e fazer  $\|F(z)\|$  tender a zero. Portanto,

$$\lim_{r \rightarrow 1} \sigma_{\min, SHARF} = 0. \quad (2.182)$$

Para o algoritmo PLR, temos, como visto acima, que o primeiro termo à direita em (2.180) é nulo e que o segundo termo em (2.180) só é nulo se  $R(z) \equiv 0$ , ao passo que em (2.154), os elementos da segunda somatória são dados por

$$\left\langle \frac{F(z)}{D(z)}, z^i \right\rangle = \left\langle \frac{1}{D(z)} [Q(z) - zR(z)H(z)], z^i \right\rangle, \quad (2.183)$$

e não são nulos se  $R(z) \equiv 0$ . Portanto  $\sigma_{\min, PLR}$ , neste caso, não tende a 0 quando algum polo de  $H(z)$  se aproxima da circunferência unitária.

### 2.6.7 Resumo dos resultados da análise de valores singulares

Os resultados da análise feita acima estão reunidos na Tabela 2.1. Repetindo o que está comentado na introdução deste capítulo, três aspectos importantes podem ser destacados:

- A única situação em que o espalhamento dos valores singulares de  $\mathbf{R}(\mathbf{w}^*)$  não tende a infinito quando um número  $m$  qualquer de polos de  $H(z)$  se aproxima da circunferência unitária, é quando  $H(z)$  é passa-tudo e emprega-se o algoritmo PLR.
- Para o algoritmo SHARF, apesar de  $P(z) = D(z)$  ser o ideal em termos do atendimento da condição de positividade (2.140), não é, necessariamente, o ideal em termos do condicionamento da matriz de estados (o que se reflete na velocidade de convergência local): quando  $H(z)$  é passa-tudo o condicionamento desta matriz tende a infinito

quando algum polo de  $H(z)$  se aproxima da circunferência unitária, o que não ocorre para o algoritmo PLR (que corresponde a fazer-se  $P(z) \equiv 1$ ).

- As relações obtidas entre as velocidades de crescimento ou decaimento dos valores singulares extremos da matriz de estados da ODE para os diferentes algoritmos permitem estabelecer relações entre os valores limites para estas mesmas grandezas mesmo sem especificar completamente  $H(z)$ .

O primeiro aspecto acima pode servir como uma motivação adicional para, como já comentado no início deste item, investigar as eventuais vantagens de utilizarem-se blocos passa-tudo em filtros IIR adaptativos. É importante lembrar, porém, que, como discutido no Item 2.2.4, um baixo espalhamento dos valores singulares de  $\mathbf{R}(\mathbf{w}^*)$  só indica boas propriedades locais de convergência se  $\mathbf{w}^*$  for, de fato, um ponto de convergência do algoritmo. No caso do algoritmo PLR, uma condição suficiente para tanto, como visto no Item 2.5, é a condição de positividade (2.135). Nas situações sendo consideradas aqui, onde  $H(z)$  tem polos próximos à circunferência unitária, de modo geral esta condição não é atendida, o que pode ser visto nas Figuras 2-4 e 2-3. No Item 2.7 será visto um exemplo dos problemas de convergência que podem ocorrer quando (2.135) não é atendida. Isto, porém, não torna inútil o melhor condicionamento de  $\mathbf{R}(\mathbf{w}^*)$  para o algoritmo PLR quando  $H(z)$  é passa-tudo: o resultado obtido sugere que se for empregado o algoritmo SHARF, não com  $P(z) = D(z)$  (ou  $P(z) \simeq D(z)$ , em termos práticos), mas com um  $P(z)$  o mais próximo possível de 1 que ainda atenda a condição de positividade (2.140), então a velocidade de convergência local será maior do que se for usado o algoritmo SHARF com  $P(z) = D(z)$ . Isto é de fato verificado na prática, como mostrado no item seguinte.

## 2.6.8 Exemplos numéricos do comportamento dos valores singulares

Para exemplificar o comportamento dos valores singulares de  $\mathbf{R}(\mathbf{w}^*)$  quando algum polo de  $H(z)$  se aproxima da circunferência unitária, foram consideradas as funções  $H(z)$  relacionadas na Tabela 2.2, onde, por questão de espaço, não foi indicado o argumento  $z$  das mesmas. Para maior facilidade de representação numérica são indicados na tabela os polos

	Restrição		
	$C(z)$ fixo	$\ H(z)\  = 1$	$H(z)H(z^{-1}) = 1$
RG/SMM	$\sigma_{\max} \geq \sigma_1 \xrightarrow{v_1} \infty$ $\sigma_{\min} \rightarrow k_1, m = M$ $\sigma_{\min} \leq k_1, m < M$	$\sigma_{\max} \geq \sigma_4 \xrightarrow{v_1-v_3} \infty$ $\sigma_{\min} \rightarrow k_1, m = M$ $\sigma_{\min} \xrightarrow{v_3} 0, m < M$	$\sigma_{\max} \geq \sigma_6 \xrightarrow{v_3} \infty$ $\sigma_{\min} \rightarrow k_1, m = M$ $\sigma_{\min} \leq k_1, m < M$
PLR	$\sigma_{\max} \geq \sigma_2 \xrightarrow{v_2} \infty$ $\sigma_{\min} \leq \sqrt{k_1}$	$\sigma_{\max} \geq \sigma_5 \xrightarrow{v_2-v_3} \infty$ $\sigma_{\min} \leq \sqrt{k_1}, m = M$ $\sigma_{\min} \rightarrow 0, m < M$	$\sigma_{\max} < \infty$ $\sigma_{\min} > 0$
SHARF	$\sigma_{\max} \geq \sigma_3 \xrightarrow{v_3} \infty$ $\sigma_{\min} \rightarrow 1, m = M$ $\sigma_{\min} \leq 1, m < M$	$1 \leq \sigma_{\max} \leq (1 + \sqrt{2})^2$ $0 < \sigma_{\min} \leq 1, m = M$ (*) $\sigma_{\min} \rightarrow 0$ (‡) $\sigma_{\min} \xrightarrow{v_3} 0, m < M$	$1 \leq \sigma_{\max} \leq (1 + \sqrt{2})^2$ $\sigma_{\min} \rightarrow 0$

Tabela 2.1: Comportamento dos valores singulares extremos de  $\mathbf{R}(\mathbf{w}^*)$  quando  $m$  polos de  $H(z)$  se aproximam da circunferência unitária; o símbolo  $\xrightarrow{v}$  indica comportamento dado por (2.158) com velocidade  $v$  (quando tendendo para infinito) ou  $-v$  (quando tendendo para zero); (\*):  $H(z)$  só tem polos simples; (‡):  $H(z)$  tem polos múltiplos.

Funções	Característica	Polos( $r = 0, 9; 0,99; 0,999$ ); Zeros; $K$
$H_1, H_2, H_3$	$C(z)$ fixo, $m = M$	$r \angle \pm 30^\circ; 0, 8 \angle \pm 110^\circ; K = 1, 50 \times 10^{-1}$
$H_1, H_4, H_5$	$\ H(z)\  = 1, m = M$	$r \angle \pm 30^\circ; 0, 8 \angle \pm 110^\circ$ $K = (1, 50 \times 10^{-1}; 5, 02 \times 10^{-2}; 1, 60 \times 10^{-2})$
$H_6, H_7, H_8$	passa-tudo	$r \angle \pm 30^\circ; 1/r \angle \pm 30^\circ$
$H_9, H_{10}, H_{11}$	$C(z)$ fixo, $m < M$	$(r; 0,3); 0, 8 \angle \pm 110^\circ; K = 1, 48 \times 10^{-1}$
$H_9, H_{12}, H_{13}$	$\ H(z)\  = 1, m < M$	$(r; 0,3); 0, 8 \angle \pm 110^\circ$ $K = (1, 48 \times 10^{-1}; 4, 54 \times 10^{-2}; 1, 43 \times 10^{-2})$
$H_{14}, H_{15}, H_{16}$	passa-tudo	$(r; 0,3); (1/r; 1/0,3)$

Tabela 2.2: Descrição das funções  $H(z)$  usadas para exemplificar comportamento dos valores singulares de  $\mathbf{R}(\mathbf{w}^*)$ ;  $K$  é o fator multiplicativo da função de transferência: quando na forma  $(K_1; K_2; K_3)$ , corresponde, respectivamente, aos valores 0.9, 0.99 e 0.999 de  $r$ .

e zeros de  $H(z^{-1})$ , que para os casos  $C(z)$  fixo e  $\|H(z)\| = 1$  tem então a forma

$$H(z^{-1}) = K \frac{(1 - \eta_1 z^{-1})(1 - \eta_2 z^{-1}) \dots (1 - \eta_M z^{-1})}{(1 - \nu_1 z^{-1})(1 - \nu_2 z^{-1}) \dots (1 - \nu_M z^{-1})}, \quad (2.184)$$

e para os casos passa-tudo tem a forma

$$\begin{aligned} H(z^{-1}) &= \frac{d_M + d_{M-1} z^{-1} + \dots + z^{-M}}{1 + d_1 z^{-1} + d_2 z^{-2} + \dots + d_M z^{-M}} \\ &= \frac{(z^{-1} - \nu_1)(z^{-1} - \nu_2) \dots (z^{-1} - \nu_M)}{(1 - \nu_1 z^{-1})(1 - \nu_2 z^{-1}) \dots (1 - \nu_M z^{-1})}. \end{aligned} \quad (2.185)$$

Como motivação para considerar as funções de transferência acima, apresenta-se na

	$H_1$	$H_2$	$H_3$	$H_4$	$H_5$	$H_6$	$H_7$	$H_8$
$n_{30}$	34	454	5686	345	3453	30	186	695
$n_{60}$	65	796	7952	688	6851	62	529	4135
	$H_9$	$H_{10}$	$H_{11}$	$H_{12}$	$H_{13}$	$H_{14}$	$H_{15}$	$H_{16}$
$n_{30}$	35	464	5786	346	3454	28	152	349
$n_{60}$	68	808	7961	690	6854	61	496	3758

Tabela 2.3: Número necessário de coeficientes de um filtro FIR para aproximar funções  $H(z)$  com erros  $\|H(z) - H_{FIR}(z)\|$  de  $-30$  dB e de  $-60$  dB

Tabela 2.3 o número necessário de coeficientes de um filtro FIR para aproximar, com erros  $\|H(z) - H_{FIR}(z)\|$  de  $-30$  dB e de  $-60$  dB, cada função de transferência. .

As funções  $H_1(z)$  a  $H_8(z)$  tem polos com os mesmos ângulos. Se fizermos as raízes de  $D(z)$  ter estes ângulos e o seu raio tender a 1, obtemos  $k_1$  definido por (2.166):

$$k_1 = \lim_{r \rightarrow 1} 1 / \|D(z)\|^2 = 0,2000, \quad H_1(z) \text{ a } H_8(z). \quad (2.186)$$

Da mesma maneira, para as funções  $H_9(z)$  a  $H_{16}(z)$ ,

$$k_1 = \lim_{r \rightarrow 1} 1 / \|D(z)\|^2 = 0,3597, \quad H_9(z) \text{ a } H_{16}(z). \quad (2.187)$$

Os valores singulares máximo e mínimo de  $\mathbf{R}(\mathbf{w}^*)$  para as funções  $H(z)$  relacionadas na Tabela 2.2 foram computados para os algoritmos em questão. Para o caso do algoritmo PLR foi registrada também a menor parte real dos autovalores de  $\mathbf{R}(\mathbf{w}^*)$ , fator que será usado no item seguinte. Os resultados estão apresentados nas Tabelas 2.4 e 2.5, onde pode ser observado na prática o comportamento descrito na Tabela 2.1. Em relação aos casos nos quais, na análise teórica, foi obtido apenas um limite superior ou inferior para o valor singular mínimo, alguns comentários podem ser feitos: 1) para casos nos quais  $m = M$  (todos os polos de  $H(z)$  próximos à circunferência unitária), que são os casos da Tabela 2.4, o valor de  $\sigma_{\min}$  quando os polos de  $H(z)$  se aproximam da circunferência unitária é relativamente grande, não sendo responsável por uma contribuição muito grande ao possível mau condicionamento de  $\mathbf{R}(\mathbf{w}^*)$ ; 2) para os casos nos quais  $m < M$ , que são os casos da Tabela 2.5, há duas situações distintas: para a Restrição 1 (funções  $H_9(z)$  a  $H_{11}(z)$ ) os valores de  $\sigma_{\min}$  estão consideravelmente abaixo dos limites superiores obtidos, ao passo que para a Restrição 3 (funções  $H_{14}(z)$  a  $H_{16}(z)$ ), no caso do algoritmo RG,  $\sigma_{\min}$  é relativamente próximo do limite superior, e, no caso do algoritmo PLR,  $\sigma_{\min}$  é relativamente grande. Se por



Função	Algoritmo						
	RG/SMM		PLR			SHARF	
	$\sigma_{\max}$	$\sigma_{\min}$	$\sigma_{\max}$	$\sigma_{\min}$	$\text{Re}[\lambda_{\min}]$	$\sigma_{\max}$	$\sigma_{\min}$
$H_1(z)$	$1,2 \times 10^2$	0,20	$1,1 \times 10^1$	0,29	0,53	2,0	0,09
$H_2(z)$	$8,4 \times 10^4$	0,20	$5,3 \times 10^2$	0,34	0,96	17	0,9
$H_3(z)$	$8,2 \times 10^7$	0,20	$4,9 \times 10^4$	0,34	1,00	170	1,0
$H_4(z)$	$9,5 \times 10^3$	0,20	$6,1 \times 10^1$	0,34	0,96	1,9	0,13
$H_5(z)$	$9,3 \times 10^5$	0,20	$5,5 \times 10^2$	0,34	1,00	1,9	0,13
$H_6(z)$	28	0,22	3,5	0,29	1,00	2,0	$2,9 \times 10^{-2}$
$H_7(z)$	$2,5 \times 10^2$	0,20	4,2	0,26	1,00	2,0	$2,7 \times 10^{-3}$
$H_8(z)$	$2,5 \times 10^3$	0,20	4,3	0,26	1,00	2,0	$2,7 \times 10^{-4}$

Tabela 2.4: Valores singulares extremos de  $\mathbf{R}(\mathbf{w}^*)$ .

Função	Algoritmo						
	RG/SMM		PLR			SHARF	
	$\sigma_{\max}$	$\sigma_{\min}$	$\sigma_{\max}$	$\sigma_{\min}$	$\text{Re}[\lambda_{\min}]$	$\sigma_{\max}$	$\sigma_{\min}$
$H_9(z)$	$2,4 \times 10^2$	$1,6 \times 10^{-2}$	$1,8 \times 10^1$	$1,2 \times 10^{-2}$	$1,2 \times 10^{-2}$	2,1	$8,3 \times 10^{-3}$
$H_{10}(z)$	$2,2 \times 10^5$	$2,1 \times 10^{-2}$	$1,5 \times 10^3$	$1,4 \times 10^{-2}$	$1,4 \times 10^{-2}$	21	$9,4 \times 10^{-3}$
$H_{11}(z)$	$2,2 \times 10^8$	$2,1 \times 10^{-2}$	$1,5 \times 10^5$	$1,4 \times 10^{-2}$	$1,5 \times 10^{-2}$	210	$9,5 \times 10^{-3}$
$H_{12}(z)$	$2,0 \times 10^4$	$2,1 \times 10^{-3}$	$1,4 \times 10^2$	$1,4 \times 10^{-3}$	$1,4 \times 10^{-3}$	2,0	$8,9 \times 10^{-4}$
$H_{13}(z)$	$2,0 \times 10^6$	$2,2 \times 10^{-4}$	$1,4 \times 10^3$	$1,4 \times 10^{-4}$	$1,4 \times 10^{-4}$	2,0	$8,9 \times 10^{-5}$
$H_{14}(z)$	$2,9 \times 10^1$	0,28	2,5	0,39	1,00	1,0	$4,8 \times 10^{-2}$
$H_{15}(z)$	$3,0 \times 10^2$	0,26	2,7	0,37	1,00	2,0	$4,5 \times 10^{-3}$
$H_{16}(z)$	$3,1 \times 10^3$	0,26	2,7	0,37	1,00	2,0	$4,5 \times 10^{-4}$

Tabela 2.5: Valores singulares extremos de  $\mathbf{R}(\mathbf{w}^*)$ .

um lado a análise teórica realizada sobre o comportamento dos valores singulares de  $\mathbf{R}(\mathbf{w}^*)$  foi suficiente para estabelecer os casos em que o espalhamento dos mesmos tende a infinito, uma análise mais aprofundada poderia, eventualmente, permitir uma melhor compreensão destes aspectos adicionais comentados.

Para verificar a lei de crescimento dos valores singulares dada por (2.157), foram feitas regressões lineares sobre os logaritmos dos valores singulares calculados, obtendo, para cada caso, valores da velocidade  $v_{\max}$  de crescimento do valor singular máximo de  $\mathbf{R}(\mathbf{w}^*)$ . Os valores obtidos estão na Tabela 2.6. Pode-se observar uma boa concordância das relações verificadas entre a velocidade de crescimento em cada caso com as relações previstas pelo modelamento utilizado, apresentadas na Tabela 2.1.

Finalmente, para facilitar análises adicionais sobre os resultados, o espalhamento dos valores singulares de  $\mathbf{R}(\mathbf{w}^*)$  para cada caso,  $\sigma_{\max}/\sigma_{\min}$ , já está computado nas Tabelas 2.7 e 2.8.

Funções	$u_{\max}$		
	Algoritmo		
	RG/SMM	PLR	SHARF
$H_1(z), H_2(z), H_3(z)$	2,9	1,8	0,96
$H_1(z), H_4(z), H_5(z)$	1,9	0,85	—
$H_6(z), H_7(z), H_8(z)$	0,98	—	—
$H_9(z), H_{10}(z), H_{11}(z)$	3,0	2,0	1,0
$H_9(z), H_{12}(z), H_{13}(z)$	2,0	0,95	—
$H_{14}(z), H_{15}(z), H_{16}(z)$	1	—	—

Tabela 2.6: Velocidade de crescimento de  $\sigma_{\max}$ .

Função	$\sigma_{\max}/\sigma_{\min}$		
	Algoritmo		
	RG/SMM	PLR	SHARF
$H_1(z)$	$6,0 \times 10^2$	$3,8 \times 10^1$	$2,2 \times 10^1$
$H_2(z)$	$4,2 \times 10^5$	$1,6 \times 10^3$	$1,9 \times 10^1$
$H_3(z)$	$4,1 \times 10^8$	$1,4 \times 10^5$	$1,7 \times 10^2$
$H_4(z)$	$4,8 \times 10^4$	$1,8 \times 10^2$	$1,5 \times 10^1$
$H_5(z)$	$4,7 \times 10^6$	$1,6 \times 10^3$	$1,5 \times 10^1$
$H_6(z)$	$1,3 \times 10^2$	$1,2 \times 10^1$	$6,9 \times 10^1$
$H_7(z)$	$1,3 \times 10^3$	$1,6 \times 10^1$	$7,4 \times 10^2$
$H_8(z)$	$1,3 \times 10^4$	$1,6 \times 10^1$	$7,4 \times 10^3$

Tabela 2.7: Espalhamento dos valores singulares de  $\mathbf{R}(\mathbf{w}^*)$ .

Função	$\sigma_{\max}/\sigma_{\min}$		
	Algoritmo		
	RG/SMM	PLR	SHARF
$H_9(z)$	$1,5 \times 10^4$	$1,5 \times 10^3$	$2,5 \times 10^2$
$H_{10}(z)$	$1,1 \times 10^7$	$1,1 \times 10^5$	$2,2 \times 10^3$
$H_{11}(z)$	$1,1 \times 10^{10}$	$1,1 \times 10^7$	$2,2 \times 10^4$
$H_{12}(z)$	$9,5 \times 10^6$	$1,0 \times 10^5$	$2,2 \times 10^3$
$H_{13}(z)$	$9,1 \times 10^9$	$1,0 \times 10^7$	$2,2 \times 10^4$
$H_{14}(z)$	$1,0 \times 10^2$	6,4	$4,2 \times 10^1$
$H_{15}(z)$	$1,2 \times 10^3$	7,3	$4,4 \times 10^2$
$H_{16}(z)$	$1,2 \times 10^4$	7,3	$4,4 \times 10^3$

Tabela 2.8: Espalhamento dos valores singulares de  $\mathbf{R}(\mathbf{w}^*)$ .

## 2.7 Exemplos da simulação dos algoritmos de adaptação

Neste item são verificados na prática aspectos analisados nos itens anteriores deste capítulo, relativos ao comportamento local dos algoritmos para filtros IIR adaptativos. Estes aspectos são:

- relação entre o espalhamento dos valores singulares de  $\mathbf{R}(\mathbf{w}^*)$ , a velocidade de convergência local e o tempo de convergência (ver (2.69));
- variação do valor máximo de  $\alpha = \mu\sigma_{\max}$  (ver (2.68)) conforme o sistema  $H(z)$  e o algoritmo de adaptação;
- validade da utilização de modos normais para caracterizar a convergência do algoritmo (ver (2.61));
- efeitos do não atendimento das condições de positividade (2.135) e (2.140) e da não positividade de  $\mathbf{R}(\mathbf{w}^*)$ .

Os algoritmos RG, SMM, PLR e SHARF para filtros IIR adaptativos, analisados neste capítulo, foram implementados em MatLab, e em seguida convertidos para a linguagem C usando o pacote *C Compiler* da MathWorks, Inc. A transposição para C é bastante conveniente no caso geral da implementação de algoritmos adaptativos no MatLab, pois a estrutura de malha destes programas é um dos casos típicos nos quais o MatLab é pouco eficiente. O ganho em velocidade de execução com a conversão dos programas para C foi da ordem de 15 vezes.

Os algoritmos de adaptação foram testados na identificação das funções  $H_1(z)$ ,  $H_4(z)$ ,  $H_6(z)$ ,  $H_7(z)$ ,  $H_9(z)$ ,  $H_{12}(z)$ ,  $H_{14}(z)$ ,  $H_{15}(z)$  definidas na Tabela 2.2 do item anterior, para o caso de ruído de medida nulo. Uma vez que estão sendo analisados os aspectos locais da convergência, foram escolhidos como pontos iniciais da adaptação funções  $\hat{H}(z)$  relativamente próximas de  $H(z)$ , apresentadas na Tabela 2.9.

Quando não indicado o contrário, usou-se uma entrada  $u(n)$  branca de potência unitária, sendo que todos os casos foram testados com a mesma realização de  $u(n)$ . Por uma questão de tempo de computação, os resultados levantados referem-se a uma única realização de  $u(n)$ , obtida com os comandos `randn('seed',0)` e `randn(1,N)` do MatLab. Isto é suficiente no

Função	$\hat{H}(z)$ inicial
$H_1(z), H_4(z)$	polos: $0,9\angle \pm 10^\circ$ ; zeros: $0,75\angle \pm 95^\circ$ ; $\ \hat{H}(z)\  = 1$
$H_6(z), H_7(z)$	polos: $0,9\angle \pm 10^\circ$ ; zeros: $1/0,9\angle \mp 10^\circ$ (passa-tudo)
$H_9(z), H_{12}(z)$	polos: $0,8; 0,4$ ; zeros: $0,75\angle \pm 95^\circ$ ; $\ \hat{H}(z)\  = 1$
$H_{14}(z), H_{15}(z)$	polos: $0,8; 0,4$ ; zeros: $1/0,8; 1/0,4$ (passa-tudo)

Tabela 2.9: Valores iniciais de  $\hat{H}(z)$

caso das ordens de grandeza das variações das características que estão em questão aqui. Se fosse desejada uma maior precisão na medição destas características seria necessário fazer uma média de mais realizações de  $u(n)$ .

Os resultados do algoritmo SMM, como esperado devido à proximidade do ponto inicial ao valor correto dos parâmetros, foram semelhantes aos do algoritmo RG, portanto apenas estes últimos são apresentados. O algoritmo SHARF, quando não indicado o contrário, foi testado com  $P(z) = D(z)$ . Esta escolha de  $P(z)$  é uma idealização do que é feito na prática, uma vez que  $D(z)$  não é conhecido, mas serve como indicação do que ocorre quando  $P(z) \simeq D(z)$ .

Foi usado, em todos os algoritmos, um controle para assegurar que os polos estivessem sempre dentro do domínio de estabilidade, o que no caso de sistemas de até segunda ordem é feito de maneira simples.

Para cada caso de  $H(z)$  e de algoritmo de adaptação o passo de adaptação  $\mu$  foi progressivamente incrementado até se obter um valor além do qual um pequeno incremento adicional ou levava o algoritmo a se tornar instável, isto é, a não convergir mais para algum valor qualquer de parâmetros, ou o levava de tal modo perto da instabilidade que o tempo de convergência aumentava ao invés de diminuir. Os valores de  $\mu$  assim obtidos e o número necessário de iterações, em cada caso, para obter um erro quadrático de saída  $(y(n) - \hat{y}(n))^2$  60 dB abaixo da potência de saída  $E[y^2(n)]$ , denotado  $n_{60}$ , estão apresentados na Tabela 2.10. Para maior conveniência,  $n_{60}$  será referido doravante como "tempo de convergência". Em relação às funções  $H(z)$  que constam desta tabela, como pode ser visto na Tabela 2.2, as funções  $H_1(z^{-1})$  e  $H_4(z^{-1})$  tem norma  $L_2$  unitária, e um par de polos complexos de ângulo  $30^\circ$  e raio  $0,9$  e  $0,99$ , respectivamente; as funções  $H_6(z^{-1})$  e  $H_7(z^{-1})$  são passa-tudo e tem os mesmos polos de  $H_1(z^{-1})$  e  $H_4(z^{-1})$ , respectivamente; as funções  $H_9(z^{-1})$  e  $H_{12}(z^{-1})$  tem norma  $L_2$  unitária e dois polos reais, um em  $0,3$  e o outro em  $0,9$  e  $0,99$ , respectivamente; as funções  $H_{14}(z^{-1})$  e  $H_{15}(z^{-1})$  são passa-tudo e tem os mesmos polos de  $H_9(z^{-1})$  e  $H_{12}(z^{-1})$ ,

Função	Algoritmo					
	RG		PLR		SHARF	
	$\mu$	$n_{60}$	$\mu$	$n_{60}$	$\mu$	$n_{60}$
$H_1(z)$	$1,5 \times 10^{-3}$	$1,5 \times 10^4$	$1,5 \times 10^{-2}$	$1,8 \times 10^3$	$4 \times 10^{-2}$	$1,5 \times 10^3$
$H_4(z)$	$1,5 \times 10^{-5}$	$1,5 \times 10^6$	$3 \times 10^{-4}$	$8,3 \times 10^4$	$1,5 \times 10^{-2}$	$4,5 \times 10^3$
$H_6(z)$	$8 \times 10^{-3}$	$3,0 \times 10^3$	$3 \times 10^{-2}$	$1,6 \times 10^3$	$3 \times 10^{-2}$	$7,0 \times 10^3$
$H_7(z)$	$1,5 \times 10^{-4}$	$1,9 \times 10^5$	$2 \times 10^{-3}$	$6,9 \times 10^5$	$8 \times 10^{-3}$	$3 \times 10^5$
$H_9(z)$	$1,5 \times 10^{-3}$	$1,8 \times 10^5$	$5,0 \times 10^{-2}$	$8,5 \times 10^3$	$4 \times 10^{-2}$	$1,2 \times 10^4$
$H_{12}(z)$	(*)	(*)	(*)	(*)	$1,5 \times 10^{-2}$	$3 \times 10^5$
$H_{14}(z)$	$7,0 \times 10^{-3}$	$2,2 \times 10^3$	$3 \times 10^{-2}$	$3,6 \times 10^2$	$5 \times 10^{-2}$	$2,0 \times 10^3$
$H_{15}(z)$	$1,0 \times 10^{-4}$	$1,9 \times 10^5$	$8 \times 10^{-3}$	$1,8 \times 10^4$	$1,5 \times 10^{-2}$	$1,8 \times 10^5$

Tabela 2.10: Resultados da execução dos algoritmos RG, PLR e SHARF;  $\mu$ : passo de adaptação máximo;  $n_{60}$ : número de iterações para erro quadrático de saída ficar 60 dB abaixo da potência de saída; (\*): execução não completada devido ao longo tempo de convergência ( $> 10^6$ );

Função	Número da figura		
	RG	PLR	SHARF
$H_1(z)$	2-5	2-7	2-10
$H_4(z)$	2-6	2-8	2-11
$H_6(z)$	2-12	2-14	2-17
$H_7(z)$	2-13	2-15, 2-16	2-18
$H_9(z)$	2-19	2-20	2-21
$H_{12}(z)$	-	-	2-22
$H_{14}(z)$	2-23	2-25	2-27
$H_{15}(z)$	2-24	2-26	2-28

Tabela 2.11: Número da figura com gráficos da execução dos algoritmos para cada caso de  $H(z)$

respectivamente.

São apresentados também os gráficos com a adaptação dos parâmetros e a evolução do erro quadrático  $(y(n) - \hat{y}(n))^2$ . O número da figura para cada caso está na Tabela 2.11.

**Observação:** para poder-se extrair mais informação dos gráficos com a adaptação dos parâmetros foi adotado um fator de escala específico para cada caso, cujo valor é indicado no próprio gráfico. Para obter-se o valor correto do erro de cada parâmetro em um dado instante deve-se subtrair o valor lido naquele instante do valor correto do parâmetro, indicado por um traço mais claro, e dividir o resultado pelo fator de escala.

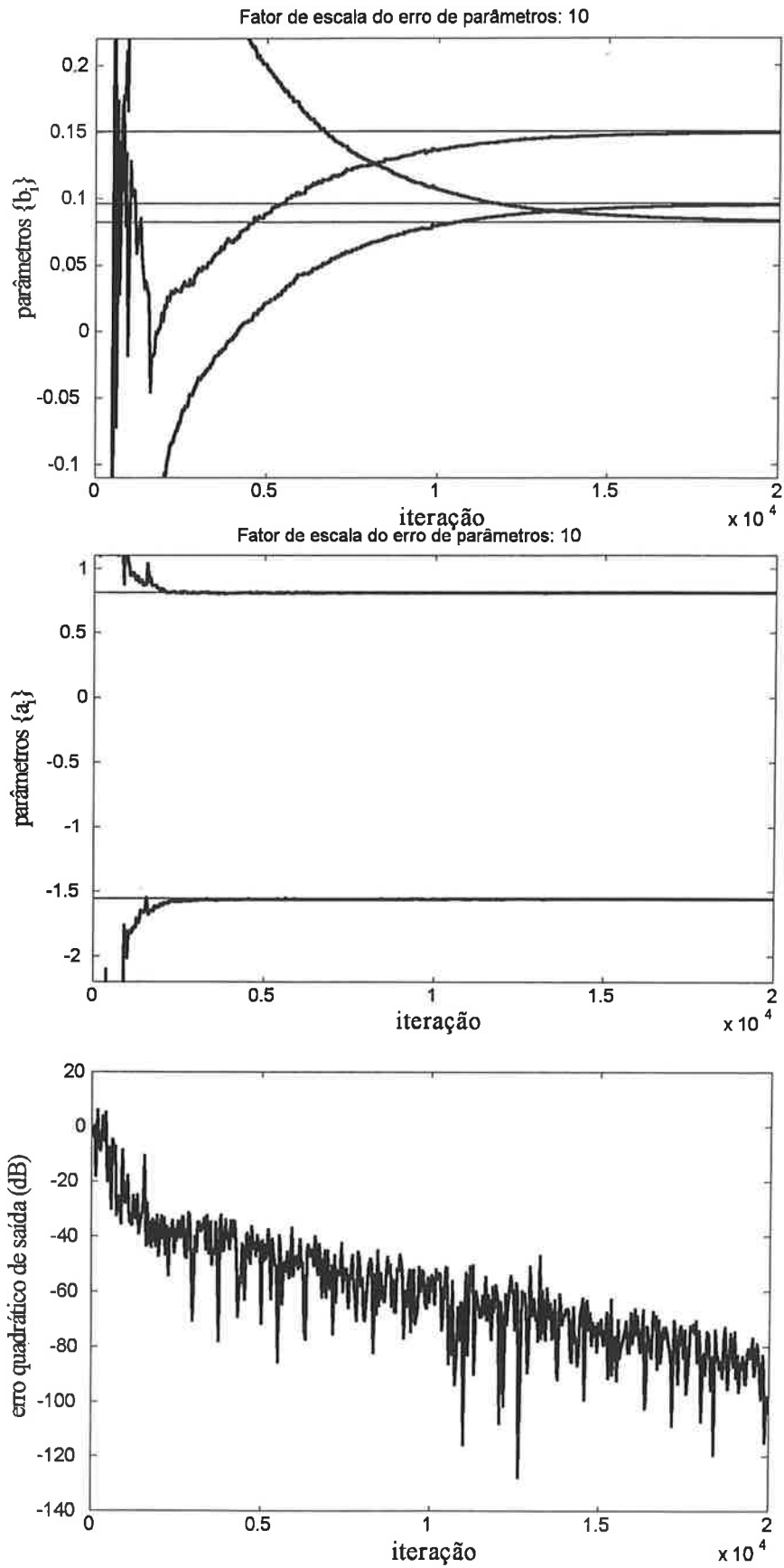


Figura 2-5: Algoritmo RG, sistema  $H_1(z)$  (parâmetros em traços claros)

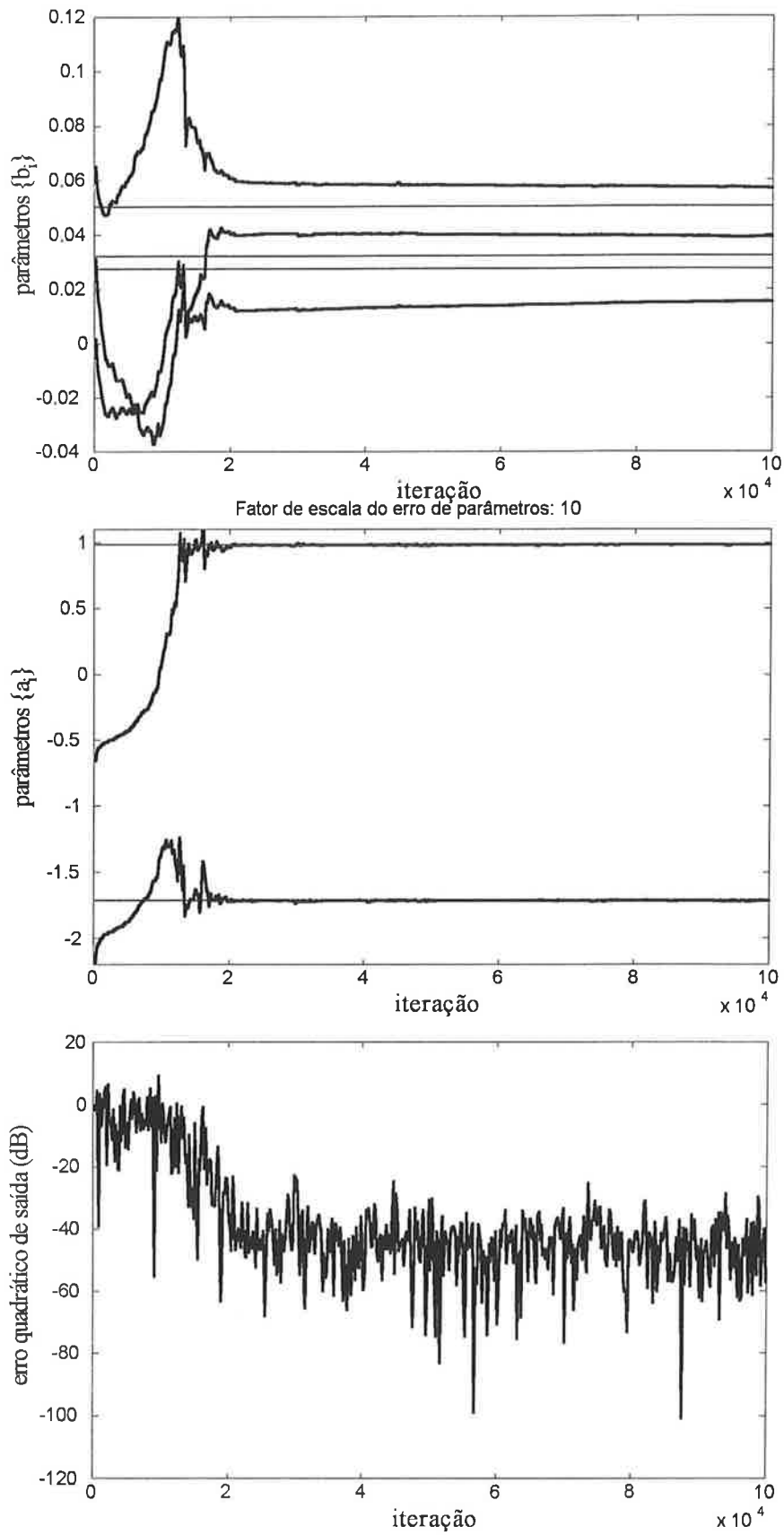


Figura 2-6: Algoritmo RG, sistema  $H_4(z)$  (parâmetros em traços claros)

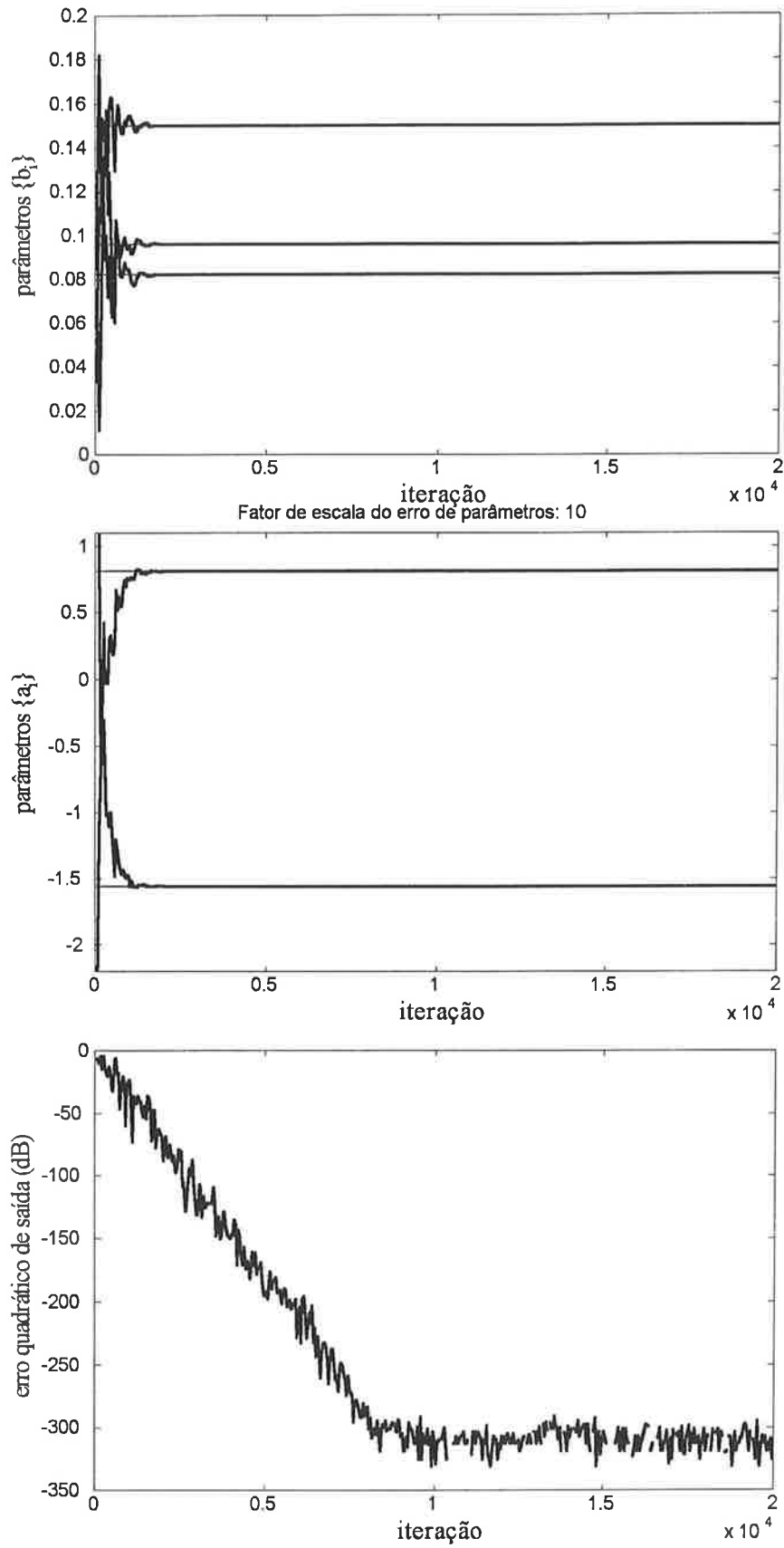


Figura 2-7: Algoritmo PLR, sistema  $H_1(z)$  (parâmetros em traços claros)



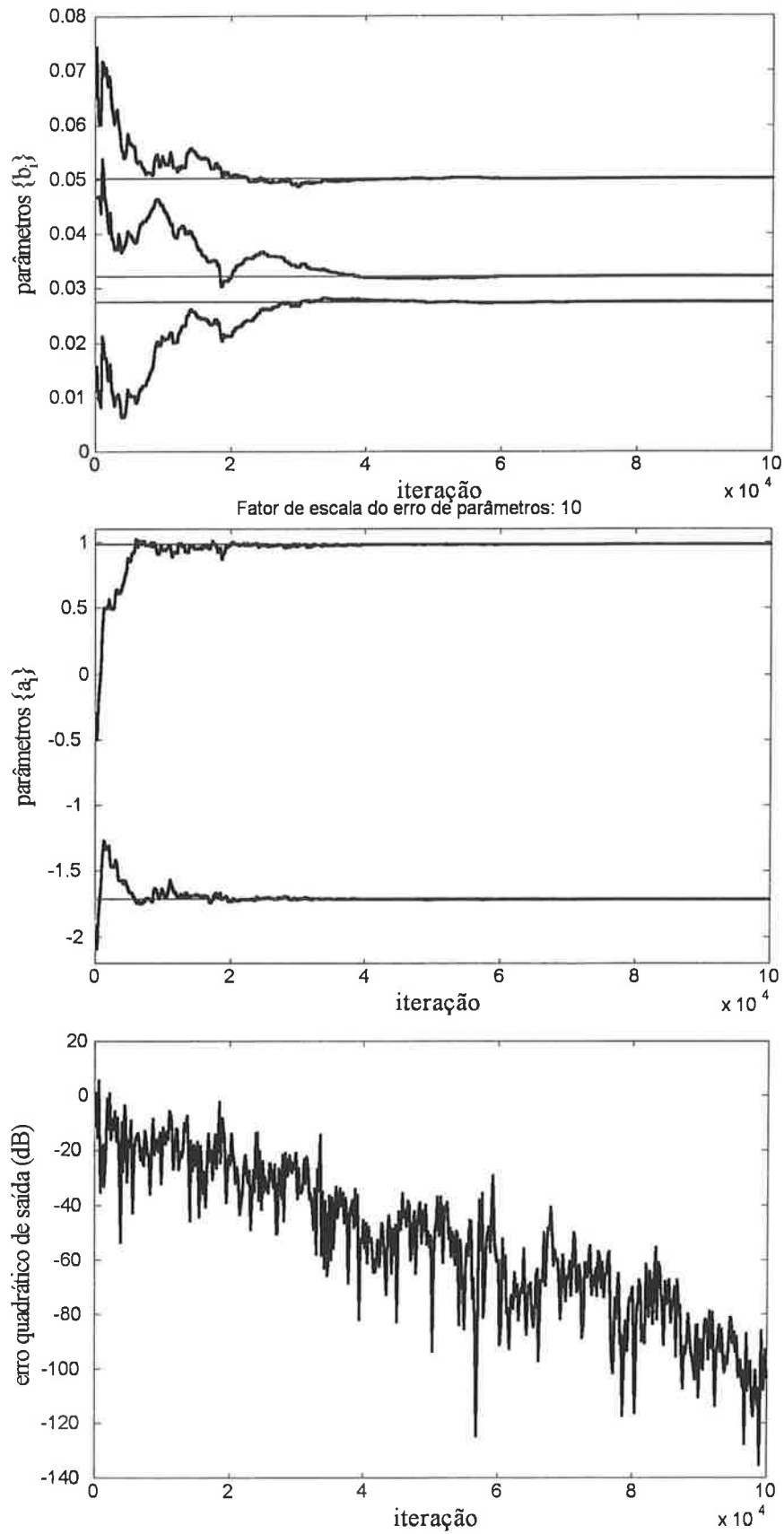


Figura 2-8: Algoritmo PLR, sistema  $H_4(z)$  (parâmetros em traços claros)

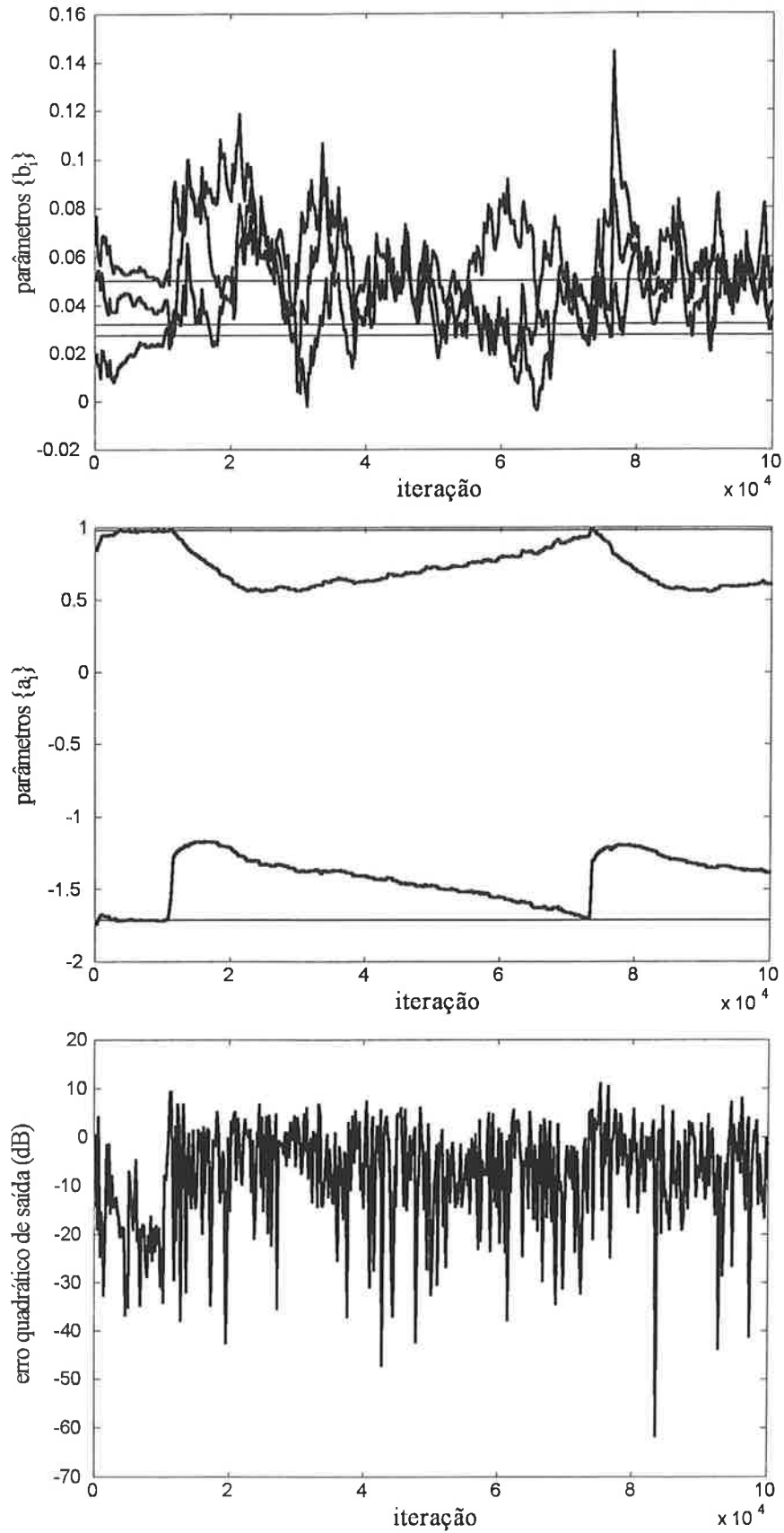


Figura 2-9: Algoritmo PLR,  $\mu = 0,0005$ ; sistema  $H_4(z)$  (parâmetros em traços claros)

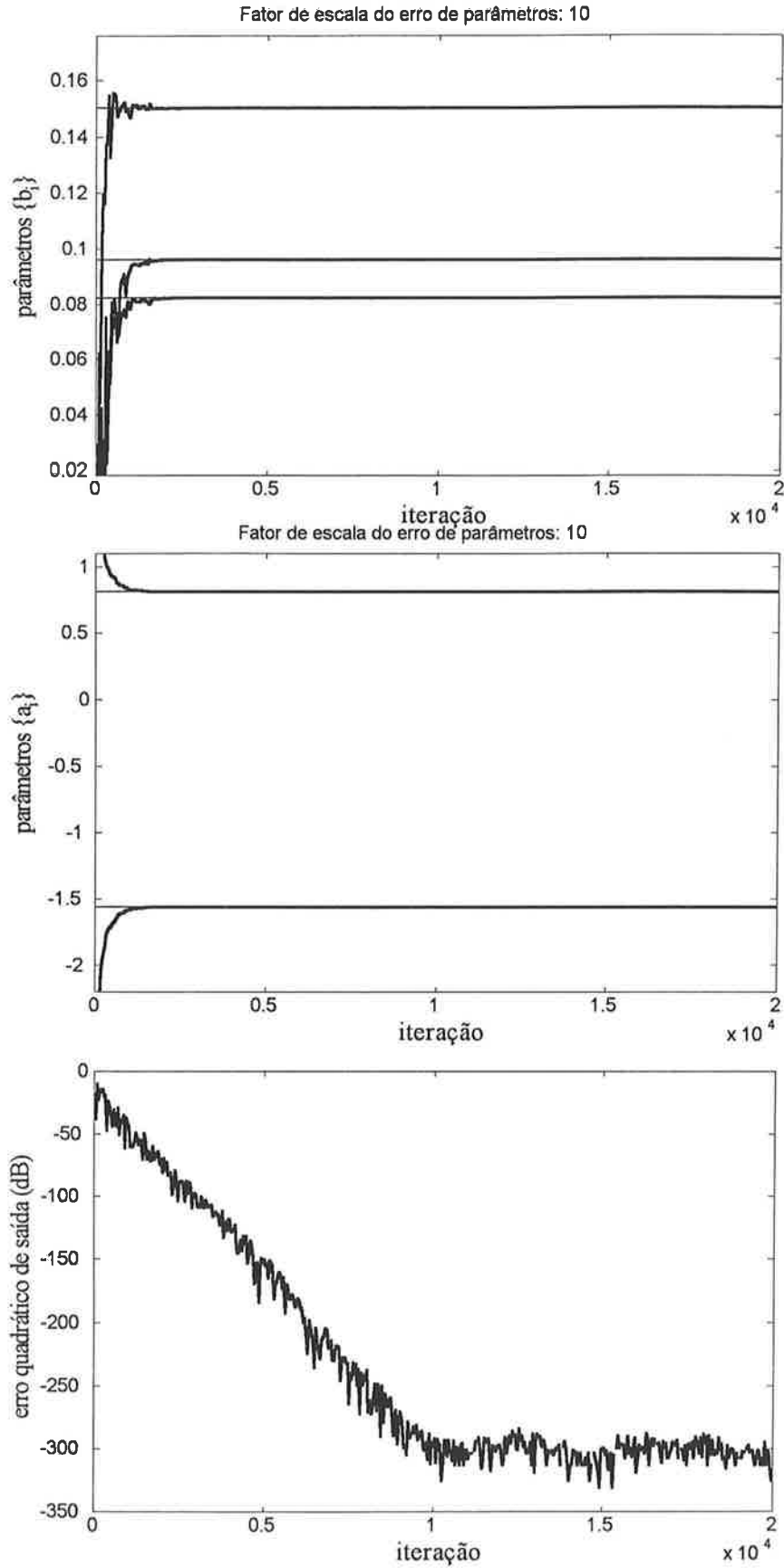


Figura 2-10: Algoritmo SHARF, sistema  $H_1(z)$  (parâmetros em traços claros)

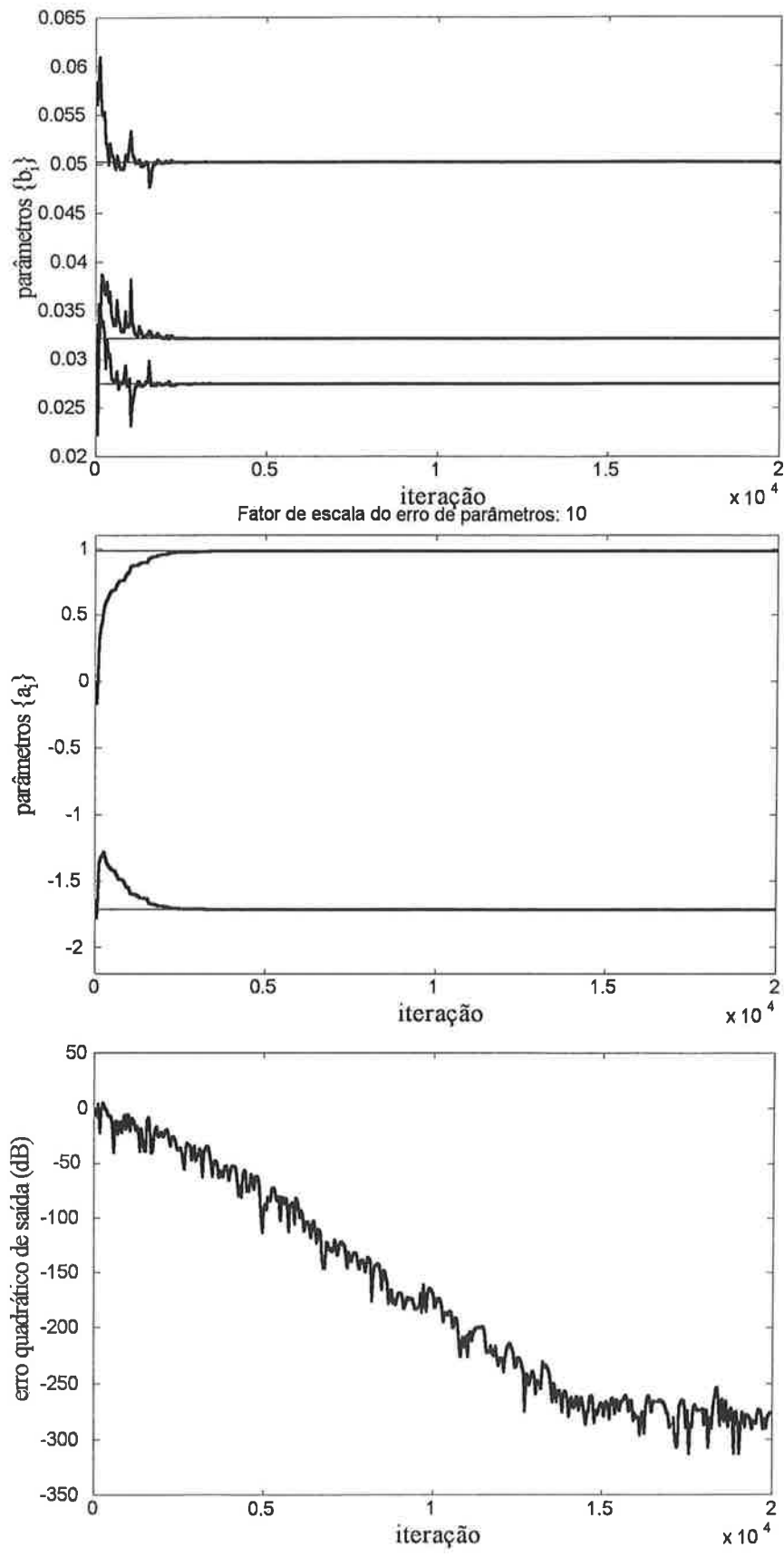


Figura 2-11: Algoritmo SHARF, sistema  $H_4(z)$  (parâmetros em traços claros)

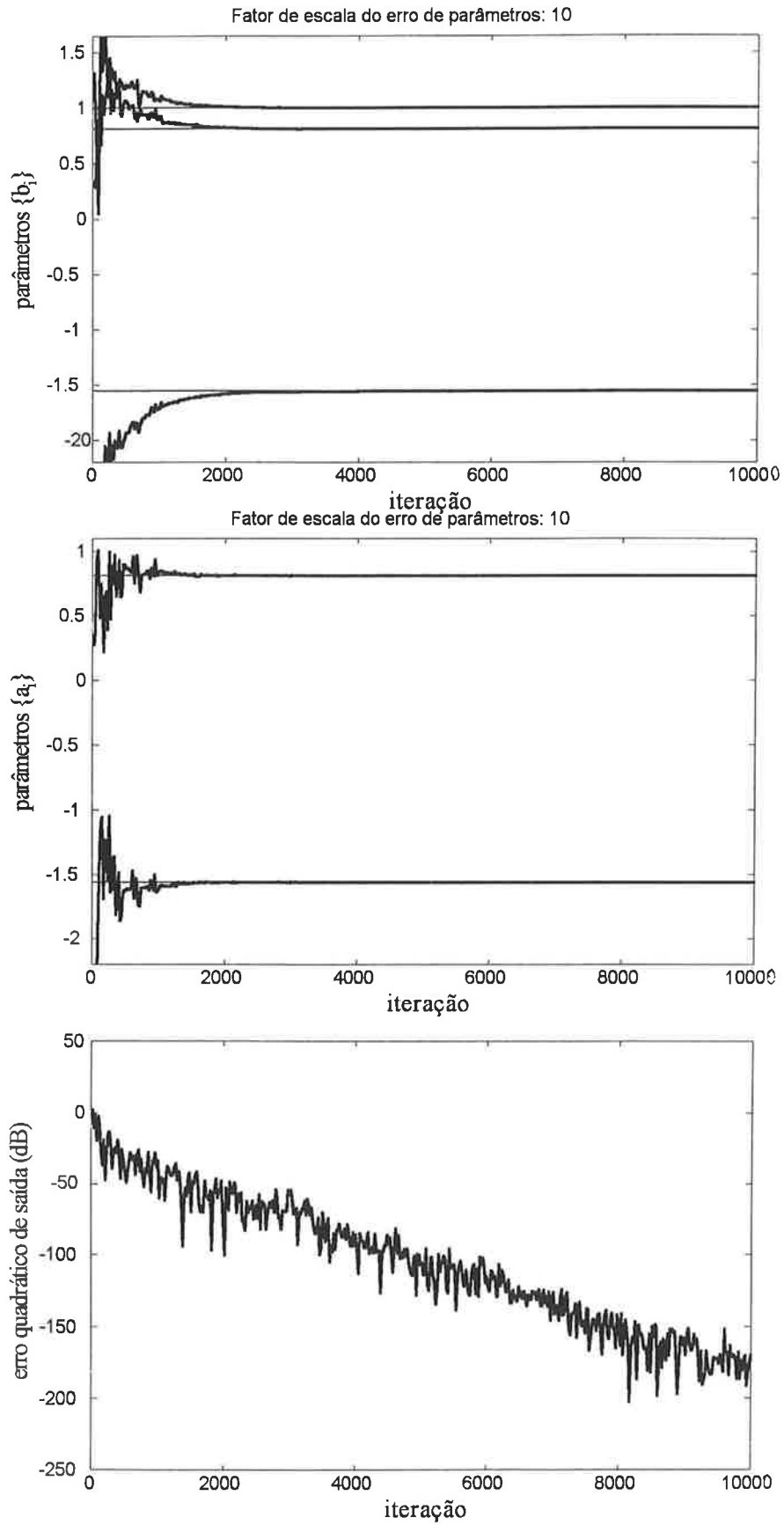


Figura 2-12: Algoritmo RG, sistema  $H_6(z)$  (parâmetros em traços claros)

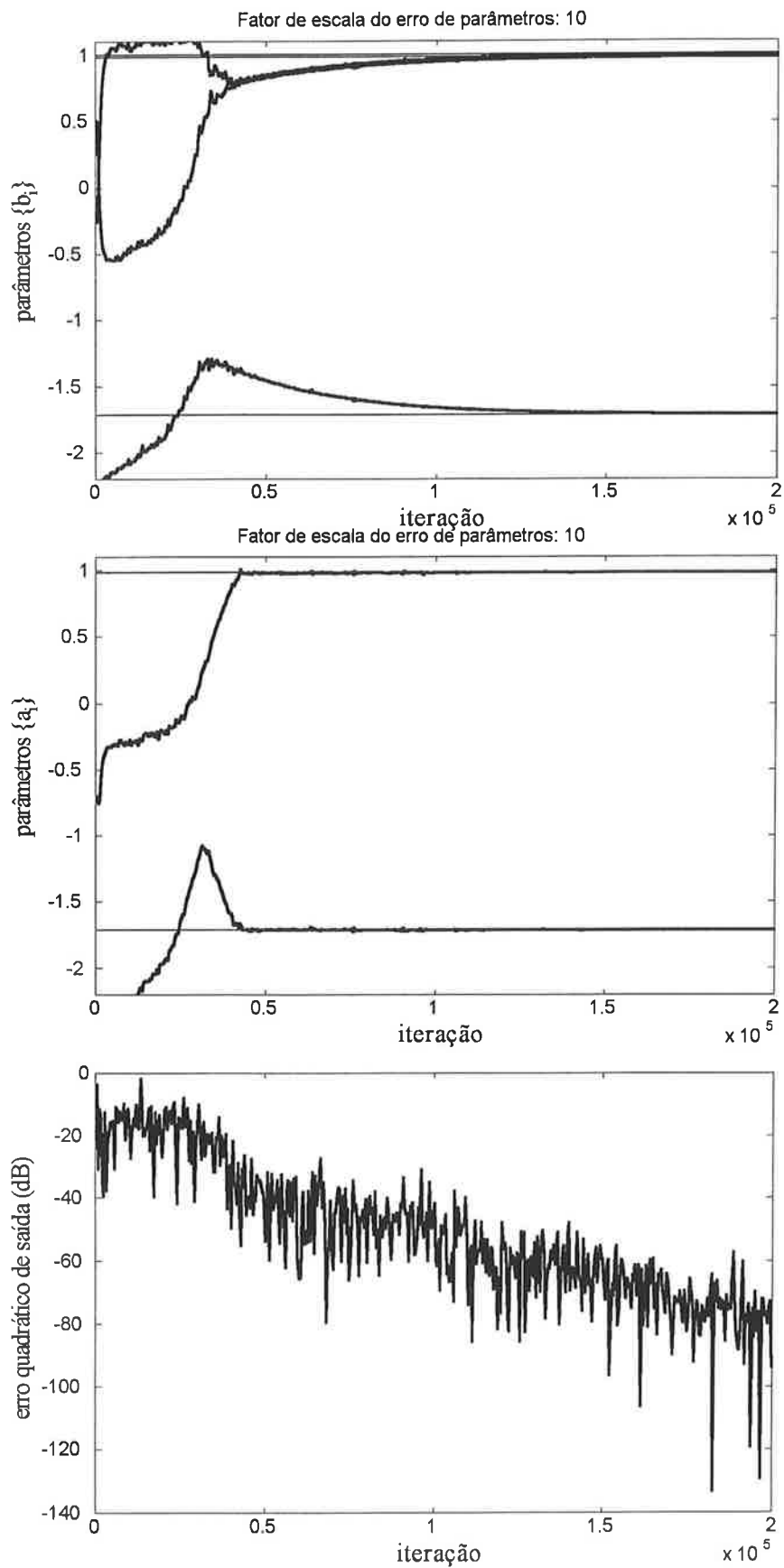


Figura 2-13: Algoritmo RG, sistema  $H_7(z)$  (parâmetros em traços claros)

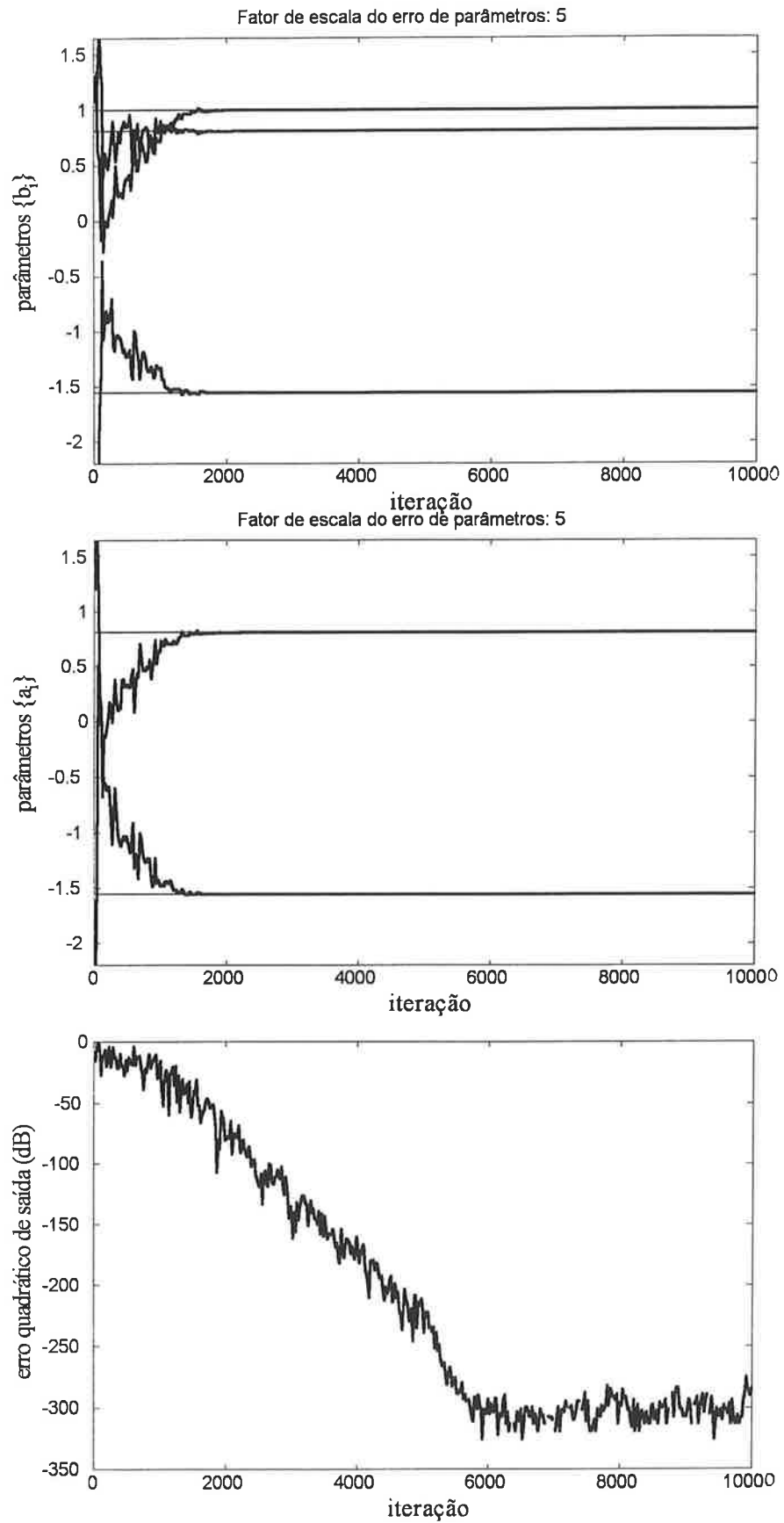


Figura 2-14: Algoritmo PLR, sistema  $H_6(z)$  (parâmetros em traços claros)

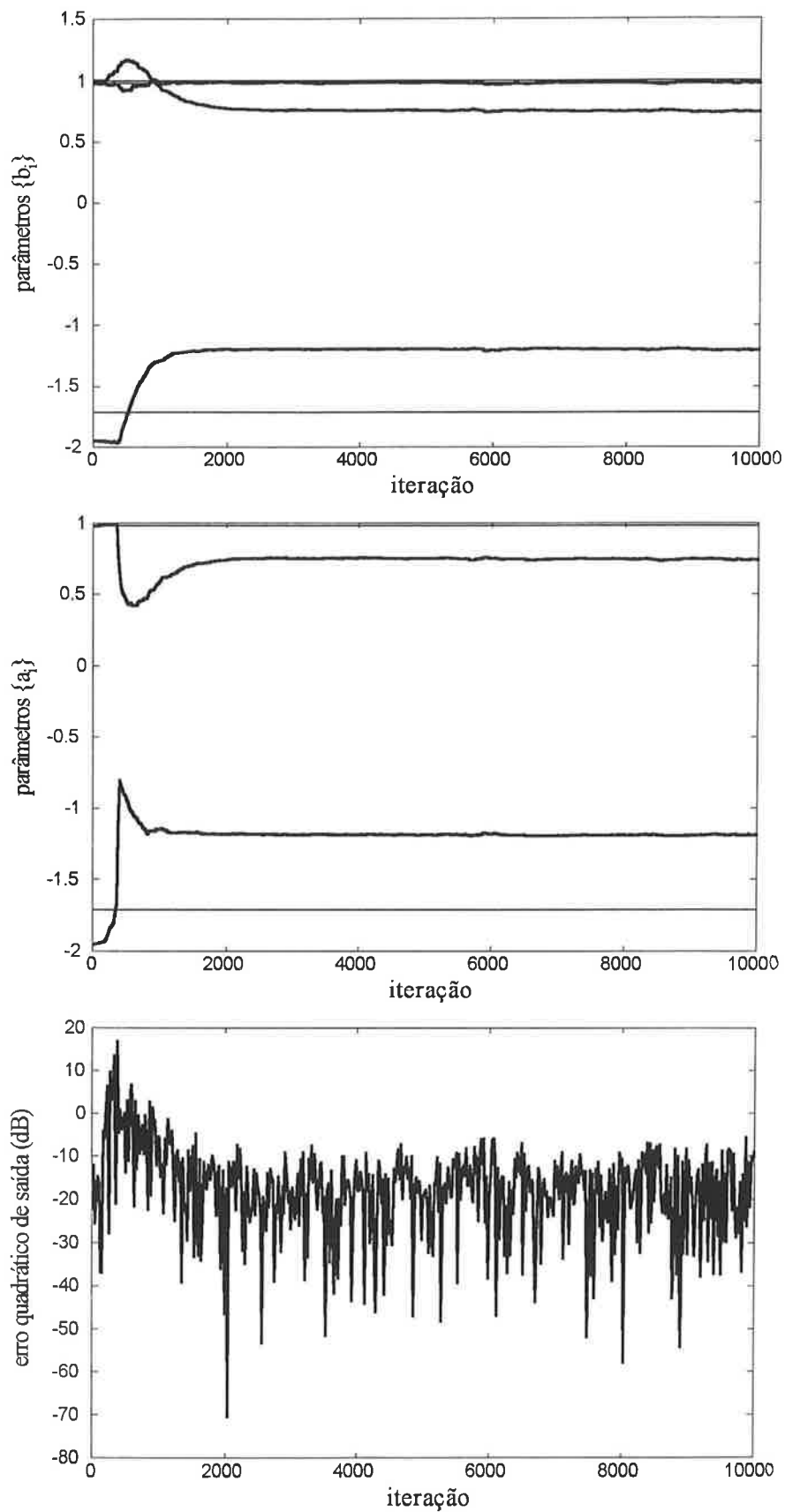


Figura 2-15: Algoritmo PLR, sistema  $H_7(z)$  (parâmetros em traços claros)



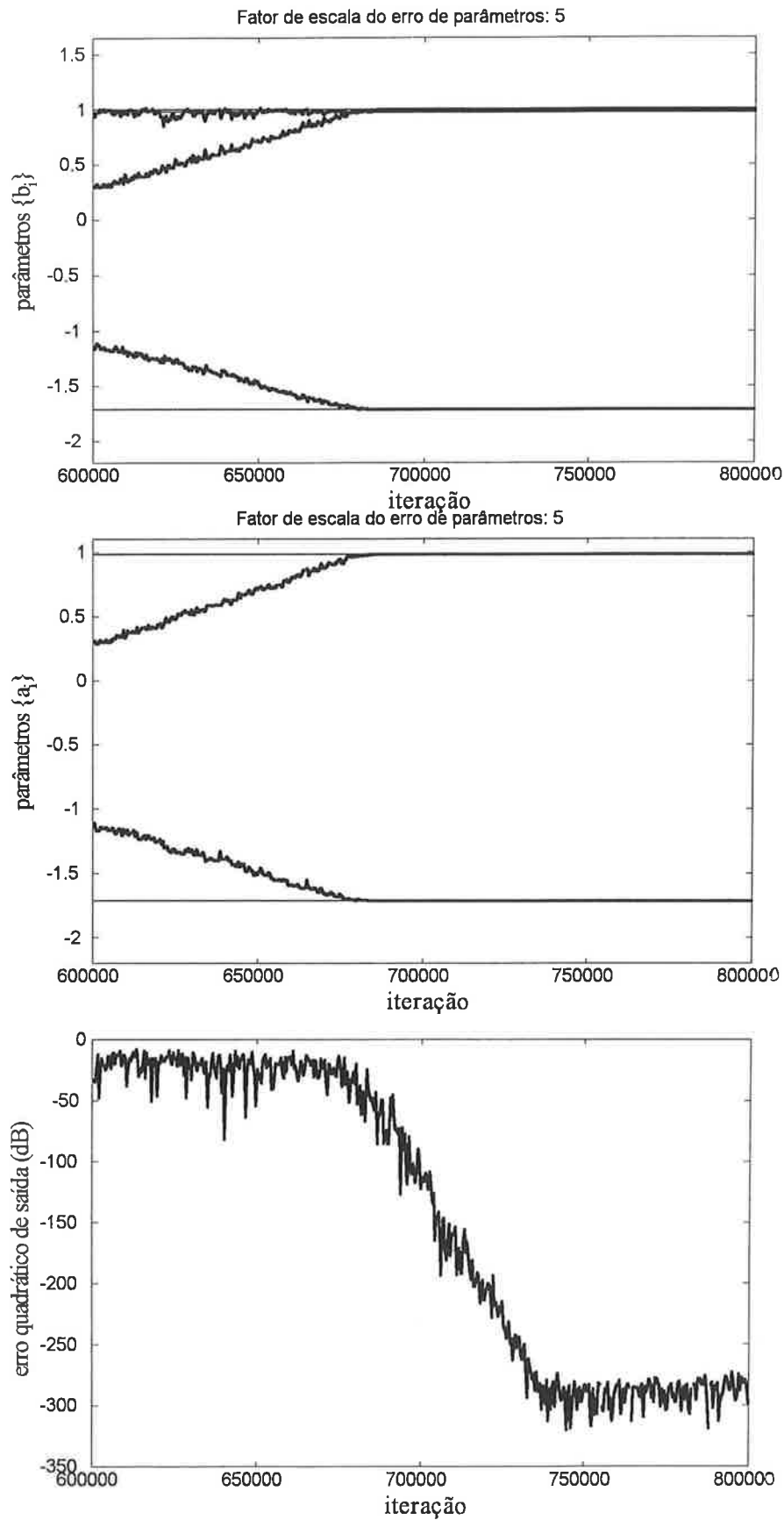


Figura 2-16: Algoritmo PLR, sistema  $H_7(z)$  (parâmetros em traços claros)

### 2.7.1 Verificação da análise local

Antes de proceder a comentários sobre os resultados das simulações dos algoritmos apresentados na Tabela 2.10, é de interesse relacionar estes resultados com a análise sobre convergência local feita no Item 2.2.4. Para tanto, duas grandezas foram computadas para cada caso da Tabela 2.10, como descrito a seguir.

Para verificar a relação entre  $\sigma_{\max}$ , o valor singular máximo de  $\mathbf{R}(\mathbf{w}^*)$  (a matriz de estados da equação diferencial associada ao algoritmo), e o valor máximo do passo de adaptação, foi computado o produto  $\mu\sigma_{\max} = \alpha$ . O papel de  $\alpha$  na velocidade de convergência local pode ser visto em (2.69).

Para verificar a relação entre  $\mu$ , a menor parte real dos autovalores de  $\mathbf{R}(\mathbf{w}^*)$  (denotada  $\text{Re}[\lambda_m]$ ) e a velocidade de convergência local foi computado o produto

$$\bar{\delta} \triangleq 20 \log(e)\mu \text{Re}[\lambda_m] \quad (2.188)$$

( para os algoritmos RG e SHARF com  $P(z) = D(z)$  temos que  $\lambda_m$  é real e igual ao menor valor singular de  $\mathbf{R}(\mathbf{w}^*)$ ,  $\sigma_{\min}$ ). Conforme (2.71), quando, para  $\mu$  suficientemente pequeno, valer a aproximação de (2.61) por (2.56),  $\bar{\delta}$  fornece a velocidade de convergência local. Esta velocidade foi também medida a partir dos resultados obtidos em cada execução: para tanto foi feita uma regressão linear de  $10 \log[e^2(n)]$  com  $n$ , tomando como ponto inicial o valor de  $n$  a partir do qual o decaimento do erro não mudava mais de inclinação e (se necessário), como ponto final, o valor de  $n$  a partir do qual o limite de precisão numérica do MatLab era atingido.

Nas Tabelas 2.12 e 2.13 estão os valores de  $\mu\sigma_{\max}$ ,  $\bar{\delta}$  e  $\delta$ . Podemos ver que, tomando todos os casos apresentados,  $\mu\sigma_{\max}$  está entre 238 e 2,2 vezes abaixo do limite superior  $\mu\sigma_{\max} = 2$  dado por (2.53). Dada esta grande faixa de variação, vemos que, como comentado em seguida à expressão (2.69), há interesse em uma análise teórica mais aprofundada sobre este aspecto. Em relação a  $\bar{\delta}$  e  $\delta$ , podemos ver que para os algoritmos RG e SHARF há uma concordância satisfatória entre ambas. Isto indica que o valor de  $\mu$  que, em cada caso, permitiu o menor tempo de convergência era pequeno o suficiente para valer a aproximação por modos normais feita no Item 2.2.4. Para o algoritmo PLR, porém, isto não é verdade, o que exemplifica o efeito da não-simetria da matriz  $\mathbf{R}(\mathbf{w}^*)$ , também discutido no Item 2.2.4. Esta questão será analisada ainda a seguir, juntamente com comentários específicos sobre cada caso dos

Função	Algoritmo					
	RG			SHARF		
	$\mu\sigma_{\max}$	$\bar{\delta}$	$\delta$	$\mu\sigma_{\max}$	$\bar{\delta}$	$\delta$
$H_1(z)$	0,18	$2,6 \times 10^{-3}$	$2,7 \times 10^{-3}$	0,080	$3,1 \times 10^{-2}$	$2,9 \times 10^{-2}$
$H_4(z)$	0,14	$2,6 \times 10^{-5}$	(**)	0,029	$1,7 \times 10^{-2}$	$2,0 \times 10^{-2}$
$H_6(z)$	0,22	$1,5 \times 10^{-2}$	$1,6 \times 10^{-2}$	0,060	$7,6 \times 10^{-3}$	$7,3 \times 10^{-3}$
$H_7(z)$	0,038	$2,6 \times 10^{-4}$	$2,7 \times 10^{-4}$	0,016	$1,9 \times 10^{-4}$	$1,5 \times 10^{-4}$
$H_9(z)$	0,36	$2,1 \times 10^{-4}$	$2,0 \times 10^{-4}$	0,084	$2,9 \times 10^{-3}$	$2,7 \times 10^{-3}$
$H_{12}(z)$	(*)	(*)	(*)	0,03	$1,2 \times 10^{-4}$	$1,1 \times 10^{-4}$
$H_{14}(z)$	0,20	$1,7 \times 10^{-2}$	$1,9 \times 10^{-2}$	0,10	$2,1 \times 10^{-2}$	$2,0 \times 10^{-2}$
$H_{15}(z)$	0,030	$2,3 \times 10^{-4}$	$2,4 \times 10^{-4}$	0,030	$5,9 \times 10^{-4}$	$5,6 \times 10^{-4}$

Tabela 2.12: Valores de  $\mu\sigma_{\max}$ ,  $\bar{\delta}$  e  $\delta$  obtidos das execuções dos algoritmos RG e SHARF ( $\delta$  e  $\bar{\delta}$  em db/iteração); (\*): execução não completada devido ao longo tempo de convergência ( $> 10^6$ ); (\*\*): dados insuficientes

Função	Algoritmo		
	PLR		
	$\mu\sigma_{\max}$	$\bar{\delta}$	$\delta$
$H_1(z)$	0,17	$6,9 \times 10^{-2}$	$3,6 \times 10^{-2}$
$H_4(z)$	0,018	$2,5 \times 10^{-3}$	$8,8 \times 10^{-4}$
$H_6(z)$	0,11	$2,6 \times 10^{-1}$	$5,4 \times 10^{-2}$
$H_7(z)$	0,0084	$1,7 \times 10^{-2}$	$4,6 \times 10^{-3}$
$H_9(z)$	0,9	$5,2 \times 10^{-3}$	$4,2 \times 10^{-3}$
$H_{12}(z)$	(*)	(*)	(*)
$H_{14}(z)$	0,075	$2,6 \times 10^{-1}$	$9,6 \times 10^{-2}$
$H_{15}(z)$	0,022	$6,9 \times 10^{-2}$	$9,2 \times 10^{-3}$

Tabela 2.13: Valores de  $\mu\sigma_{\max}$ ,  $\bar{\delta}$  e  $\delta$  obtidos das execuções do algoritmo PLR ( $\delta$  e  $\bar{\delta}$  em db/iteração); (\*): execução não completada devido ao longo tempo de convergência ( $> 10^6$ )

resultados apresentados.

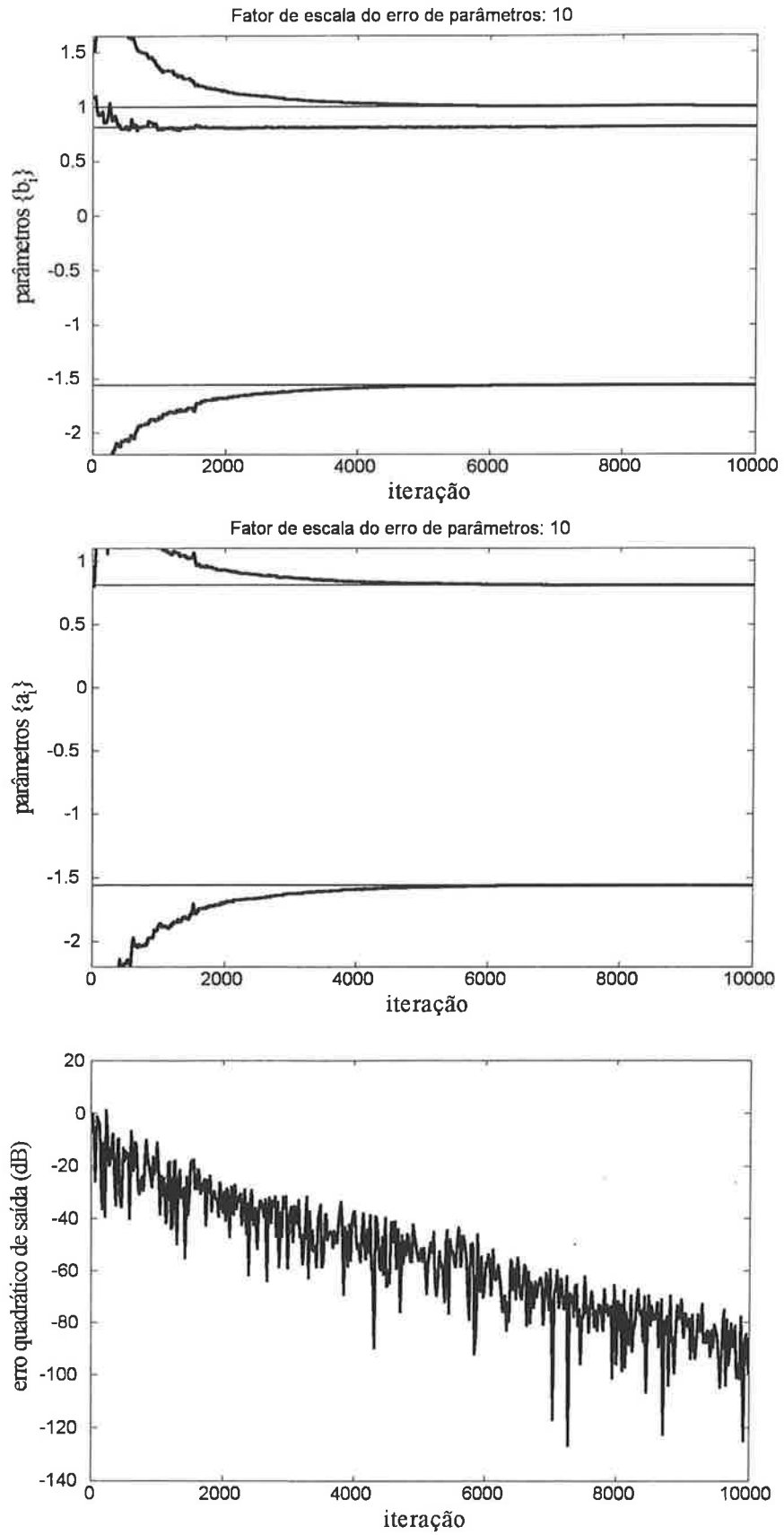


Figura 2-17: Algoritmo SHARF, sistema  $H_6(z)$  (parâmetros em traços claros)

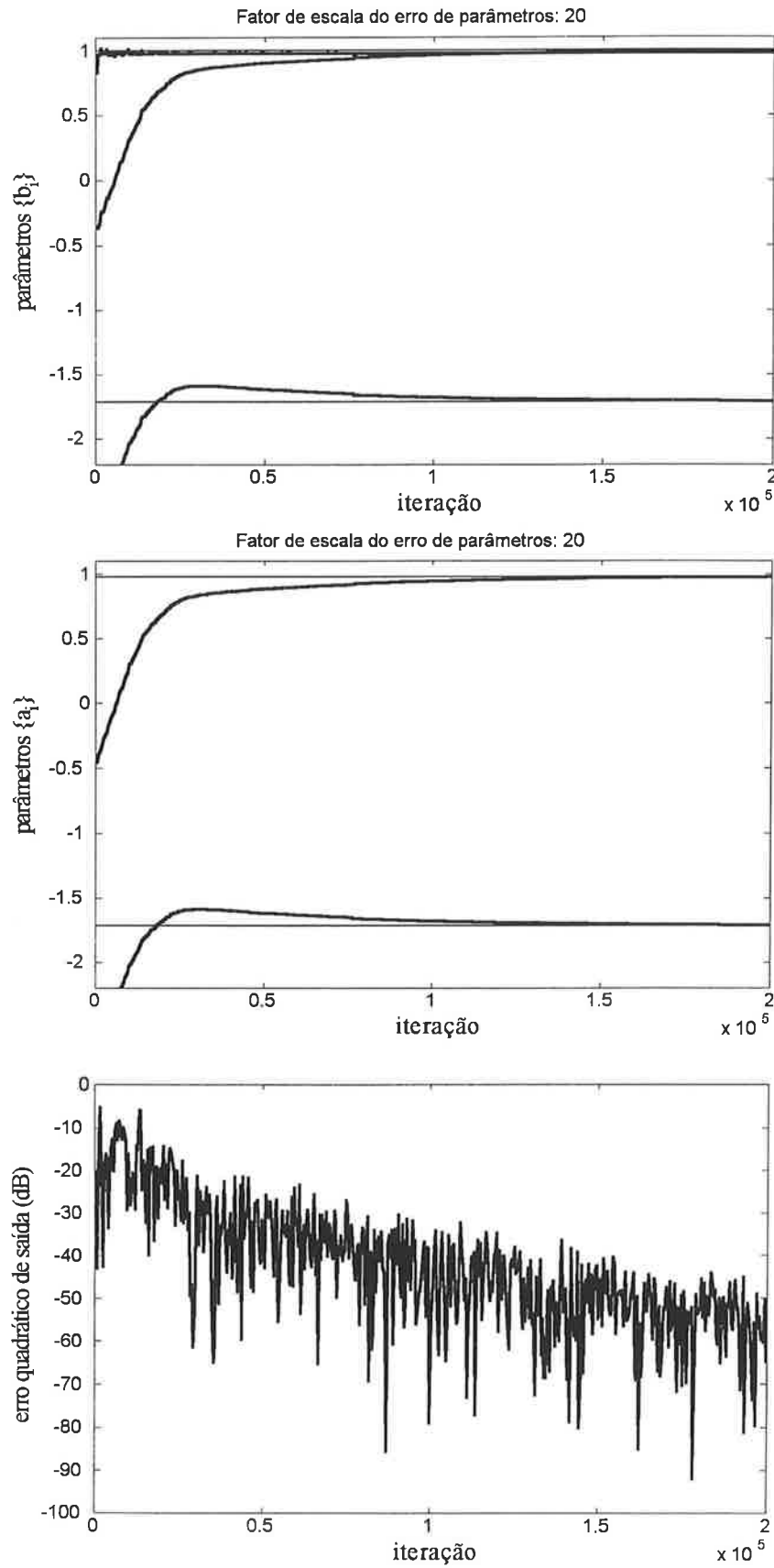


Figura 2-18: Algoritmo SHARF, sistema  $H_7(z)$  (parâmetros em traços claros)

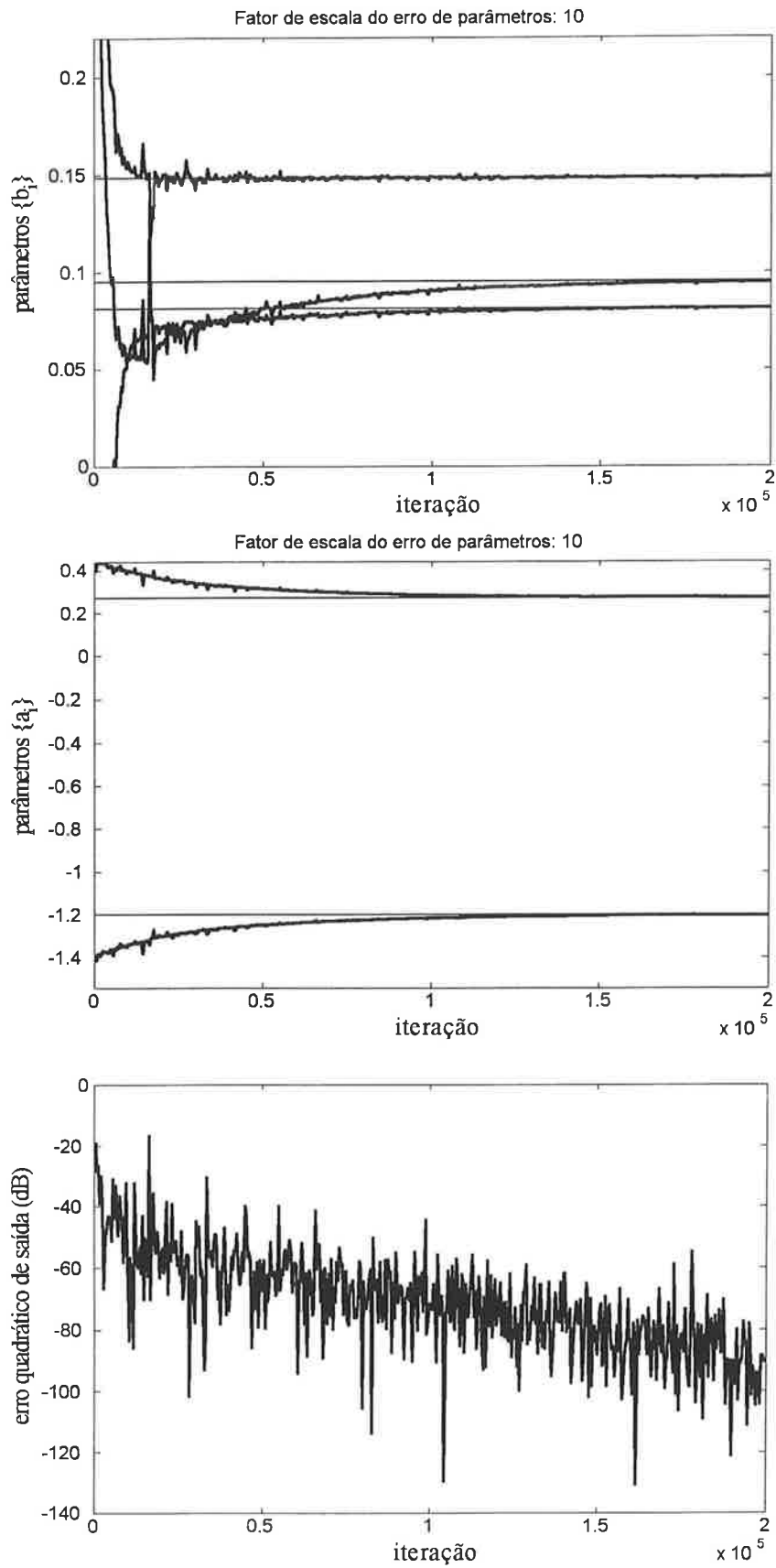


Figura 2-19: Algoritmo RG, sistema  $H_9(z)$  (parâmetros em traços claros)

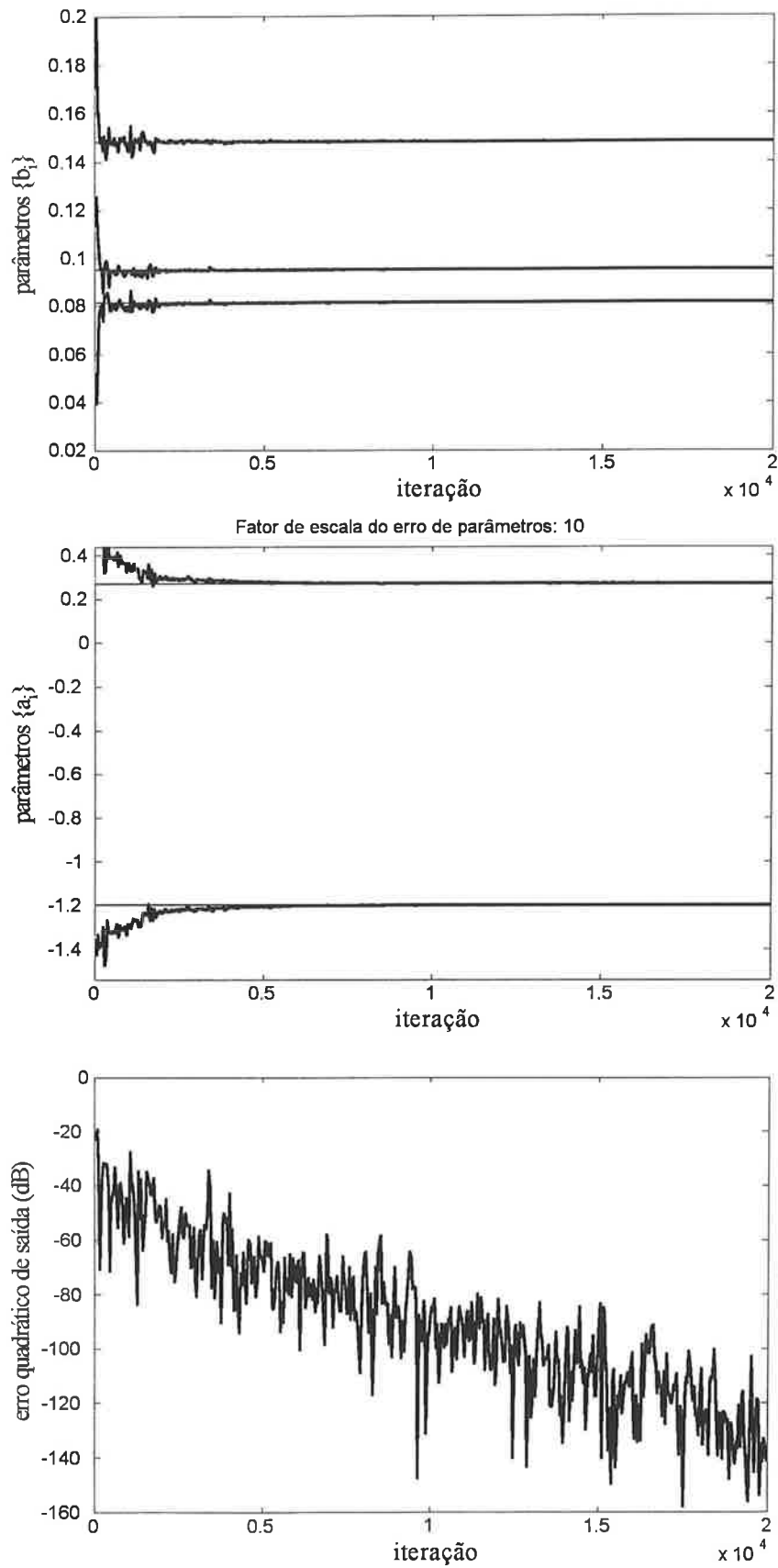


Figura 2-20: Algoritmo PLR; sistema  $H_9(z)$  (parâmetros em traços claros)

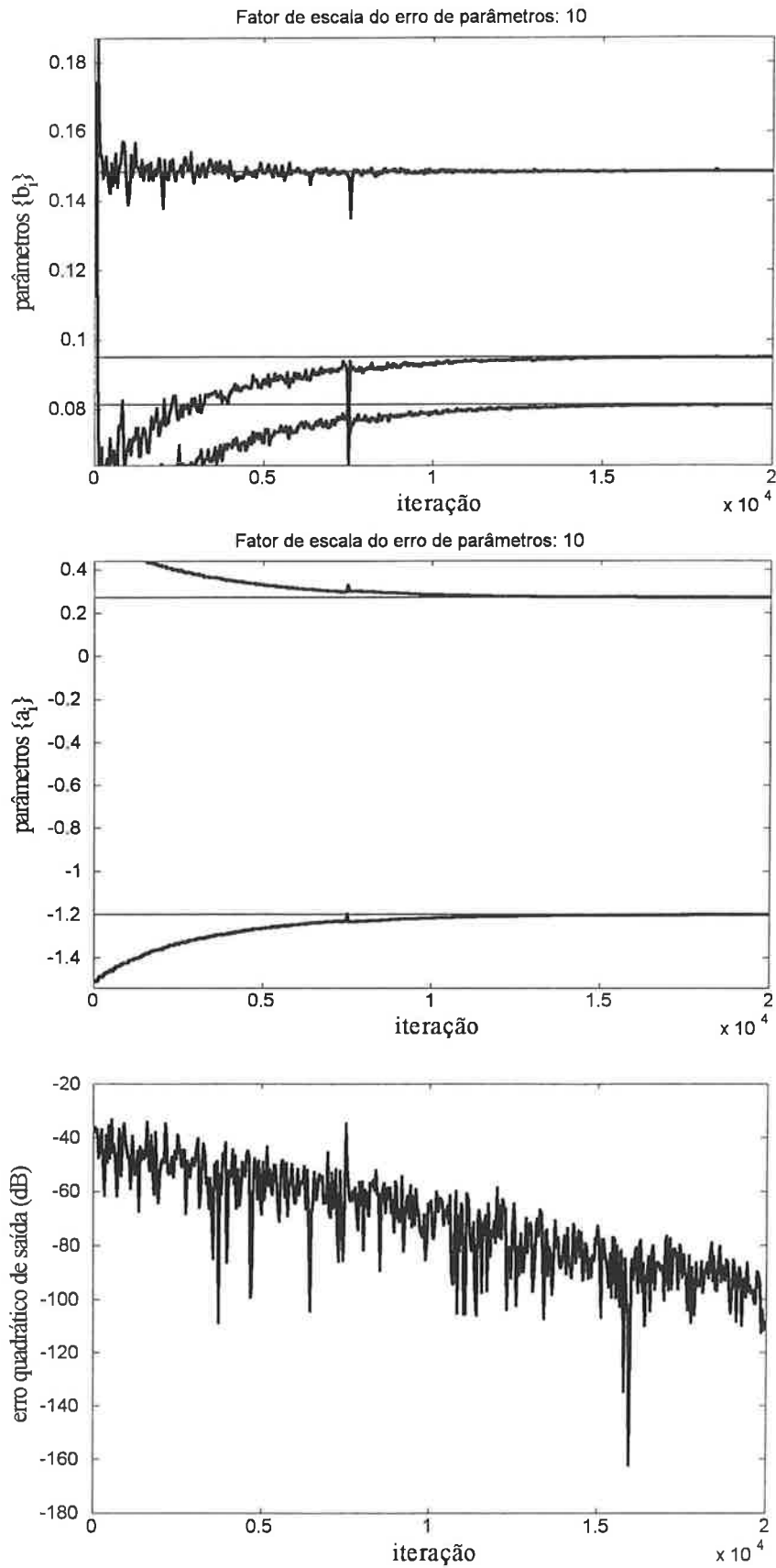


Figura 2-21: Algoritmo SHARF, sistema  $H_9(z)$  (parâmetros em traços claros)



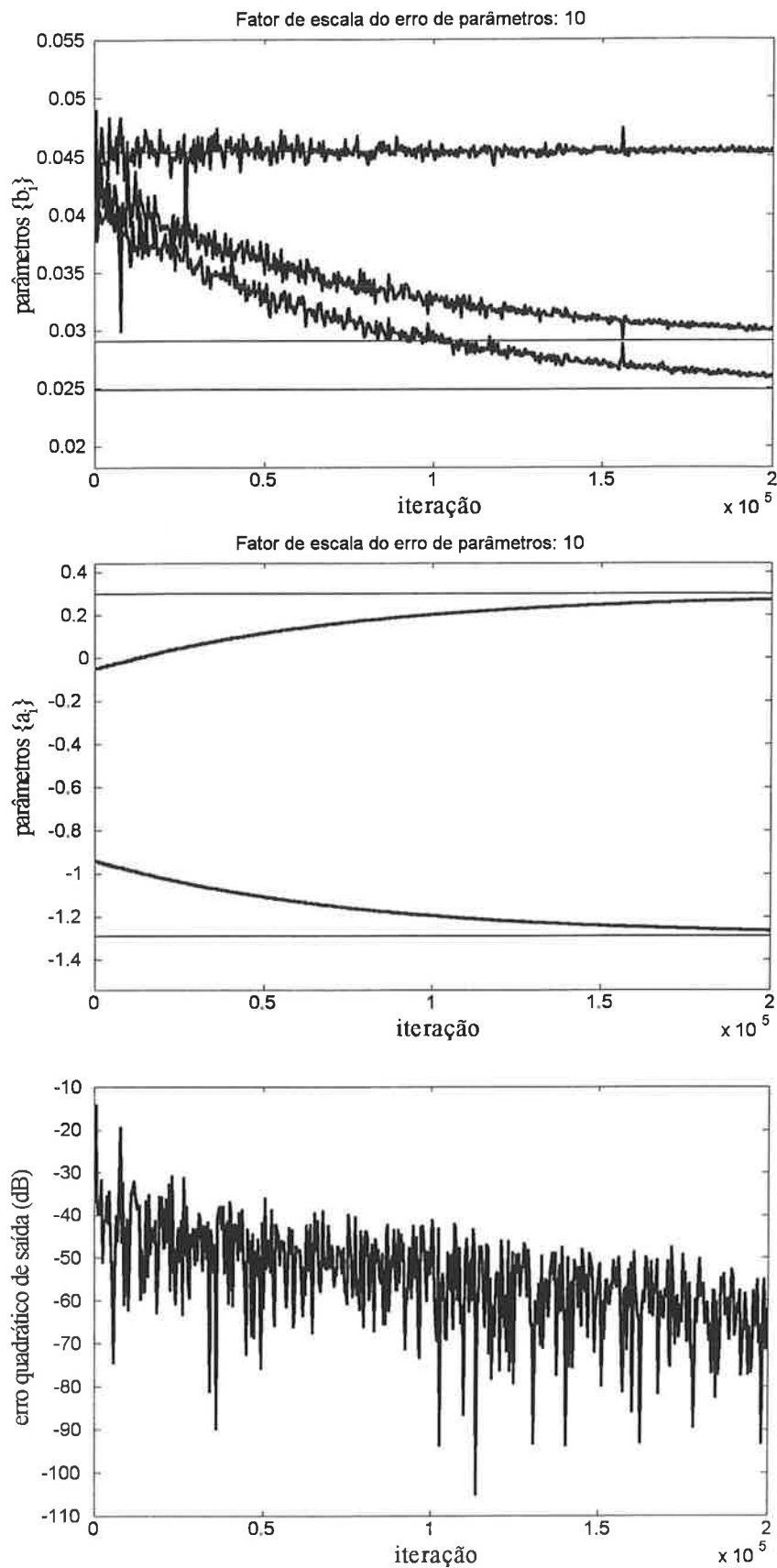


Figura 2-22: Algoritmo SHARF, sistema  $H_{12}(z)$  (parâmetros em traços claros)

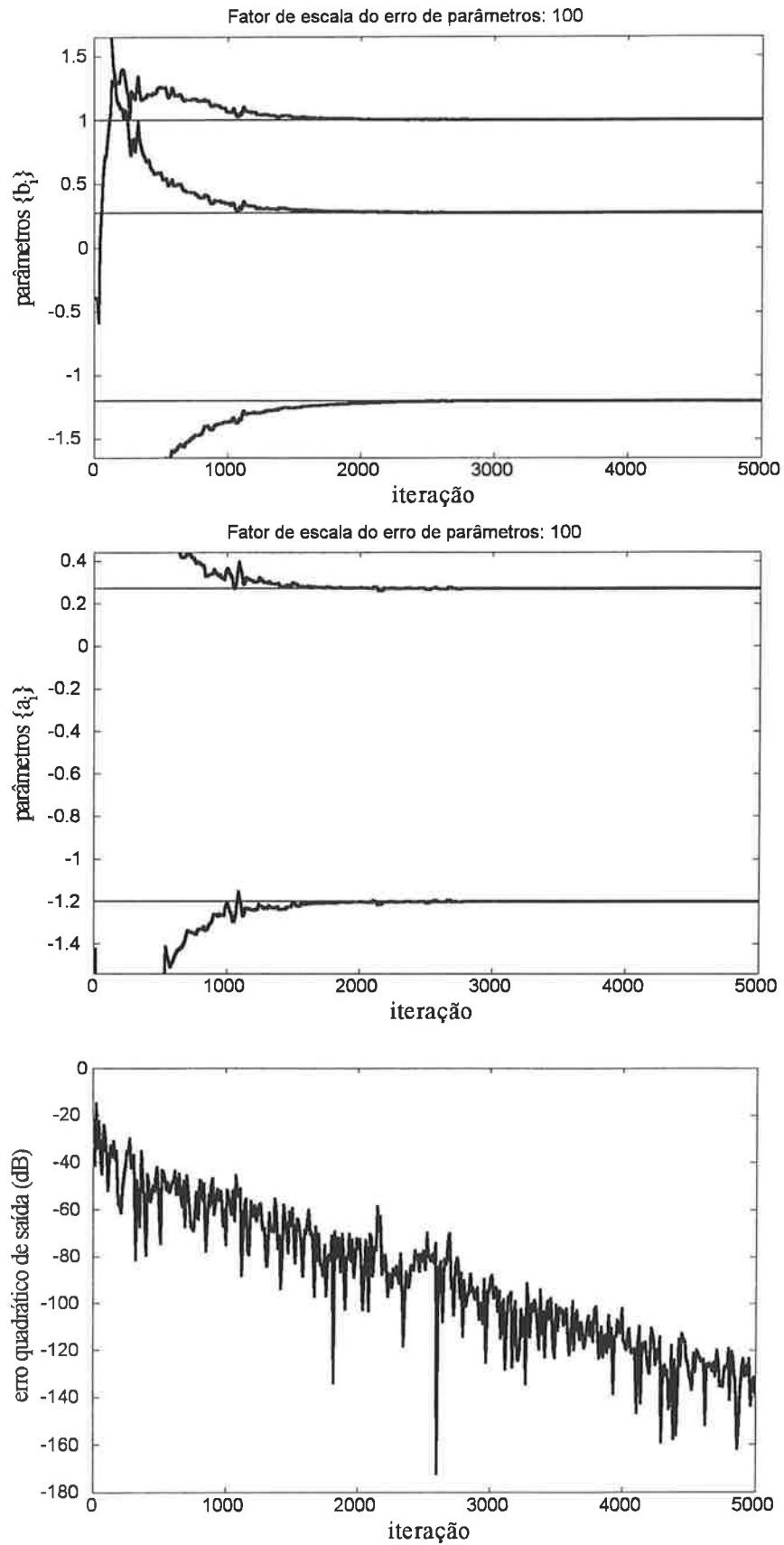


Figura 2-23: Algoritmo RG, sistema  $H_{14}(z)$  (parâmetros em traços claros)

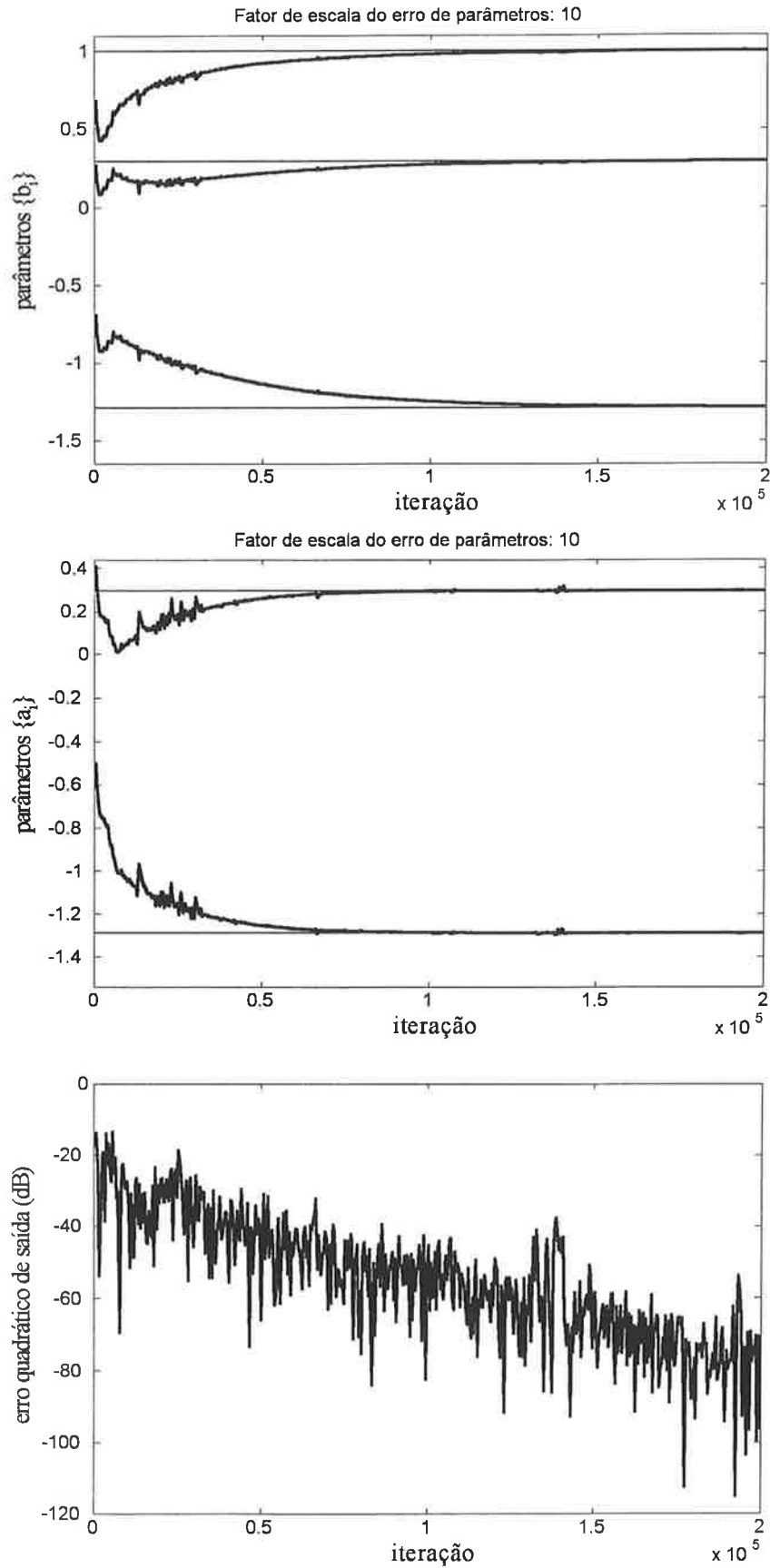


Figura 2-24: Algoritmo RG, sistema  $H_{15}(z)$  (parâmetros em traços claros)

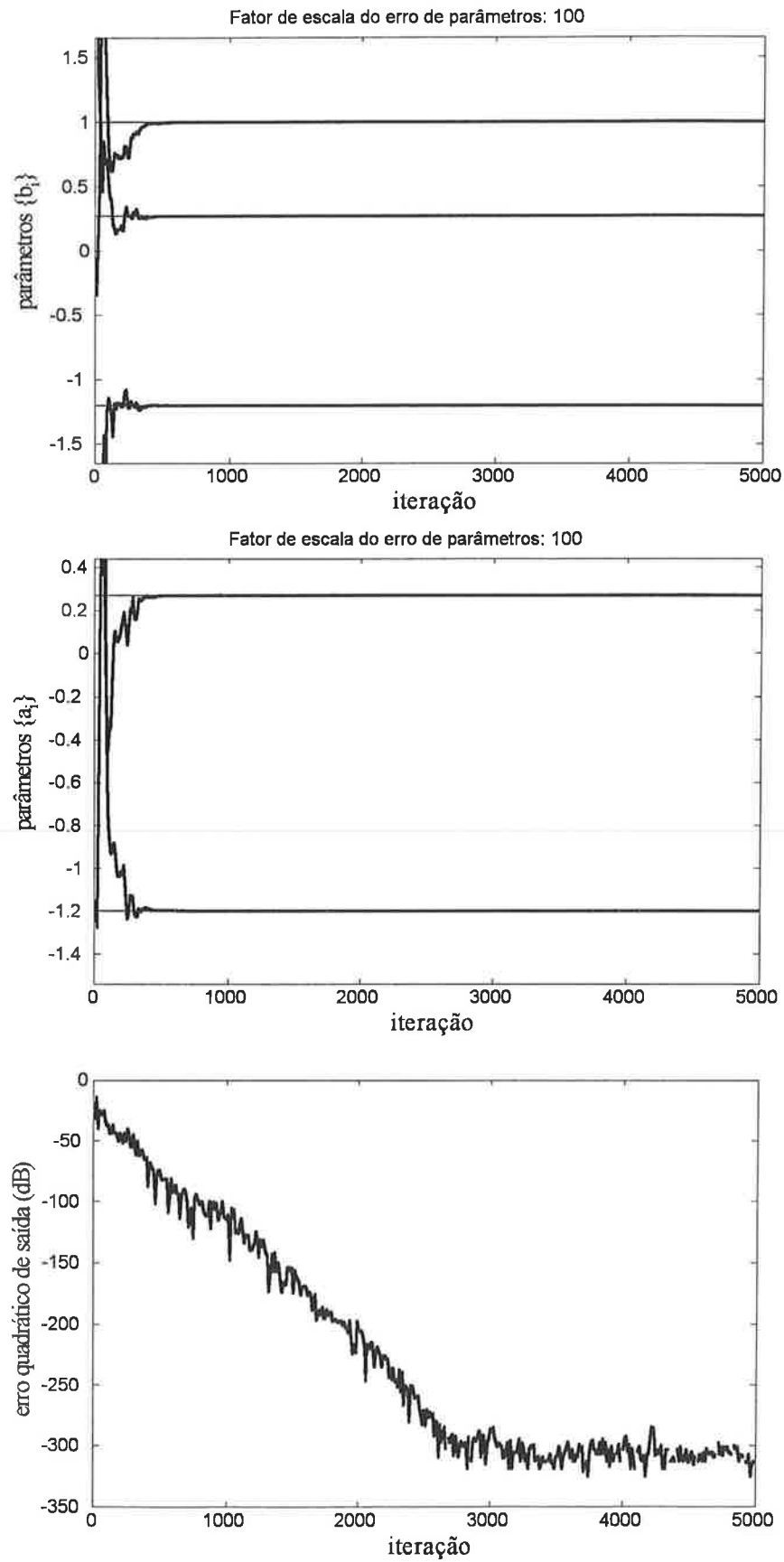


Figura 2-25: Algoritmo PLR, sistema  $H_{14}(z)$  (parâmetros em traços claros)

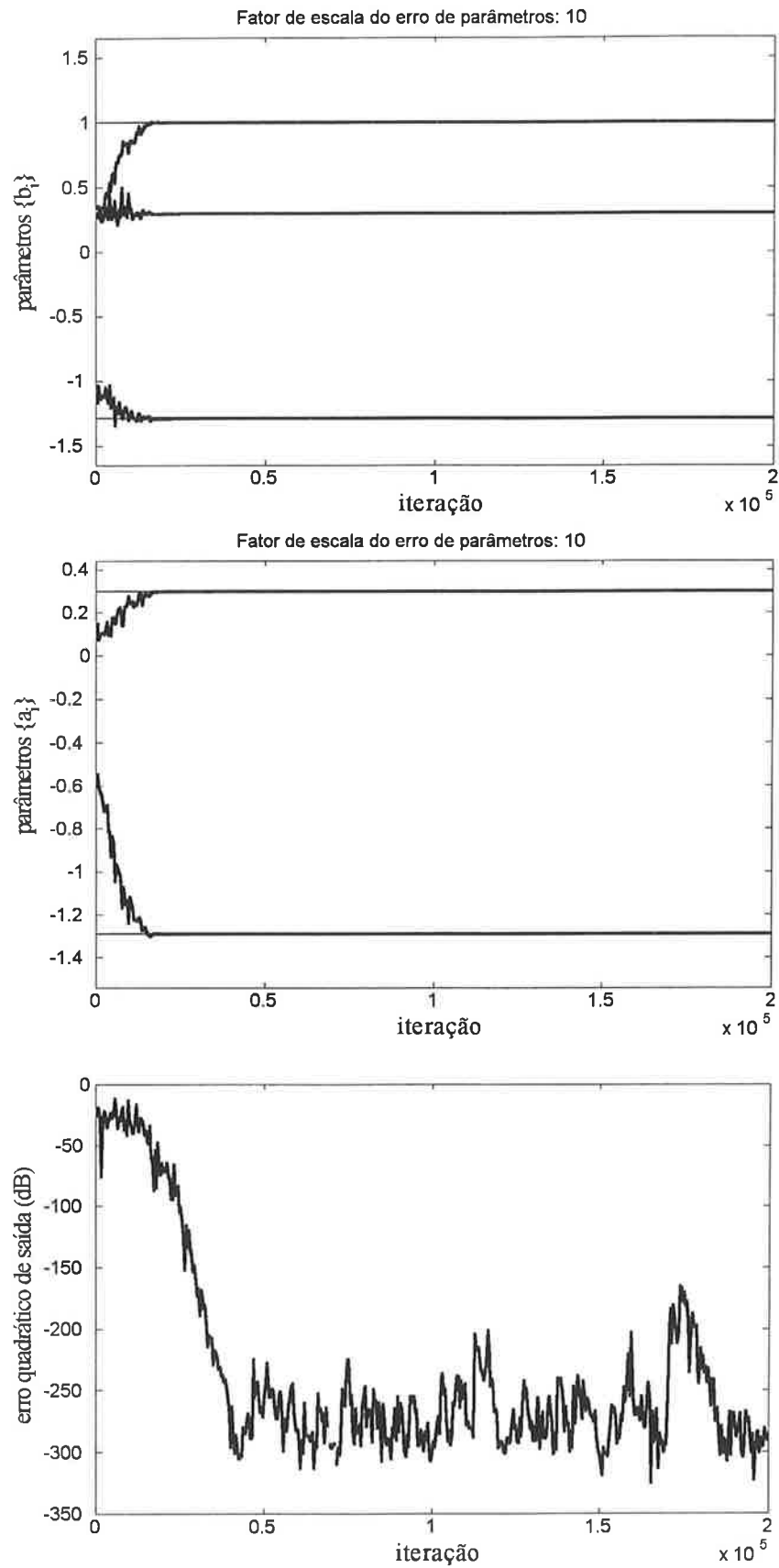


Figura 2-26: Algoritmo PLR, sistema  $H_{15}(z)$  (parâmetros em traços claros)

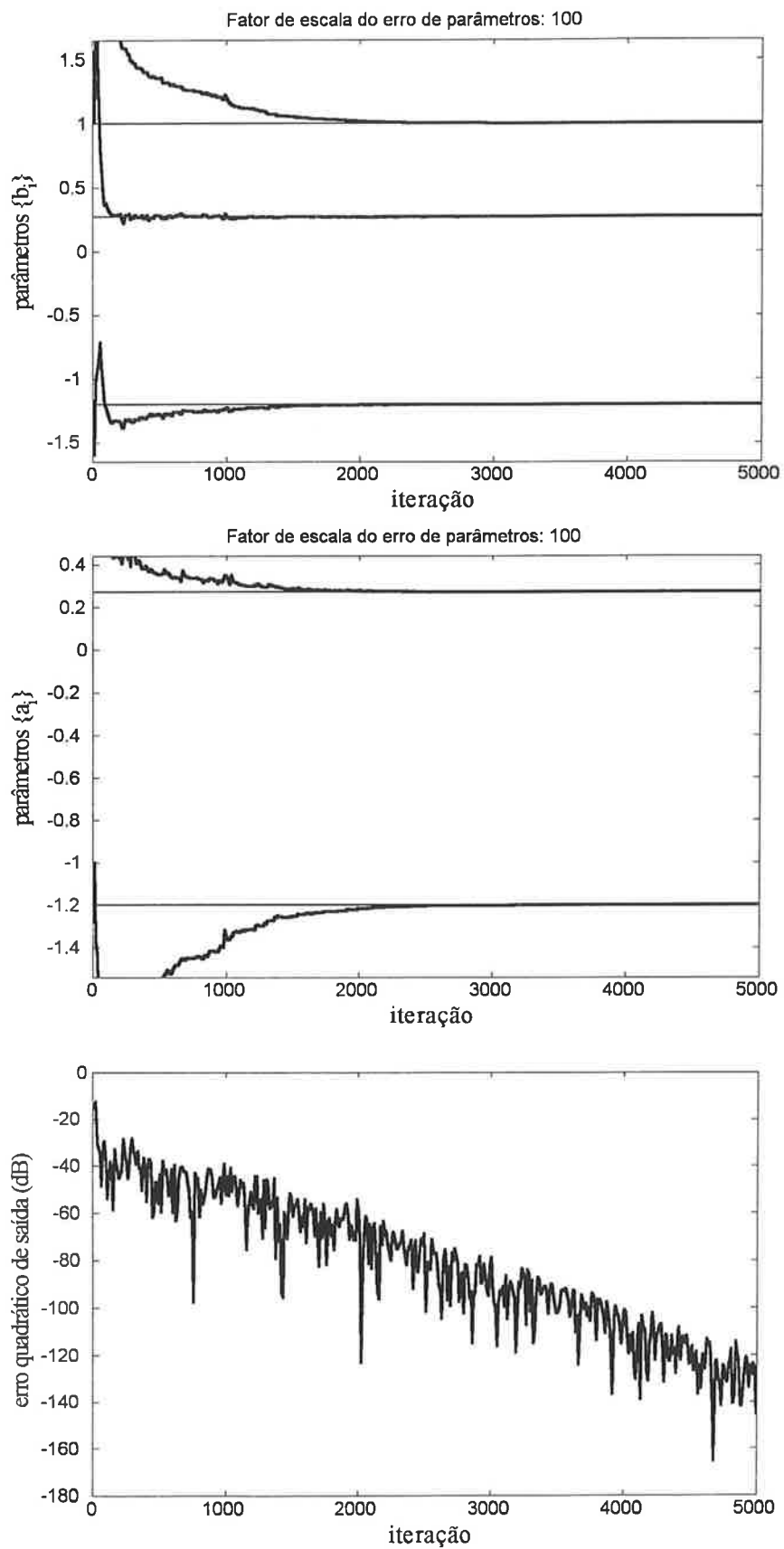


Figura 2-27: Algoritmo SHARF, sistema  $H_{14}(z)$  (parâmetros em traços claros)

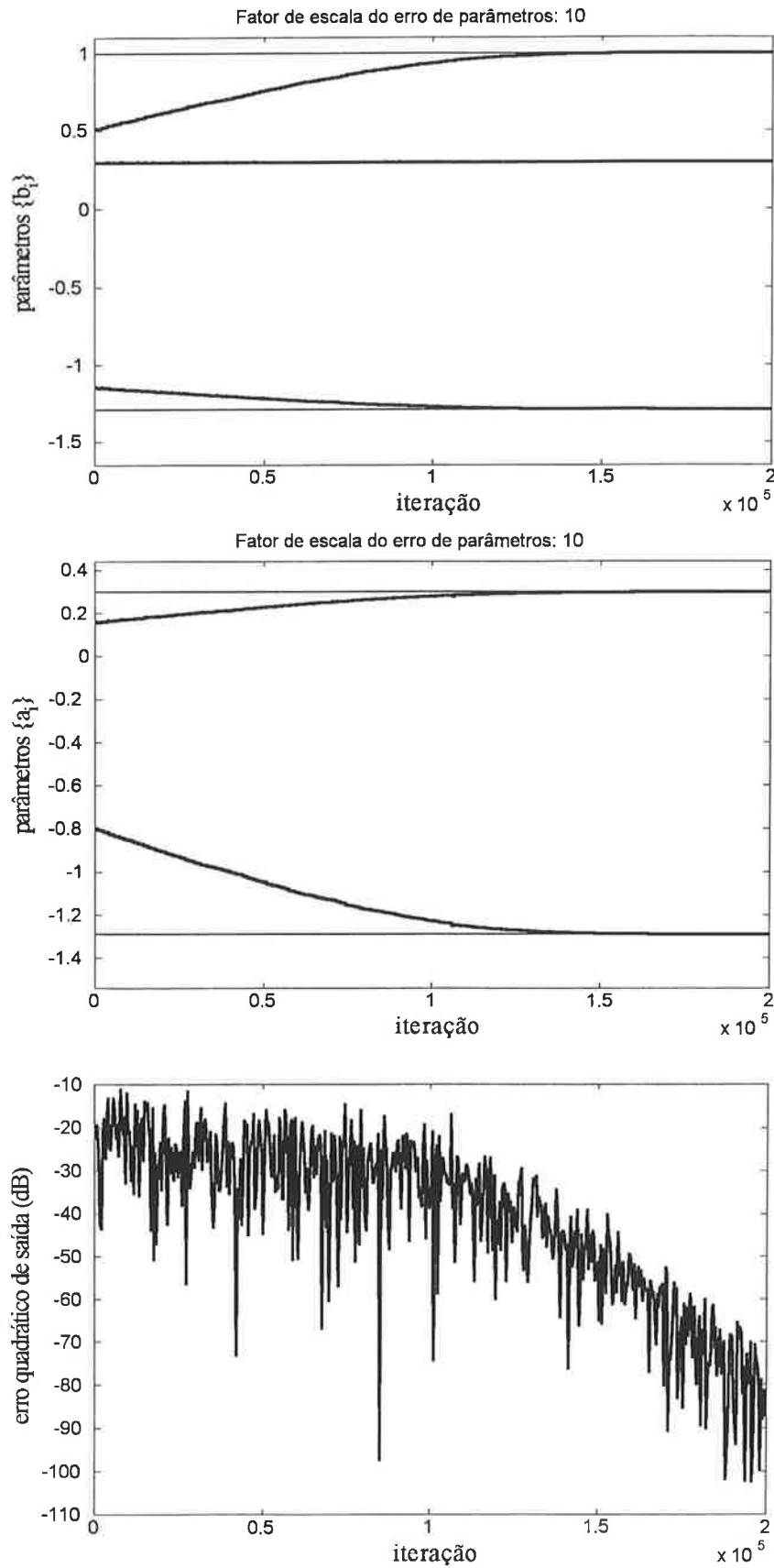


Figura 2-28: Algoritmo SHARF, sistema  $H_{15}(z)$  (parâmetros em traços claros)

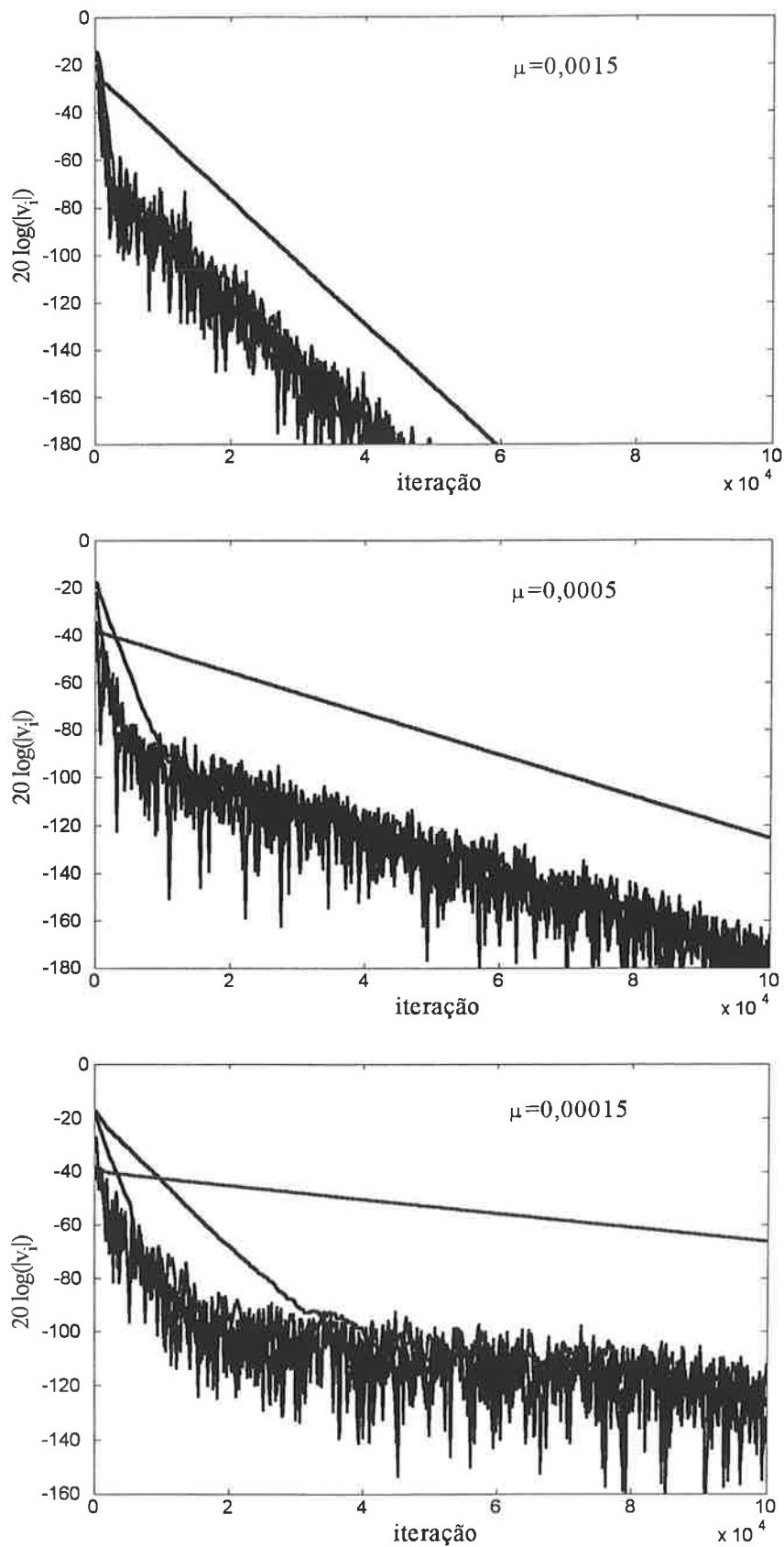


Figura 2-29: Algoritmo RG, sistema  $H_1(z)$ : verificação da aproximação por modos normais; cada traço corresponde a um elemento de  $\mathbf{v}(n) = \mathbf{P}^{-1}\tilde{\mathbf{w}}(n)$ , para diferentes valores do passo de adaptação  $\mu$ .



## 2.7.2 Discussão dos resultados

### Funções $H_1(z)$ e $H_4(z)$

Estes casos serão analisados em maior detalhe, de modo a exemplificar cada aspecto mencionado no início deste item. Este procedimento não será repetido, porém, para os demais casos.

As funções  $H_1(z^{-1})$  e  $H_4(z^{-1})$  tem norma  $L_2$  unitária, e um par de polos complexos de ângulo  $30^\circ$  e raio 0,9 e 0,99, respectivamente (Tabela 2.2). Na Tabela 2.10 podemos ver que o algoritmo RG apresentou um tempo de convergência bem maior do que o algoritmo SHARF, diferença que aumentou ao aproximar o polo de  $H(z)$  da circunferência unitária (nos gráficos do algoritmo RG para a função  $H_4(z)$  é apresentada a adaptação até  $10^5$  iterações apenas, sendo que o número de iterações necessárias para um erro de  $-60$  dB, conforme apresentado na Tabela 2.12, é de  $1,5 \times 10^6$ ). Isto já era sugerido, conforme a análise feita no Item 2.2.4, pelo maior espalhamento dos valores singulares de  $\mathbf{R}(\mathbf{w}^*)$  para o algoritmo RG e pelo maior aumento deste espalhamento ao passar de  $H_1(z)$  para  $H_4(z)$  (Tabela 2.4). Notar, porém, que o efeito deste maior espalhamento é um pouco atenuado pelo maior valor de  $\mu\sigma_{\max}$  para o algoritmo RG (Tabela 2.13).

Para exemplificar o efeito do passo de adaptação  $\mu$  na validade da utilização da expressão recursiva dos modos normais (2.56) para descrever a convergência local dos parâmetros dada por (2.61), a execução do algoritmo RG para  $H_1(z)$  foi repetida para três valores de  $\mu$ , e o vetor  $\mathbf{v}(n) = \mathbf{P}^{-1}\tilde{\mathbf{w}}(n)$  em (2.61) foi calculado para cada caso. Os gráficos com o módulo dos elementos de  $\mathbf{v}(n)$  estão na Figura 2-29. Podemos ver que, conforme discutido no Item 2.2.4, quanto menor o valor de  $\mu$  maior é a amplitude do valor do elemento de  $\mathbf{v}(n)$  associado ao modo normal mais lento de (2.37) em relação à dos demais elementos.

Consideremos agora o caso do algoritmo PLR. Podemos ver que ao passar de  $H_1(z)$  para  $H_4(z)$  há um aumento de 4,7 vezes no espalhamento dos valores singulares (Tabela 2.4), uma diminuição de 9,4 vezes no valor de  $\alpha = \mu\sigma_{\max}$  (Tabela 2.13) e um aumento de 46 vezes no tempo de convergência. A instabilidade do algoritmo ao usar-se um valor maior de  $\mu$  (0,0005) é mostrada na Figura 2-9. Constatou-se que, neste último caso, a divergência dos parâmetros que ocorre por volta das iterações 10000 e 70000 está associada a tentativas de saída do domínio de estabilidade. A diminuição de  $\alpha$  ao passar de  $H_1(z)$  para  $H_4(z)$  parece estar relacionada a este fato e à positividade de  $\mathbf{R}(\mathbf{w}^*)$ . Pode-se verificar que para  $H_1(z)$

Algoritmo PLR, Função $H_1(z)$					
$\mu$	$10^{-2}$	$5 \times 10^{-3}$	$2 \times 10^{-3}$	$1 \times 10^{-3}$	$5 \times 10^{-4}$
$\bar{\delta}$	$4,6 \times 10^{-2}$	$2,3 \times 10^{-2}$	$9,2 \times 10^{-3}$	$4,6 \times 10^{-3}$	$2,3 \times 10^{-3}$
$\delta$	$2,9 \times 10^{-2}$	$1,9 \times 10^{-2}$	$8,3 \times 10^{-3}$	$4,7 \times 10^{-3}$	$2,3 \times 10^{-3}$

Tabela 2.14: Valores de  $\bar{\delta}$  e  $\delta$ , em dB/iteração, para o algoritmo PLR e  $H_1(z)$

Algoritmo PLR, Função $H_4(z)$			
$\mu$	$10^{-4}$	$5 \times 10^{-5}$	$2 \times 10^{-5}$
$\bar{\delta}$	$8,3 \times 10^{-4}$	$4,2 \times 10^{-4}$	$1,7 \times 10^{-4}$
$\delta$	$4,6 \times 10^{-4}$	$2,8 \times 10^{-4}$	$9,6 \times 10^{-5}$

Tabela 2.15: Valores de  $\bar{\delta}$  e  $\delta$ , em dB/iteração, para o algoritmo PLR e  $H_4(z)$

e  $H_4(z)$  a condição (2.135), suficiente para a positividade de  $\mathbf{R}(\mathbf{w}^*)$  no caso do algoritmo PLR, não é atendida (a positividade de uma matriz pode ser verificada pela positividade dos autovalores de sua parte simétrica  $(\mathbf{R}(\mathbf{w}^*) + \mathbf{R}^\top(\mathbf{w}^*))/2$ ). Apesar de (2.135) não ser, dado um sinal de entrada  $u(n)$ , uma condição sempre necessária para a positividade de  $\mathbf{R}(\mathbf{w}^*)$ , neste caso resulta que mesmo com  $u(n)$  branco,  $\mathbf{R}(\mathbf{w}^*)$  não é positiva definida (seus autovalores, no entanto, tem ainda a parte real positiva, o que é necessário para  $\mathbf{w}^*$  ser um ponto de convergência). Como visto na expressão (2.48) do Item 2.2.3, isto significa que a norma do erro de parâmetros não é necessariamente sempre decrescente em uma vizinhança de  $\mathbf{w}^*$ . Mais especificamente, se em uma dada iteração o erro de parâmetros está em uma direção associada a um autovalor negativo de  $\mathbf{R}(\mathbf{w}^*) + \mathbf{R}^\top(\mathbf{w}^*)$ , então, naquela iteração, a norma do erro de parâmetros aumenta. Pode-se constatar que para  $H_1(z)$  a matriz  $\mathbf{R}(\mathbf{w}^*) + \mathbf{R}^\top(\mathbf{w}^*)$  tem um autovalor negativo de valor  $-0,030$  e para  $H_4(z)$ , um autovalor negativo de valor  $-0,81$ . Pode-se conjecturar então que é necessário um valor de  $\mu$  menor no segundo caso para evitar que, caindo o erro de parâmetros na direção associada àquele autovalor, ocorra uma tentativa de saída do domínio de estabilidade.

Para o algoritmo PLR ainda, em relação aos valores de  $\bar{\delta}$  e  $\delta$ , como resultado da não-simetria da matriz  $\mathbf{R}(\mathbf{w}^*)$  temos que o valor de  $\mu$  que, em cada caso, propicia o menor tempo de convergência não é pequeno o suficiente para valer a aproximação feita na análise da velocidade de convergência local no Item 2.2.4. Conforme pode ser visto nas Tabelas 2.14 e 2.15, à medida em que  $\mu$  é diminuído (o que, evidentemente, não seria feito na prática, uma vez que o tempo de convergência aumentaria) os valores de  $\bar{\delta}$  e  $\delta$  se aproximam.

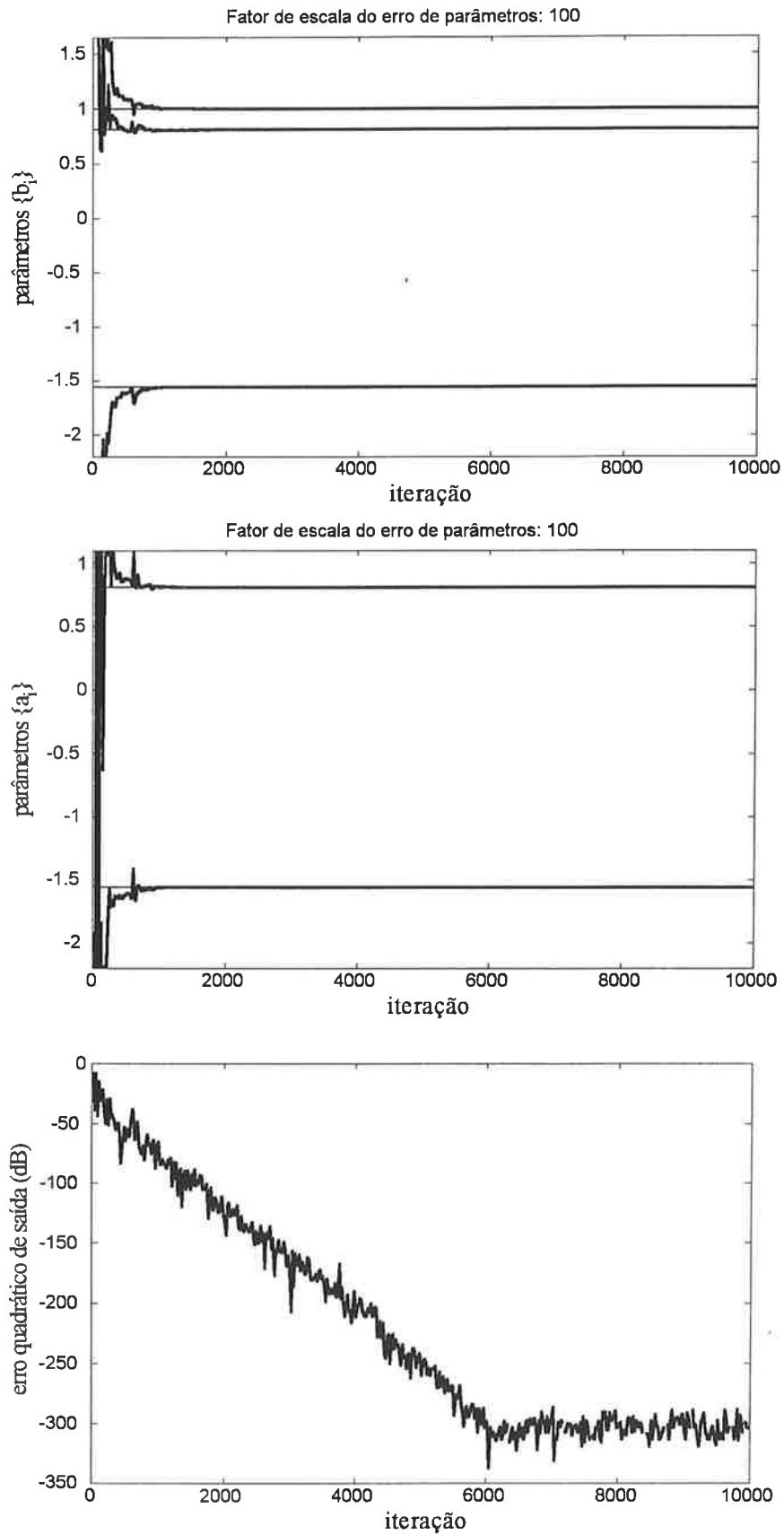


Figura 2-30: Algoritmo SHARF,  $P(z) = 1 - 0,4z$ , sistema  $H_6(z)$  (parâmetros em traços claros).

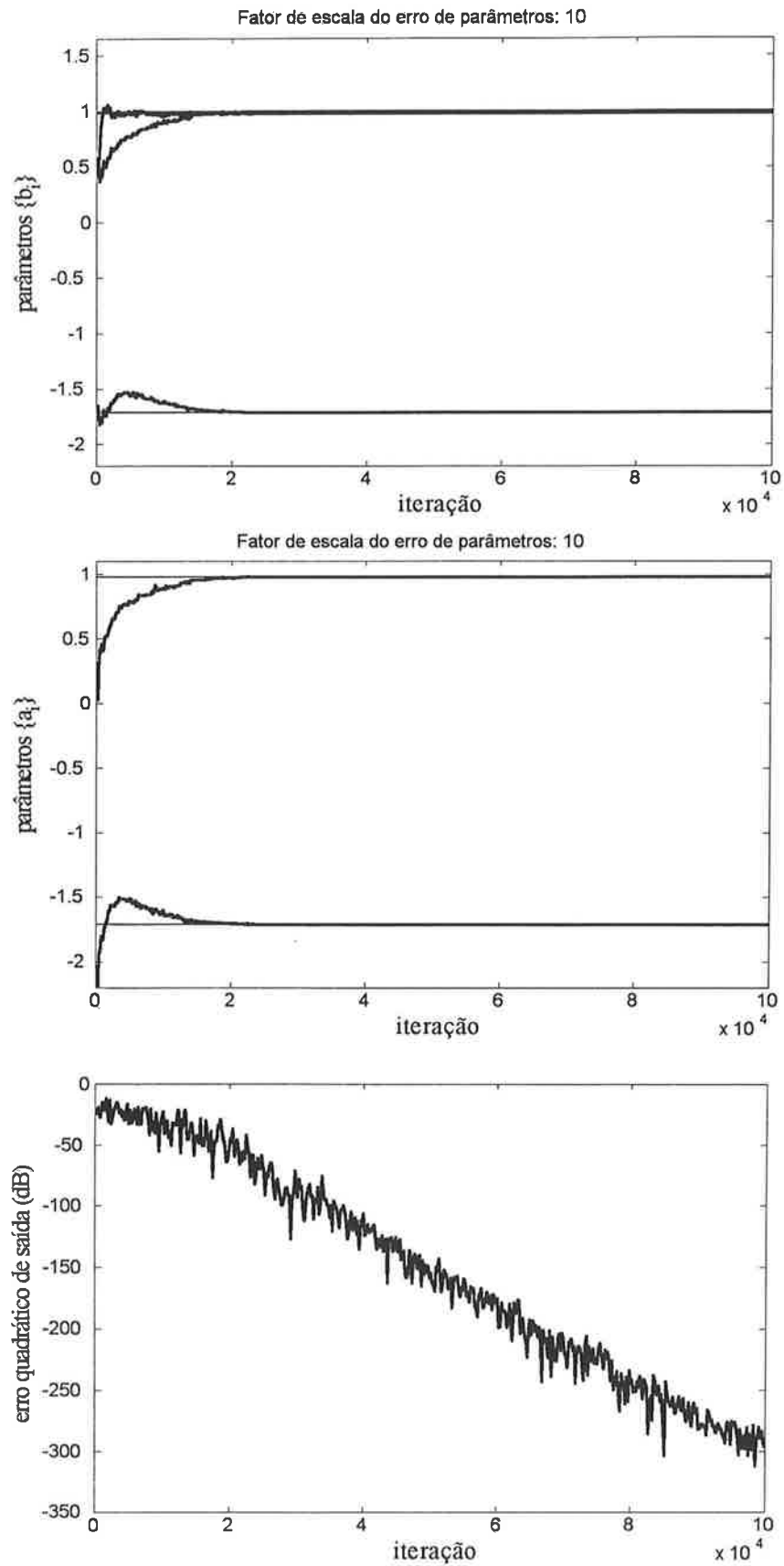


Figura 2-31: Algoritmo SHARF,  $P(z) = 1 - 0,75z$ , sistema  $H_7(z)$  (parâmetros em traços claros).

## Funções $H_6(z)$ e $H_7(z)$

As funções  $H_6(z^{-1})$  e  $H_7(z^{-1})$  são passa-tudo, e tem um par de polos complexos de ângulo  $30^\circ$  e raio 0,9 e 0,99, respectivamente (Tabela 2.2). Nestes casos, o algoritmo RG apresentou um tempo de convergência em torno de duas vezes menor do que o algoritmo SHARF (Tabela 2.10). Podemos ver que os tempos de convergência seguiram aproximadamente as variações do espalhamento dos valores singulares de  $\mathbf{R}(\mathbf{w}^*)$  (Tabela 2.4) e do valor de  $\alpha = \mu\sigma_{\max}$  (Tabela 2.13).

No caso do algoritmo PLR, podemos ver (Tabela 2.10) que o tempo de convergência apresentou um aumento muito grande ao passar de  $H_6(z)$  para  $H_7(z)$  e o ganho no tempo de convergência que poderia se esperar pelo melhor condicionamento de  $\mathbf{R}(\mathbf{w}^*)$  para o algoritmo PLR (Tabela 2.4) não se concretizou. Através das Figuras 2-15 e 2-16 podemos ver que a razão disto é que os parâmetros permanecem um longo tempo em uma região onde a convergência tem características diferentes das que assume localmente em torno de  $\mathbf{w}^*$ . Este comportamento, assim como o visto para os casos  $H_1(z)$  e  $H_4(z)$ , parece estar relacionado com a positividade de  $\mathbf{R}(\mathbf{w}^*)$ , como discutido a seguir.

Para as funções  $H_6(z)$  e  $H_7(z)$  a condição de positividade (2.135) não é obedecida e mesmo para  $u(n)$  branco  $\mathbf{R}(\mathbf{w}^*)$  não é positiva definida. Para tentar tirar proveito do melhor condicionamento de  $\mathbf{R}(\mathbf{w}^*)$  para o algoritmo PLR, foi empregado então, conforme discutido no Item 2.6.7, o algoritmo SHARF não com  $P(z) = D(z)$ , mas com  $P(z)$  de primeira ordem e o mais próximo possível de 1 para o qual a condição de positividade (2.140) fosse obedecida (com  $P(z) \equiv 1$  recai-se no algoritmo PLR). Este procedimento alcançou bons resultados, conforme podemos ver pelo resultados apresentados na Tabela 2.16. Podemos ver que para o caso  $H_6(z)$  o tempo de convergência é 3,6 vezes menor e 8,3 vezes menor do que com os algoritmos RG e SHARF com  $P(z) = D(z)$  (Tabela 2.10), respectivamente, e que no caso  $H_7(z)$  o tempo de convergência é 5,6 vezes menor e 9,1 vezes menor do que com os algoritmos RG e SHARF com  $P(z) = D(z)$ , respectivamente. Os gráficos estão nas Figuras 2-30 e 2-31.

A partir dos efeitos negativos que resultaram da não positividade de  $\mathbf{R}(\mathbf{w}^*)$ , para este caso e o anterior do algoritmo PLR, parece razoável conjecturar que, apesar da positividade de  $\mathbf{R}(\mathbf{w}^*)$  poder não ser, para uma dada densidade espectral de potência de  $u(n)$ , uma condição necessária para  $\mathbf{w}^*$  ser um ponto de convergência local do algoritmo, é uma condição necessária para que o mesmo seja "bem-comportado" localmente.

Algoritmo							
SHARF, $P(z) \neq D(z)$							
Função	$P(z)$	$\text{Re}[\lambda_m]$	$\sigma_{\min}$	$\sigma_{\max}$	$\mu$	$\mu\sigma_{\max}$	$n_{60}$
$H_6(z)$	$1 - 0,4z$	0,18	0,16	2,8	$4 \times 10^{-2}$	0,11	840
$H_7(z)$	$1 - 0,75z$	0,077	0,075	2,7	$5 \times 10^{-3}$	0,014	$2,5 \times 10^4$

Tabela 2.16: Algoritmo SHARF com  $P(z) \neq A(z)$

A questão do efeito da densidade espectral de potência do sinal de entrada  $u(n)$  quando não há o atendimento das condições de positividade (2.135) e (2.140), discutido anteriormente nos Itens 2.5 e 2.5.3, também pode ser exemplificada aqui. Se para a função  $H_7(z)$ , ao invés de tomar-se  $P(z) = 1 - 0,75z$  como acima, tomar-se  $P(z) = 1 - 0,3z$ , então  $\mathbf{R}(\mathbf{w}^*)$  será positiva para  $u(n)$  branco mas (2.140) não será obedecida. Ou seja, pode-se obter  $u(n)$  tal que  $\mathbf{R}(\mathbf{w}^*)$  não seja positiva definida, e, ainda pior, tenha algum autovalor com parte real negativa. Por exemplo, se for usado  $u(n)$  gerado passando ruído branco de potência unitária por um filtro com polos em  $0,85 \angle \pm 35^\circ$ , zeros em  $0,9 \angle \pm 20^\circ$  e norma unitária, pode-se ver na Figura 2-32 que a densidade espectral de potência  $u(n)$  está concentrada na região onde  $\text{Re}[P(e^{j\omega})/D(e^{j\omega})] < 0$ . Isto faz com que tenha  $\mathbf{R}(\mathbf{w}^*)$  autovalores com parte real negativa, e, conseqüentemente,  $\mathbf{w}^*$  não seja um ponto de convergência do algoritmo (ver Propriedade 2.2). Podemos comprovar isto na prática pelos gráficos da Figura 2-33, onde foi usada a entrada  $u(n)$  mencionada, e, para mostrar bem a instabilidade do algoritmo, iniciou-se a adaptação com  $B(z) = 0,9999C(z)$  e  $A(z) = 0,9999D(z)$ , e usou-se  $\mu = 0,0005$  ao invés de  $\mu = 0,005$ .

### Funções $H_9(z)$ e $H_{12}(z)$

As funções  $H_9(z^{-1})$  e  $H_{12}(z^{-1})$  tem norma  $L_2$  unitária e tem um polo real em 0,9 e 0,99, respectivamente, e um polo real em 0,3. Podemos ver que para  $H_9(z)$  o espalhamento dos valores singulares de  $\mathbf{R}(\mathbf{w}^*)$  (Tabela 2.4) para o algoritmo PLR é 5,9 vezes maior do que para o algoritmo SHARF, e o valor de  $\alpha$ , por outro lado é 10,7 vezes maior, o que contribui para que o tempo de convergência seja 1,4 vezes menor. Para  $H_{12}(z)$  a velocidade de convergência para os algoritmos RG e PLR era muito baixa, conforme poderia-se esperar do espalhamento dos valores singulares de  $\mathbf{R}(\mathbf{w}^*)$  (Tabela 2.4), e as execuções dos algoritmos não foram levadas até a convergência.

### Funções $H_{14}(z)$ e $H_{15}(z)$

As funções  $H_{14}(z^{-1})$  e  $H_{15}(z^{-1})$  são passa-tudo e tem um polo real em 0,9 e 0,99, respectivamente, e um polo real em 0,3. Nestes casos a condição de positividade (2.135) para o algoritmo PLR é atendida e não ocorrem os problemas vistos nos casos anteriores quando  $\mathbf{R}(\mathbf{w}^*)$  não era positiva definida. A expectativa de uma maior velocidade do algoritmo PLR, sugerida pelo melhor condicionamento de sua matriz  $\mathbf{R}(\mathbf{w}^*)$  (Tabela 2.4), neste caso é confirmada: podemos ver na Tabela 2.10 que, para  $H_{14}(z)$ , o tempo de convergência do algoritmo PLR foi 6,1 e 5,6 vezes menor do que o dos algoritmos RG e SHARF, respectivamente, e para  $H_{15}(z)$  foi 10,6 e 10 vezes menor do que o dos algoritmos RG e SHARF, respectivamente. Para  $H_{15}(z)$ , no entanto, o ganho em tempo de convergência não foi tão grande quanto poderia se imaginar a partir do condicionamento de  $\mathbf{R}(\mathbf{w}^*)$  para cada algoritmo. A razão disto, como pode ser visto pela grande diferença entre  $\bar{\delta}$  e  $\delta$  para este caso (ver Tabela 2.13), é que o valor de  $\mu$  que propiciou a maior velocidade de convergência local está longe do valor para o qual é válida a caracterização da convergência por meio de modos normais (ver Item 2.2.4).

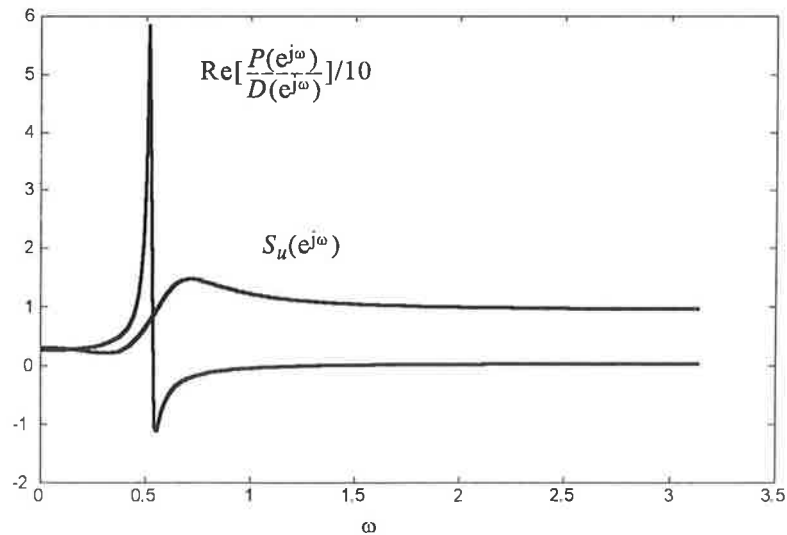


Figura 2-32: Sistema  $H_7(z)$ ,  $P(z) = 1 - 0,3z$ ,  $u(n)$  não-branco

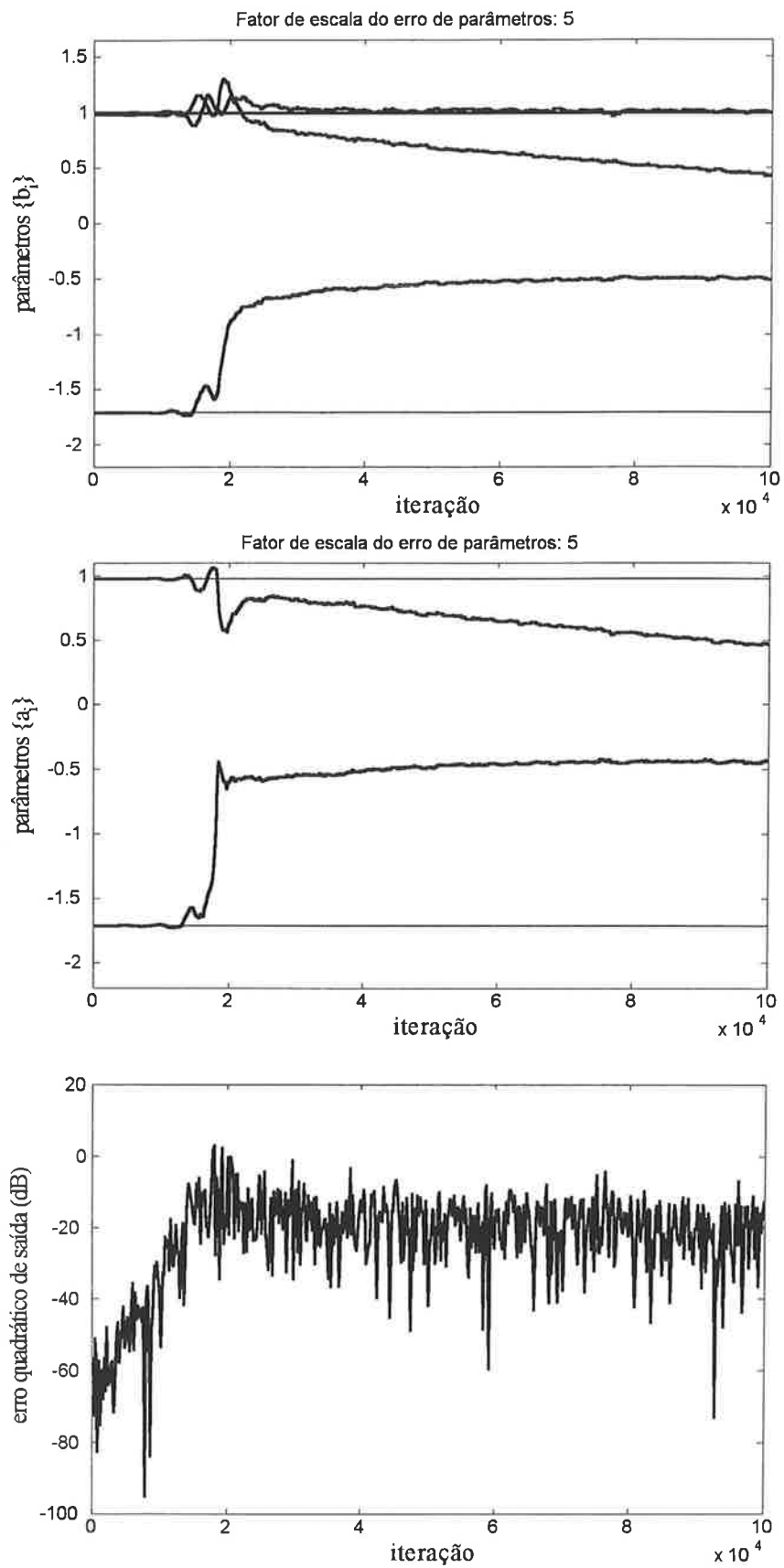


Figura 2-33: Algoritmo SHARF, sistema  $H_7(z)$ ,  $P(z) = 1 - 0,3z$ ,  $u(n)$  não-branco.



## Capítulo 3

# Análise da superfície de erro reduzida para filtros IIR adaptativos

Neste capítulo são analisadas a superfície de erro e a superfície de erro reduzida para filtros IIR adaptativos. Para algoritmos de adaptação do tipo gradiente, como o algoritmo RG, visto no Item 2.3, as propriedades destas superfícies tem relações diretas tanto com a sua convergência global (ponto inicial distante do ótimo) quanto com a sua convergência local (nas imediações do ponto ótimo). Para outros tipos de algoritmos, como os algoritmos SMM, PLR e SHARF, vistos nos Itens 2.4 a 2.5.3, estas relações não são tão diretas, porém, pode-se observar, na prática, que são importantes: como apresentado no Item 3.7, por exemplo, em regiões onde a superfície de erro é excessivamente plana a convergência de todos os algoritmos considerados é lenta.

Nos Itens 3.1 e 3.2 são introduzidas definições e propriedades necessárias nos itens seguintes. A exposição nestes dois itens é baseada em [1].

No Item 3.3 aborda-se a questão da extensão de propriedades da superfície de erro para o caso de uma implementação direta do filtro adaptativo a outras formas de implementação.

No Item 3.4, a partir da decomposição em valores singulares (SVD) da forma de Hankel do sistema sendo modelado, é apresentada uma decomposição da superfície de erro reduzida. A partir deste resultado, obtém-se então uma expressão para a matriz Hessiana no mínimo global da superfície de erro reduzida, sendo que nesta expressão existe uma separação parcial dos efeitos do sistema sendo modelado e da forma de realização do filtro adaptativo. Com base neste resultado, discute-se no Item 3.5 a possibilidade de eliminar a contribuição dada pela forma de realização do filtro adaptativo ao espalhamento dos autovalores da matriz

Hessiana no mínimo global da superfície de erro reduzida.

Neste trabalho, os resultados apresentados nos Itens 3.4 e 3.5 não puderam ser explorados quanto às suas possíveis implicações para a análise teórica e aplicação prática de filtros IIR adaptativos. Apesar disto, considerou-se apropriado incluí-los no trabalho, por parecerem resultados de interesse e que podem eventualmente servir de ponto de partida para outras investigações.

No item 3.6, é analisada uma característica global da superfície de erro reduzida: mostra-se que quando o sistema sendo identificado possui polos próximos à circunferência unitária a superfície de erro reduzida apresenta regiões planas com erro quadrático médio elevado. A existência destas regiões resulta em uma baixa velocidade de convergência global de algoritmos de passo constante, como exemplificado pelos resultados de simulações apresentados no Item 3.7.

## 3.1 Tópicos de sistemas lineares

### 3.1.1 Matriz e gramiano de controlabilidade

Consideremos a descrição em espaço de estados de um sistema linear com entrada  $u(n)$  e saída  $\hat{y}(n) = \hat{H}(z)u(n)$ :

$$\begin{aligned} \mathbf{x}(n+1) &= \mathbf{A}\mathbf{x}(n) + \mathbf{b}u(n) \\ y(n) &= \mathbf{c}^\top \mathbf{x}(n) + du(n), \end{aligned} \quad (3.1)$$

onde o vetor de estados  $\mathbf{x}(n)$  tem  $M$  elementos. A matriz de controlabilidade do sistema é definida por:

$$\mathcal{C} \triangleq [\mathbf{b} \quad \mathbf{A}\mathbf{b} \quad \mathbf{A}^2\mathbf{b} \quad \dots], \quad (3.2)$$

e o par  $(\mathbf{A}, \mathbf{b})$  é dito completamente controlável se  $\mathcal{C}$  tiver posto  $M$ , ou seja, se tiver  $M$  colunas linearmente independentes.

Consideremos que  $\mathbf{A}$  tem todos os autovalores dentro do círculo unitário. O gramiano de controlabilidade formado a partir de  $\mathcal{C}$  segundo  $\mathbf{K} = \mathcal{C}\mathcal{C}^\top$  pode então ser escrito como[1]:

$$\mathbf{K} = \sum_{k=0}^{\infty} \mathbf{A}^k \mathbf{b} (\mathbf{A}^k \mathbf{b})^\top$$

$$\begin{aligned}
&= \sum_{k=1}^{\infty} \mathbf{A}^k \mathbf{b} (\mathbf{A}^k \mathbf{b})^\top + \mathbf{b} \mathbf{b}^\top \\
&= \mathbf{A} \mathbf{K} \mathbf{A}^\top + \mathbf{b} \mathbf{b}^\top.
\end{aligned} \tag{3.3}$$

Se um par  $(\mathbf{A}, \mathbf{b})$  é completamente controlável segue diretamente que o seu gramiano de controlabilidade  $\mathbf{K}$  é uma matriz positiva definida.

Tomando agora uma raiz quadrada simétrica de  $\mathbf{K}^{-1}$ , temos que ela é uma transformação de similaridade  $\mathbf{T}$  que leva  $(\mathbf{A}, \mathbf{b}, \mathbf{K})$  a  $(\bar{\mathbf{A}} = \mathbf{T} \mathbf{A} \mathbf{T}^{-1}, \bar{\mathbf{b}} = \mathbf{T} \mathbf{b}, \bar{\mathbf{K}} = \mathbf{T} \mathbf{K} \mathbf{T}^\top)$  tal que  $\bar{\mathbf{K}} = \mathbf{I}$ . Considerando que  $\mathbf{A}$  tem todos os autovalores dentro do círculo unitário e usando (3.3) temos então que

$$\bar{\mathbf{A}} \bar{\mathbf{A}}^\top + \bar{\mathbf{b}} \bar{\mathbf{b}}^\top = \mathbf{I}. \tag{3.4}$$

Desta igualdade decorre diretamente que  $[\bar{\mathbf{A}} \ \bar{\mathbf{b}}]$  tem  $M$  linha ortonormais. Portanto, existem, um vetor  $\mathbf{u}$  e um escalar  $v_0$  únicos de tal modo que a matriz

$$\begin{bmatrix} \bar{\mathbf{A}} & \bar{\mathbf{b}} \\ \mathbf{u}^\top & v_0 \end{bmatrix}$$

seja ortogonal [1]. Além disso, pode-se mostrar que a função de transferência  $V(z)$  do sistema  $(\bar{\mathbf{A}}, \bar{\mathbf{b}}, \mathbf{u}, v_0)$ , dada por

$$V(z) = v_0 + \sum_{k=1}^{\infty} \mathbf{u}^\top \bar{\mathbf{A}}^{k-1} \bar{\mathbf{b}} z^k = v_0 + z \mathbf{u}^\top (\mathbf{I} - z \bar{\mathbf{A}})^{-1} \bar{\mathbf{b}}, \tag{3.5}$$

é uma função passa-tudo, ou seja,  $V(z)V(z^{-1}) = 1$ , para todo  $z$ .

### 3.1.2 Vetor de controlabilidade e complemento passa-tudo

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

$$\mathbf{x}(n+1) = \mathcal{C}(z)u(n), \tag{3.6}$$

com  $\mathcal{C}(z)$  dado por

$$\mathcal{C}(z) \triangleq \sum_{k=0}^{\infty} \mathbf{A}^k \mathbf{b} z^k = (\mathbf{I} - z \mathbf{A})^{-1} \mathbf{b}. \tag{3.7}$$

Como veremos a seguir,  $\mathcal{C}(z)$  e  $V(z)$  definido no item anterior tem uma relação de ortogonalidade importante. Antes porém, definimos a nomenclatura utilizada:

**Definição 3.1** *Dado um par  $(\mathbf{A}, \mathbf{b})$  completamente controlável, o vetor  $\mathcal{C}(z)$  de funções de transferência dado por (3.7) é denominado de vetor de controlabilidade de  $(\mathbf{A}, \mathbf{b})$ , e a função de transferência  $V(z)$  dada por (3.5) é denominada de complemento passa-tudo de  $\mathcal{C}(z)$ .*

Pode-se ver que, como  $\mathbf{A}$  e  $\overline{\mathbf{A}}$  tem os mesmos autovalores, cada linha de  $\mathcal{C}(z)$  tem os mesmos polos que  $V(z)$ . Por outro lado, para ser uma função passa-tudo, os zeros de  $V(z)$  devem ser recíprocos de seus polos [8]. Portanto, se cada função de  $\mathcal{C}(z)$  é da forma  $c_k(z)/A(z)$ , com  $A(z) = 1 + a_1z + \dots + a_M z^M$ , então  $V(z)$  é dado por

$$V(z) = \frac{z^M A(z^{-1})}{A(z)} = \frac{z^M + a_1 z^{M-1} + \dots + a_M}{1 + a_1 z^1 + \dots + a_M z^M}. \quad (3.8)$$

O vetor de controlabilidade de um sistema linear, por sua vez, depende da realização específica. No caso em que o par  $(\mathbf{A}, \mathbf{b})$  tem a forma de uma realização direta (canônica controlável), temos  $\mathbf{b} = [1 \ 0 \ 0 \ \dots \ 0]^\top$  e o vetor de controlabilidade é dado por

$$\mathcal{C}(z) = [1/A(z) \ z/A(z) \ \dots \ z^{M-1}/A(z)]^\top. \quad (3.9)$$

Conforme mostrado em [1] uma função  $f(z) \in \mathcal{H}_2$  é ortogonal a um vetor de controlabilidade  $\mathcal{C}(z)$  se, e somente se,  $f(z)$  for causalmente divisível pelo complemento passa-tudo de  $\mathcal{C}(z)$ :

**Propriedade 3.1** *Se  $\mathcal{C}(z)$  é um vetor de controlabilidade e  $f(z) \in \mathcal{H}_2$  então*

$$\langle \mathcal{C}(z), f(z) \rangle = \mathbf{0} \iff f(z) = g(z)V(z), \quad g(z) \in \mathcal{H}_2,$$

onde  $V(z)$  é o complemento passa-tudo de  $\mathcal{C}(z)$ .

Usando (1.16), (3.3) e (3.7) o gramiano de controlabilidade pode ser escrito agora como

$$\mathbf{K} = \mathcal{C} \mathcal{C}^\top = \langle \mathcal{C}(z), \mathcal{C}(z) \rangle.$$

Portanto, quando  $\mathbf{K} = \mathbf{I} = \mathbf{A} \mathbf{A}^\top + \mathbf{b} \mathbf{b}^\top$  as  $M$  funções de  $\mathcal{C}(z)$  são ortonormais. Se observarmos ainda que

$$\langle z^k V(z), z^l V(z) \rangle = \langle z^k, z^l V(z) V(z^{-1}) \rangle = \langle z^k, z^l \rangle = \begin{cases} 1, & k = l, \\ 0, & k \neq l, \end{cases}$$

decorre desta igualdade e da propriedade anterior a seguinte propriedade:

**Propriedade 3.2** *Dado um par  $(\mathbf{A}, \mathbf{b})$  completamente controlável tal que  $\mathbf{A} \mathbf{A}^\top + \mathbf{b} \mathbf{b}^\top = \mathbf{I}$ , as  $M$  funções do seu vetor de controlabilidade  $\mathcal{C}(z)$  mais o conjunto  $\{z^k V(z)\}_{k=0}^\infty$ , onde  $V(z)$  é o complemento passa-tudo de  $\mathcal{C}(z)$ , formam uma base ortonormal completa para o espaço  $\mathcal{H}_2$ .*

## 3.2 Análise da superfície de erro com entrada branca

### 3.2.1 Forma do erro de aproximação

Consideremos que a saída do sistema a ser identificado é dada por  $y(n) = H(z)u(n) + \zeta(n)$ , e a saída do filtro adaptativo por  $\hat{y}(n) = \hat{H}(z)u(n)$ . Para uma entrada  $u(n)$  branca de média nula e potência unitária e um ruído de medida  $\zeta(n)$  independente de  $u(n)$ , o erro quadrático médio de saída é dado por

$$E\{[y(n) - \hat{y}(n)]^2\} = \left\| H(z) - \hat{H}(z) \right\|^2 + E\{\zeta^2(n)\}. \quad (3.10)$$

Uma vez que a potência de ruído de medida  $E\{\zeta^2(n)\}$  não é afetada pelo filtro adaptativo  $\hat{H}(z)$ , podemos considerar apenas a minimização do erro quadrático médio de aproximação  $\left\| H(z) - \hat{H}(z) \right\|^2$  (no que segue, esta grandeza poderá ser referida, para maior concisão, simplesmente como "erro de aproximação").

Um ponto estacionário do erro de aproximação é definido como segue:

**Definição 3.2**  $\hat{H}(z)$  é um ponto estacionário de  $\left\| H(z) - \hat{H}(z) \right\|^2$  se para qualquer  $w_i$  do conjunto  $\{w_i\}$  de parâmetros a serem adaptados de  $\hat{H}(z)$  valer

$$\frac{\partial}{\partial w_i} \left\| H(z) - \hat{H}(z) \right\|^2 = 0. \quad (3.11)$$

Notar que uma condição necessária para que  $\hat{H}(z)$  seja um ponto de mínimo do erro de aproximação é que seja um ponto estacionário do mesmo. Pontos de máximo e pontos de sela, porém, também são pontos estacionários.

Para prosseguir na análise do erro de aproximação, consideremos que o filtro adaptativo  $\hat{H}(z)$  é implementado na forma direta:

**Definição 3.3** A função de transferência  $\hat{H}(z)$  é implementada na forma direta com  $2M+1$  coeficientes a serem adaptados dados por  $\{a_k\}_{k=1}^M$  e  $\{b_k\}_{k=0}^M$ :

$$\hat{H}(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1z + \dots + b_Mz^M}{1 + a_1z + \dots + a_Mz^M}. \quad (3.12)$$

Apesar desta restrição a  $\hat{H}(z)$ , como será visto, certas propriedades deduzidas para a forma direta podem ser extendidas a outras formas de implementação.

Com  $\hat{H}(z)$  dado pela Definição 3.3 e utilizando a Propriedade 3.1 e a Definição 1.1, as seguintes propriedades do erro de aproximação podem ser deduzidas [1]:

**Propriedade 3.3** Quando os polos de  $\hat{H}(z)$  dada pela Definição 3.3 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 seguinte forma:

$$H(z) - \hat{H}(z) = V(z)g(z), \quad z^{-1}g(z) \in \mathcal{H}_2, \quad (3.13)$$

para algum  $g(z)$ , com  $V(z)$  dado por (3.8).

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

**Propriedade 3.4** (Teorema de Walsh) Se  $\deg[\hat{H}(z)] = M$  então  $\hat{H}(z)$  dada pela Definição 3.3 é um ponto estacionário de  $\|H(z) - \hat{H}(z)\|^2$  se e somente se

$$H(z) - \hat{H}(z) = z[V(z)]^2Q(z), \quad (3.14)$$

para algum  $Q(z) \in \mathcal{H}_2$ .

Notar que com  $\hat{H}(z)$  dado por (3.12) a condição  $\deg[\hat{H}(z)] = M$  significa que nenhum zero de  $A(z)$  é cancelado por algum zero de  $B(z)$ .

Um aspecto de interesse para algoritmos de adaptação, em especial para aqueles baseados no gradiente do erro de saída, é o da existência de pontos estacionários da superfície de

erro. A forma do erro de aproximação estipulada pelo teorema acima permite relacionar a existência destes pontos estacionários com o grau do sistema  $H(z)$  e o do filtro adaptativo  $\hat{H}(z)$ , como visto a seguir.

Uma solução, "trivial", por assim dizer, de (3.14) é o caso em que  $Q(z) \equiv 0$  e portanto  $\hat{H}(z) = H(z)$ . Para determinar a existência de pontos estacionários com  $H(z) \neq \hat{H}(z)$  notamos inicialmente que o grau do lado esquerdo de (3.14) obedece a

$$\deg[H(z) - \hat{H}(z)] \leq \deg[H(z)] + \deg[\hat{H}(z)] = \deg[H(z)] + M \quad (3.15)$$

No lado direito de (3.14) podemos ver de (3.8) que se  $\hat{H}(z)$  é estável então  $V(z)$  tem exatamente  $M$  zeros dentro do círculo unitário e até  $M$  finitos polos fora do círculo unitário. Notar ainda que nenhum dos  $2M + 1$  zeros de  $z[V(z)]^2$  pode ser cancelado por um polo de  $Q(z)$  pois  $Q(z) \in \mathcal{H}_2$ . Devemos ter então

$$\deg[H(z) - \hat{H}(z)] \geq 2M + 1, \quad Q(z) \in \mathcal{H}_2, \quad Q(z) \neq 0 \quad (3.16)$$

Portanto, para que (3.15) e (3.16) sejam atendidas simultaneamente é necessário que  $\deg[H(z)] > M$ . Esta análise pode ser sintetizada na seguinte propriedade:

**Propriedade 3.5** *Se  $\hat{H}(z)$  é dada pela Definição 3.3 e  $\deg[\hat{H}(z)] = \deg[H(z)] = M$ , então o único ponto estacionário de  $\|H(z) - \hat{H}(z)\|^2$  é o seu mínimo global  $\hat{H}(z) = H(z)$ ; não existem pontos estacionários de  $\|H(z) - \hat{H}(z)\|^2$  com  $\deg[\hat{H}(z)] = M > \deg[H(z)]$ .*

É possível, ainda, extendendo a análise do erro de aproximação dado por (3.14), eliminar a exigência  $\deg[\hat{H}(z)] = M$  da propriedade acima e concluir que, com  $\hat{H}(z)$  dado por (3.12), se  $\deg[H(z)] \leq M$  então para todos os pontos estacionários de  $\|H(z) - \hat{H}(z)\|^2$  vale  $\hat{H}(z) = H(z)$ , ou seja todos são mínimos globais do erro de aproximação. Este resultado é conhecido há bastante tempo (ver [6] e [5]), porém o procedimento geral usado aqui para obtê-lo é diferente do procedimento seguido nestas referências. A demonstração deste resultado não será incluída aqui pois pode ser considerada como um caso particular do resultado obtido para o filtro adaptativo polifásico, a ser visto no capítulo 4.

### 3.2.2 Modelamento suficiente

A forma de  $\widehat{H}(z)$  dada por (3.12) não é a forma usualmente considerada na literatura sobre filtros IIR adaptativos (por exemplo, [6], [4] e [7]) pois se impõe aqui que o número de coeficientes do seu numerador seja igual ao número de coeficientes do seu denominador mais um. A forma usualmente considerada é mais genérica, sendo dada por

$$\widehat{H}(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1z + \dots + b_{M_b}z^{M_b}}{1 + a_1z + \dots + a_{M_a}z^{M_a}}. \quad (3.17)$$

Neste contexto, se o sistema sendo identificado é dado por

$$H(z) = \frac{C(z)}{D(z)} = \frac{c_0 + c_1z + \dots + c_{M_c}z^{M_c}}{1 + d_1z + \dots + d_{M_d}z^{M_d}}, \quad (3.18)$$

define-se em [4] uma situação na qual  $M_b \geq M_c$  e  $M_a \geq M_d$  como uma situação de "modelamento suficiente". Em [4] se conjecturava ainda que em uma condição de modelamento suficiente todos os pontos estacionários de  $\|H(z) - \widehat{H}(z)\|^2$  corresponderiam a  $\widehat{H}(z) = H(z)$ . Do resultado obtido em [5] pode-se concluir, porém, [7] que para tanto é necessária uma restrição adicional, e que uma restrição suficiente (porém não necessária) é

$$M_b \geq M_d - 1. \quad (3.19)$$

Como uma justificativa para a adoção da forma de  $\widehat{H}(z)$  dada por (3.12) notamos que a restrição acima é, no caso de "modelamento suficiente", sempre obedecida. Além disto, outras propriedades úteis da superfície de erro decorrem de se considerar  $\widehat{H}(z)$  dado por (3.12), conforme será visto posteriormente.

### 3.2.3 Superfície de erro reduzida

Uma forma útil de exprimir a função  $g(z)$  em (3.13) é obtida [1] re-escrevendo inicialmente esta expressão como

$$V(z^{-1})H(z) - V(z^{-1})\widehat{H}(z) = V(z^{-1})V(z)g(z) = g(z), \quad z^{-1}g(z) \in \mathcal{H}_2, \quad (3.20)$$



onde  $V(z^{-1})V(z) \equiv 1$  pois  $V(z)$  é passa-tudo. Consideremos a parcela

$$V(z^{-1})\widehat{H}(z) = \frac{z^{-M}A(z)B(z)}{A(z^{-1})A(z)} = \frac{z^{-M}B(z)}{A(z^{-1})} = \frac{b_M + b_{M-1}z^{-1} + \dots + b_0z^{-M}}{A(z^{-1})}. \quad (3.21)$$

Como discutido em relação a (1.15), se  $\widehat{H}(z)$  é estável  $1/A(z^{-1})$  é estável e anti-causal. Portanto a parcela  $V(z^{-1})\widehat{H}(z) = z^{-M}B(z)/A(z^{-1})$  também é anti-causal pois  $z^{-M}B(z)$  não tem potências positivas de  $z$ . Agora, como  $g(z)$  é estritamente causal, segue que  $g(z)$  é dado apenas pela projeção estritamente causal de  $V(z^{-1})H(z)$  [1]:

**Propriedade 3.6** *A função  $g(z)$  na Propriedade 3.3 pode ser escrita como*

$$g(z) = [V(z^{-1})H(z)]_+ \quad (3.22)$$

Da Propriedade 3.3 decorre que quando os zeros de  $\widehat{H}(z)$  são otimizados em função dos seus polos temos

$$\|H(z) - \widehat{H}(z)\|^2 = \|V(z)g(z)\|^2 = \|g(z)\|^2, \quad (3.23)$$

pois  $V(z)$  é passa-tudo. De onde segue a definição:

**Definição 3.4** *A superfície de erro reduzida de  $H(z) - \widehat{H}(z)$  é dada por  $\|g(z)\|^2$  como função dos parâmetros de  $A(z)$ , com  $g(z)$  dado pela Propriedade 3.6.*

Uma outra forma útil de exprimir  $g(z)$ [1] pode ser obtida notando que com  $g(z) = \sum_{k=1}^{\infty} g_k z^k$ ,  $V(z^{-1}) = \sum_{k=0}^{\infty} v_k z^{-k}$  e  $H(z) = \sum_{k=0}^{\infty} h_k z^k$  temos:

$$\begin{aligned} \sum_{k=1}^{\infty} g_k z^k &= \left[ \sum_{l=0}^{\infty} \sum_{k=0}^{\infty} v_l h_k z^{k-l} \right]_+ = \left[ \sum_{l=0}^{\infty} \sum_{k=-l}^{\infty} v_l h_{k+l} z^k \right]_+ \\ &= \sum_{l=0}^{\infty} \sum_{k=1}^{\infty} v_l h_{k+l} z^k = \sum_{k=1}^{\infty} \sum_{l=0}^{\infty} v_l h_{k+l} z^k. \end{aligned} \quad (3.24)$$

Portanto os coeficientes  $g_k$  de  $g(z)$  são dados por

$$g_k = \sum_{l=0}^{\infty} v_l h_{k+l}, \quad k = 1, 2, \dots, \quad (3.25)$$

o que pode ser expresso matricialmente por

$$\begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \\ \vdots \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 & h_4 & \cdots \\ h_2 & h_3 & h_4 & \cdots & \\ h_3 & h_4 & \cdots & & \\ h_4 & \vdots & & & \\ \vdots & & & & \end{bmatrix} \begin{bmatrix} v_0 \\ v_1 \\ v_2 \\ v_3 \\ \vdots \end{bmatrix}. \quad (3.26)$$

Na expressão acima, a forma matricial dos  $h_k$  é a forma de Hankel do sistema  $H(z)$ , denotada por  $\Gamma_H$ , e portanto uma expressão equivalente a (3.22) é

$$\mathbf{g} = \Gamma_H \mathbf{v}, \quad (3.27)$$

e uma expressão para a superfície de erro reduzida, equivalente a (3.23), é

$$\|H(z) - \hat{H}(z)\|^2 = \mathbf{g}^\top \mathbf{g} = \mathbf{v}^\top \Gamma_H^2 \mathbf{v}. \quad (3.28)$$

### 3.3 Outras formas de implementação

Neste item, propriedades do erro quadrático médio de aproximação apresentadas nos itens anteriores são extendidas para outras formas de implementação.

#### 3.3.1 Gradiente e Hessiana após mudança de variáveis

Na literatura referente a filtros IIR adaptativos as propriedades a seguir estão formuladas em [25] e [30]. Porém, a apresentação aqui é mais detalhada.

Seja  $\mathbf{w}$  um vetor de  $N$  elementos reais,

$$\mathbf{w} = \begin{bmatrix} w_1 & w_2 & \cdots & w_N \end{bmatrix}^\top, \quad (3.29)$$

e uma função  $f_w(\mathbf{w})$  de  $R^N$  em  $R$ . O vetor gradiente  $\nabla_w(\mathbf{w})$  de  $f_w$  é definido por

$$\nabla_w(\mathbf{w}) = \begin{bmatrix} \frac{\partial f_w}{\partial w_1} & \frac{\partial f_w}{\partial w_2} & \cdots & \frac{\partial f_w}{\partial w_N} \end{bmatrix}^\top, \quad (3.30)$$

e a matriz Hessiana  $\mathbf{H}_w(\mathbf{w})$  de  $f_w$  é definida por

$$\mathbf{H}_w(\mathbf{w}) = \begin{bmatrix} \frac{\partial^2 f_w}{\partial w_1^2} & \frac{\partial^2 f_w}{\partial w_1 \partial w_2} & \cdots & \frac{\partial^2 f_w}{\partial w_1 \partial w_N} \\ \frac{\partial^2 f_w}{\partial w_2 \partial w_1} & & & \vdots \\ \vdots & & & \\ \frac{\partial^2 f_w}{\partial w_N \partial w_1} & \cdots & & \frac{\partial^2 f_w}{\partial w_N \partial w_N} \end{bmatrix}. \quad (3.31)$$

Consideremos agora que  $\mathbf{w}$  é dado por

$$\mathbf{w} = g(\mathbf{p}) \triangleq \begin{bmatrix} g_1(\mathbf{p}) & g_2(\mathbf{p}) & g_N(\mathbf{p}) \end{bmatrix}^\top, \quad (3.32)$$

onde  $\mathbf{p}$  também é um vetor de  $N$  elementos reais e as funções  $g_l(\mathbf{p})$ ,  $l = 1, 2, \dots, N$ , são de  $R^N$  em  $R$  e possuem todas as derivadas parciais  $\partial g_l / \partial w_j$ . Consideremos ainda uma função  $f_p(\mathbf{p})$  tal que

$$f_p(\mathbf{p}) = f_w(g(\mathbf{p})). \quad (3.33)$$

Queremos agora exprimir o vetor gradiente  $\nabla_p(\mathbf{p})$  de  $f_p$  em função de  $\nabla_w(g(\mathbf{p}))$ . Devido à igualdade acima podemos escrever

$$\frac{\partial f_p(\mathbf{p})}{\partial p_j} = \frac{\partial f_w(g(\mathbf{p}))}{\partial p_j}. \quad (3.34)$$

Cada elemento do vetor gradiente  $\nabla_p(\mathbf{p})$  pode então ser calculado pela regra da cadeia como

$$\frac{\partial f_p}{\partial p_j} = \sum_{l=1}^N \frac{\partial f_w}{\partial w_l} \Big|_{\mathbf{w}=g(\mathbf{p})} \frac{\partial g_l}{\partial p_j} \quad j = 1, 2, \dots, N, \quad (3.35)$$

o que em forma matricial resulta:

$$\nabla_p(\mathbf{p}) = \begin{bmatrix} \frac{\partial f_p}{\partial p_1} \\ \frac{\partial f_p}{\partial p_2} \\ \vdots \\ \frac{\partial f_p}{\partial p_N} \end{bmatrix} = \begin{bmatrix} \frac{\partial g_1}{\partial p_1} & \frac{\partial g_2}{\partial p_1} & \cdots & \frac{\partial g_N}{\partial p_1} \\ \frac{\partial g_1}{\partial p_2} & & & \vdots \\ \vdots & & & \\ \frac{\partial g_1}{\partial p_N} & \cdots & & \frac{\partial g_N}{\partial p_N} \end{bmatrix} \begin{bmatrix} \frac{\partial f_w}{\partial w_1} \\ \frac{\partial f_w}{\partial w_2} \\ \vdots \\ \frac{\partial f_w}{\partial w_N} \end{bmatrix} \Big|_{\mathbf{w}=g(\mathbf{p})}$$

Ou seja, temos a seguinte relação entre os vetores gradiente:

$$\nabla_{\mathbf{p}}(g(\mathbf{p})) = \mathbf{J}^T \nabla_w(g(\mathbf{p})), \quad (3.36)$$

onde  $\mathbf{J}$  é a matriz Jacobiana de  $g(\mathbf{p})$ , para a qual, por simplicidade, não indicamos a dependência em relação a  $\mathbf{p}$ .

Queremos também exprimir a matriz Hessiana  $\mathbf{H}_p(\mathbf{p})$  de  $f_p$  em função de  $\mathbf{H}_w(g(\mathbf{p}))$  e  $\nabla_w(g(\mathbf{p}))$ . Para tanto, usando (3.35), cada elemento de  $\mathbf{H}_p(\mathbf{p})$  pode ser escrito como

$$\frac{\partial^2 f_p}{\partial p_j \partial p_k} = \frac{\partial}{\partial p_j} \sum_{l=1}^N \frac{\partial f_w}{\partial w_l} \Big|_{\mathbf{w}=g(\mathbf{p})} \frac{\partial g_l}{\partial p_k} \quad j, k = 1, 2, \dots, N. \quad (3.37)$$

Aplicando agora a regra da cadeia e a do produto temos:

$$\frac{\partial^2 f_p}{\partial p_j \partial p_k} = \sum_{l=1}^N \left[ \sum_{m=1}^N \frac{\partial^2 f_w}{\partial w_m \partial w_l} \Big|_{\mathbf{w}=g(\mathbf{p})} \frac{\partial g_m}{\partial p_j} \right] \frac{\partial g_l}{\partial p_k} + \sum_{l=1}^N \frac{\partial f_w}{\partial w_l} \Big|_{\mathbf{w}=g(\mathbf{p})} \frac{\partial^2 g_l}{\partial p_j \partial p_k}, \quad (3.38)$$

o que resulta, em forma matricial:

$$\begin{aligned} \mathbf{H}_p(\mathbf{p}) &= \begin{bmatrix} \frac{\partial^2 f_p}{\partial p_1^2} & \frac{\partial^2 f_p}{\partial p_1 \partial p_2} & \dots & \frac{\partial^2 f_p}{\partial p_1 \partial p_N} \\ \frac{\partial^2 f_p}{\partial p_2 \partial p_1} & & & \vdots \\ \vdots & & & \\ \frac{\partial^2 f_p}{\partial p_N \partial p_1} & \dots & & \frac{\partial^2 f_p}{\partial p_N^2} \end{bmatrix} = \\ &= \begin{bmatrix} \frac{\partial g_1}{\partial p_1} & \dots & \frac{\partial g_N}{\partial p_1} \\ \vdots & & \vdots \\ \frac{\partial g_1}{\partial p_N} & \dots & \frac{\partial g_N}{\partial p_N} \end{bmatrix} \begin{bmatrix} \frac{\partial^2 f_w}{\partial w_1^2} & \frac{\partial^2 f_w}{\partial w_1 \partial w_2} & \dots & \frac{\partial^2 f_w}{\partial w_1 \partial w_N} \\ \frac{\partial^2 f_w}{\partial w_2 \partial w_1} & & & \vdots \\ \vdots & & & \\ \frac{\partial^2 f_w}{\partial w_N \partial w_1} & \dots & & \frac{\partial^2 f_w}{\partial w_N \partial w_N} \end{bmatrix} \begin{bmatrix} \frac{\partial g_1}{\partial p_1} & \dots & \frac{\partial g_1}{\partial p_N} \\ \vdots & & \vdots \\ \frac{\partial g_N}{\partial p_1} & \dots & \frac{\partial g_N}{\partial p_N} \end{bmatrix} + \mathbf{R}, \end{aligned}$$

onde os elementos  $r_{i,j}$  de  $\mathbf{R}$  são dados por

$$r_{i,j} = \begin{bmatrix} \frac{\partial f_w}{\partial w_1} & \frac{\partial f_w}{\partial w_2} & \cdots & \frac{\partial f_w}{\partial w_N} \end{bmatrix} \underbrace{\begin{bmatrix} \frac{\partial^2 g_1}{\partial p_j \partial p_k} \\ \frac{\partial^2 g_2}{\partial p_j \partial p_k} \\ \vdots \\ \frac{\partial^2 g_m}{\partial p_j \partial p_k} \end{bmatrix}}_{\triangleq \delta_g(i,j)}$$

$$= \nabla_w^\top(g(\mathbf{p})) \delta_g(i,j).$$

Portanto, a relação entre as Hessianas é

$$\mathbf{H}_p(\mathbf{p}) = \mathbf{J}^\top \mathbf{H}_w(g(\mathbf{p})) \mathbf{J} + \mathbf{R}. \quad (3.39)$$

A partir das relações obtidas acima seguem as seguintes propriedades:

**Propriedade 3.7** *Sejam duas funções  $f_p(\mathbf{p})$  e  $f_w(\mathbf{w})$  com  $\mathbf{p}$  e  $\mathbf{w}$  pertencendo a certos domínios  $D_p$  e  $D_w$ , respectivamente, e tais que  $f_p(\mathbf{p}) = f_w(g(\mathbf{p}))$ , onde a matriz jacobiana  $\mathbf{J}$  de  $g(\mathbf{p})$  existe e é não-singular para todo  $\mathbf{p} \in D_p$ . Então, em um ponto  $\mathbf{p}_0 \in D_p$ , temos que  $\nabla_p(\mathbf{p}_0) = \mathbf{0}$  se, e somente se,  $\nabla_w(g(\mathbf{p}_0)) = \mathbf{0}$ .*

*Prova:* Se  $f_p(\mathbf{p}) = f_w(g(\mathbf{p}))$  então conforme (3.36) temos  $\nabla_p(\mathbf{p}) = \mathbf{J}^\top \nabla_w(g(\mathbf{p}))$ . Se  $\mathbf{J}$  é não-singular segue diretamente a propriedade desejada. ■

**Propriedade 3.8** *Nas condições da propriedade anterior, para um ponto  $\mathbf{p}_0$  no qual  $\nabla_p(\mathbf{p}_0) = \mathbf{0}$  temos que  $\mathbf{H}_p(\mathbf{p}_0)$  é positiva(negativa) definida se, e somente se,  $\mathbf{H}_w(g(\mathbf{p}_0))$  for positiva(negativa) definida.*

*Prova:* Se  $\mathbf{H}_p(\mathbf{p}_0)$  é positiva definida então para qualquer vetor  $\mathbf{x} \neq \mathbf{0}$  temos  $\mathbf{x}^\top \mathbf{H}_p(\mathbf{p}_0) \mathbf{x} > 0$ . Da propriedade anterior temos que se  $\nabla_p(\mathbf{p}_0) = \mathbf{0}$  então  $\nabla_w^\top(g(\mathbf{p}_0)) = \mathbf{0}$  e, consequentemente, usando (3.39) temos  $\mathbf{H}_p(\mathbf{p}_0) = \mathbf{J}^\top \mathbf{H}_w(g(\mathbf{p}_0)) \mathbf{J}$ . Portanto

$$0 < \mathbf{x}^\top \mathbf{H}_p(\mathbf{p}_0) \mathbf{x} = \mathbf{x}^\top \mathbf{J}^\top \mathbf{H}_w(g(\mathbf{p}_0)) \mathbf{J} \mathbf{x} = \mathbf{y}^\top \mathbf{H}_w(g(\mathbf{p}_0)) \mathbf{y},$$

e como  $\mathbf{y} = \mathbf{J}\mathbf{x} \neq \mathbf{0}$  para qualquer  $\mathbf{x} \neq \mathbf{0}$ , pois  $\mathbf{J}$  é não-singular, temos que  $\mathbf{H}_w(g(\mathbf{p}_0))$  é positiva definida. As provas para  $\mathbf{H}_p(\mathbf{p}_0)$  negativa definida e para os sentidos inversos seguem analogamente. ■

### 3.3.2 Aproximação local da superfície de erro

Consideremos agora que  $\mathbf{w}$  é o vetor de parâmetros do filtro adaptativo  $\hat{H}(z)$  e que a função  $f_w(\mathbf{w})$  considerada acima é o erro de aproximação

$$f_w(\mathbf{w}) = \left\| H(z) - \hat{H}(z) \right\|^2.$$

Consideremos ainda que em torno de um ponto qualquer  $\mathbf{w}_0$  do espaço de parâmetros a superfície de erro possa ser aproximada por uma superfície quadrática em  $\Delta\mathbf{w} = \mathbf{w} - \mathbf{w}_0$  como

$$f(\mathbf{w}) \simeq f(\mathbf{w}_0) + \Delta\mathbf{w}^\top \nabla_w(\mathbf{w}_0) + \frac{1}{2} \Delta\mathbf{w}^\top \mathbf{H}_w(\mathbf{w}_0) \Delta\mathbf{w}, \quad (3.40)$$

onde  $\nabla_w(\mathbf{w})$  e  $\mathbf{H}_w(\mathbf{w})$  são o vetor gradiente e a matriz Hessiana definidos em (3.30) e (3.31).

Se  $\mathbf{w}_0$  é um ponto estacionário da superfície de erro (Definição 3.2) temos  $\nabla_w(\mathbf{w}_0) = \mathbf{0}$  e as características locais da superfície de erro são determinadas pela matriz Hessiana  $\mathbf{H}_w(\mathbf{w}_0)$ : se  $\mathbf{H}_w(\mathbf{w}_0)$  é positiva definida então  $\mathbf{w}_0$  é um ponto de mínimo; se for negativa definida então  $\mathbf{w}_0$  é um ponto de máximo e se não for definida  $\mathbf{w}_0$  é um ponto de sela, isto é, um ponto de máximo em relação a alguns parâmetros e um ponto de mínimo em relação aos demais.

A mesma aproximação de segunda ordem pode ser feita para a superfície de erro reduzida (Definição 3.4), devendo-se notar que os pontos estacionários da mesma não podem ser pontos de máximo, só pontos de mínimo ou pontos de sela, uma vez que por definição os parâmetros que definem os zeros de  $\hat{H}(z)$  estão em um ponto de mínimo.

### 3.3.3 Extensão de propriedades do erro de aproximação

Vejamos agora como as definições e propriedades relativas ao erro de aproximação, deduzidas anteriormente para a realização na forma direta, são extendidas para outras formas de realização.

Na terminologia do item anterior,  $\mathbf{w}$  são os parâmetros de uma realização na forma direta do filtro adaptativo  $\hat{H}(z)$ ;  $\mathbf{p}$  são os parâmetros de uma outra forma de realização relacionada

com  $\mathbf{w}$  por  $\mathbf{w} = g(\mathbf{p})$ , com  $g$  tal que a matriz jacobiana  $\mathbf{J}$  de  $g(\mathbf{p})$  existe e é não-singular;  $f_w$  e  $f_p$  são o erro de aproximação  $\|H(z) - \hat{H}(z)\|^2$  em cada realização, e é evidente que, sendo realizações equivalentes, temos  $f_p(\mathbf{p}) = f_w(g(\mathbf{p}))$ .

Deve ser ressaltado que nem toda realização equivalente a uma realização na forma direta possui parâmetros relacionados com os da forma direta por uma função com as características exigidas acima para a função  $g(\mathbf{p})$ . De imediato vemos que para realizações não-mínimas, isto é, realizações com um número maior de parâmetros do que a realização direta, a matriz  $\mathbf{J}$  não é quadrada e portanto é singular. Porém, mesmo realizações mínimas podem não se enquadrar nas condições para a função  $g(\mathbf{p})$ . Para as realizações nas formas paralela e cascata, por exemplo, o jacobiano  $\mathbf{J}$  não existe em todos os pontos do domínio de estabilidade do espaço de parâmetros, devido à não-unicidade da transformação entre estas formas e a realização direta [25].

Feitas estas ressalvas, e usando as Propriedades 3.7 e 3.8 podemos constatar que: 1)  $\hat{H}(z)$  é um ponto estacionário (Definição 3.2) do erro de aproximação na realização equivalente se, e somente se, for um ponto estacionário do erro de aproximação na realização na forma direta; 2) as Propriedades do erro de aproximação 3.3, 3.4 e 3.5 valem para a realização equivalente; 3) Um ponto estacionário tem a mesma natureza (ponto de mínimo, de máximo ou de sela) na realização direta e na realização equivalente.

## 3.4 Análise SVD da superfície de erro reduzida

### 3.4.1 Decomposição SVD de $\|g(z)\|$

De (3.27) temos que a superfície de erro reduzida pode ser escrita como

$$\begin{aligned} \|g(z)\|^2 &= \mathbf{g}^\top \mathbf{g} = \mathbf{v}^\top \Gamma_H^\top \Gamma_H \mathbf{v} \\ &= \mathbf{v}^\top \Gamma_H \Gamma_H^\top \mathbf{v}, \end{aligned} \quad (3.41)$$

onde  $\Gamma_H$  é a forma de Hankel de  $H(z)$  e  $\mathbf{v}$  é o vetor formado pelos coeficientes da série de potência de  $V(z)$ . Se  $\deg[H(z)] = M$  então  $\Gamma_H$  pode ser escrita em função dos seus valores

singulares  $\sigma_k$  e pares de Schmidt  $(\zeta_k, \eta_k)$  como [29]

$$\Gamma_H = \sum_{k=1}^M \zeta_k \sigma_k \eta_k^\top, \quad (3.42)$$

onde  $\sigma_k > 0$  e  $\{\eta_k\}$  e  $\{\zeta_k\}$  são ortonormais ( $\eta_k^\top \eta_l = \zeta_k^\top \zeta_l = \delta_{k,l}$ ). Resulta portanto que

$$\begin{aligned} \|g(z)\|^2 &= \sum_{k=1}^M \sum_{l=1}^M \mathbf{v}^\top \zeta_k \sigma_k \underbrace{\eta_k^\top \eta_l}_{\delta_{k,l}} \sigma_l \zeta_l^\top \mathbf{v} \\ &= \sum_{k=1}^M \mathbf{v}^\top \zeta_k \sigma_k^2 \zeta_k^\top \mathbf{v}. \end{aligned} \quad (3.43)$$

A partir da expressão acima podemos agora voltar à representação em  $z$ , a superfície de erro reduzida sendo escrita então como

$$\|g(z)\|^2 = \sum_{k=1}^M \sigma_k^2 \langle \zeta_k(z), V(z) \rangle^2. \quad (3.44)$$

Notar que, como  $\zeta_k(z)$  e  $V(z)$  tem norma unitária, então  $\langle \zeta_k(z), V(z) \rangle \leq 1$ . As funções  $\zeta_k(z)$  são dadas por

$$\zeta_k(z) = \sigma_k^{-1/2} \mathbf{e}_k (\mathbf{I} - z\mathbf{A})^{-1} \mathbf{b}_b, \quad k = 1, 2, \dots, M, \quad (3.45)$$

onde  $\mathbf{e}_k$  é um vetor linha de zeros com exceção de um 1 na coluna  $k$ :

$$\mathbf{e}_k \triangleq [0 \quad \dots \quad 0 \quad \underset{\substack{\uparrow \\ \text{coluna } k}}{1} \quad 0 \quad \dots \quad 0], \quad (3.46)$$

e  $\mathbf{b}_b$  é obtido de uma realização balanceada de  $H(z)$  [29]. Comparando a expressão acima com (3.7) resulta que podemos escrever ainda

$$\zeta_k(z) = \sigma_k^{-1/2} C_{b,k}(z), \quad k = 1, 2, \dots, M, \quad (3.47)$$

onde

$$C_{b,k}(z) \triangleq \mathbf{e}_k (\mathbf{I} - z\mathbf{A})^{-1} \mathbf{b}_b \quad (3.48)$$



é a  $k$ -ésima linha do vetor de funções de controlabilidade de uma realização balanceada de  $H(z)$ . De (3.45) decorre que, com  $H(z)$  dado por

$$H(z) = \frac{C(z)}{D(z)} = \frac{c_0 + c_1z + \dots + c_Mz^M}{d_0 + d_1z + \dots + d_Mz^M}, \quad (3.49)$$

onde  $C(z)$  e  $D(z)$  não tem raízes comuns,  $\zeta_k(z)$  é da forma

$$\zeta_k(z) = \frac{\zeta_{k,0} + \zeta_{k,1}z + \dots + \zeta_{k,M-1}z^{M-1}}{D(z)}, \quad (3.50)$$

A expansão de  $\|g(z)\|^2$  dada por (3.44) será utilizada a seguir, assim como a propriedade:

**Propriedade 3.9** Se  $F(z)$  e  $q(z)$  são funções racionais estáveis e causais com  $F(z)$  dada por

$$F(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1z + \dots + b_{M-1}z^{M-1}}{1 + a_1z + \dots + a_Mz^M} \quad (3.51)$$

e  $V(z)$  uma função passa-tudo obtida de  $F(z)$ :

$$V(z) = \frac{z^M A(z^{-1})}{A(z)}, \quad (3.52)$$

então o produto interno  $\langle F(z), q(z)V(z) \rangle$  é nulo.

*Prova:* Conforme (1.16) e (1.17) o produto interno  $\langle F(z), q(z)V(z) \rangle$  é dado por

$$\begin{aligned} \langle F(z), q(z)V(z) \rangle &= \frac{1}{2\pi j} \oint_{|z|=1} q(z)V(z)F(z^{-1})\frac{dz}{z} \\ &= \frac{1}{2\pi j} \oint_{|z|=1} q(z) \frac{z^M A(z^{-1})}{A(z)} \frac{B(z^{-1})}{A(z^{-1})} \frac{dz}{z} \\ &= \frac{1}{2\pi j} \oint_{|z|=1} zq(z) \frac{b_0z^{M-1} + \dots + b_{M-1}}{A(z)} \frac{dz}{z} \\ &= \frac{1}{2\pi j} \oint_{|z|=1} q(z) \frac{b_0z^{M-1} + \dots + b_{M-1}}{A(z)} dz = 0. \end{aligned} \quad (3.53)$$

A última igualdade decorre do fato de que  $F(z)$  e  $q(z)$  são estáveis e causais e portanto os zeros de  $A(z)$  e os polos de  $q(z)$  estão fora do círculo unitário (ver item 1.5.1) e o resíduo do integrando é nulo.

Com a propriedade acima, podemos ver que, como seria de se esperar, com o sistema sendo identificado dado por (3.49), quando  $V(z)$ , dado por (3.52), é tal que  $A(z) = D(z)$ , o que corresponde ao mínimo global  $\hat{H}(z) = H(z)$ , o erro dado por (3.44) é nulo.

### 3.4.2 Forma da Hessiana no mínimo global

Consideremos que na superfície de erro reduzida o erro de aproximação é função de um vetor de parâmetros  $\mathbf{w} = [w_1 \ w_2 \ \dots \ w_M]^\top$  associado ao tipo de realização empregado para o filtro adaptativo. Indicamos explicitamente esta dependência usando a expressão

$$g_w(\mathbf{w}) \triangleq \|g(z, \mathbf{w})\|^2 \quad (3.54)$$

para o erro. Analogamente ao que foi visto no Item 3.3, consideramos que em torno de um ponto qualquer  $\mathbf{w}_0$  do espaço de parâmetros a superfície de erro reduzida pode ser aproximada por uma superfície quadrática em  $\Delta \mathbf{w} = \mathbf{w} - \mathbf{w}_0$  como

$$g_w(\mathbf{w}) \simeq g_w(\mathbf{w}_0) + \Delta \mathbf{w}^\top \nabla_w(\mathbf{w}_0) + \frac{1}{2} \Delta \mathbf{w}^\top \mathbf{H}_w(\mathbf{w}_0) \Delta \mathbf{w}, \quad (3.55)$$

onde  $\nabla_w(\mathbf{w}_0) = \left[ \frac{\partial g_w}{\partial w_1} \ \frac{\partial g_w}{\partial w_2} \ \dots \ \frac{\partial g_w}{\partial w_M} \right]^\top \Big|_{\mathbf{w}_0}$  é o gradiente de  $g_w$  em  $\mathbf{w}_0$  e

$$\mathbf{H}_w(\mathbf{w}_0) = \left\{ \frac{\partial^2 g_w}{\partial w_i \partial w_j} \Big|_{\mathbf{w}_0} \right\}_{i,j} \quad (3.56)$$

é a matriz Hessiana de  $g_w$  em  $\mathbf{w}_0$ . A seguir analisaremos a forma assumida pela matriz Hessiana  $\mathbf{H}_w$  no mínimo global.

As derivadas dos termos da decomposição SVD do erro dada por (3.44) são

$$\frac{\partial}{\partial w_i} \langle \zeta_k(z), V(z) \rangle^2 = 2 \langle \zeta_k(z), V(z) \rangle \left\langle \zeta_k(z), \frac{\partial}{\partial w_i} V(z) \right\rangle, \quad i, k = 1, 2, \dots, M, \quad (3.57)$$

e as derivadas segundas são dadas por

$$\begin{aligned} \frac{\partial}{\partial w_i} \frac{\partial}{\partial w_j} \langle \zeta_k(z), V(z) \rangle^2 &= \frac{\partial}{\partial w_i} 2 \langle \zeta_k(z), V(z) \rangle \left\langle \zeta_k(z), \frac{\partial}{\partial w_j} V(z) \right\rangle \\ &= 2 \left\langle \zeta_k(z), \frac{\partial}{\partial w_i} V(z) \right\rangle \left\langle \zeta_k(z), \frac{\partial}{\partial w_j} V(z) \right\rangle + \end{aligned}$$

$$+2 \langle \zeta_k(z), V(z) \rangle \left\langle \zeta_k(z), \frac{\partial}{\partial w_i} \frac{\partial}{\partial w_j} V(z) \right\rangle. \quad (3.58)$$

Lembramos que, com a função de transferência do filtro adaptativo sendo  $\widehat{H}(z) = B(z)/A(z)$ , o complemento passa tudo  $V(z)$  é dado por  $V(z) = z^M A(z^{-1})/A(z)$ . No mínimo global  $\widehat{H}(z) = H(z)$  decorre de (3.50) e do lema 3.9 que o termo  $\langle \zeta_k(z), V(z) \rangle$  da expressão acima é nulo. Temos portanto:

$$\frac{\partial}{\partial w_i} \frac{\partial}{\partial w_j} \langle \zeta_k(z), V(z) \rangle^2 \Big|_{\mathbf{w}_0} = 2 \left\langle \zeta_k(z), \frac{\partial}{\partial w_i} V(z) \Big|_{\mathbf{w}_0} \right\rangle \left\langle \zeta_k(z), \frac{\partial}{\partial w_j} V(z) \Big|_{\mathbf{w}_0} \right\rangle, \quad (3.59)$$

onde  $i, j$  e  $k$  variam de 1 a  $M$ .

Escrevendo  $V(z)$  na forma racional com  $\bar{A}(z) \triangleq z^M A(z^{-1})$ , os produtos internos acima são dados por

$$\begin{aligned} \left\langle \zeta_k(z), \frac{\partial}{\partial w_i} V(z) \Big|_{\mathbf{w}_0} \right\rangle &= \left\langle \zeta_k(z), \frac{\partial}{\partial w_i} \bar{A}(z) \Big|_{\mathbf{w}_0} \right\rangle \\ &= \left\langle \zeta_k(z), \frac{\frac{\partial \bar{A}(z)}{\partial w_i} A(z) - \frac{\partial A(z)}{\partial w_i} \bar{A}(z)}{A^2(z)} \Big|_{\mathbf{w}_0} \right\rangle \\ &= \left\langle \zeta_k(z), \frac{\partial \bar{A}(z)/\partial w_i}{A(z)} \Big|_{\mathbf{w}_0} - \frac{\partial A(z)/\partial w_i}{A(z)} V(z) \Big|_{\mathbf{w}_0} \right\rangle \\ &= \left\langle \zeta_k(z), \frac{\partial \bar{A}(z)/\partial w_i}{A(z)} \Big|_{\mathbf{w}_0} \right\rangle - \left\langle \zeta_k(z), \frac{\partial A(z)/\partial w_i}{A(z)} V(z) \Big|_{\mathbf{w}_0} \right\rangle \end{aligned} \quad (3.60)$$

Novamente temos um termo,  $\left\langle \zeta_k(z), \frac{\partial A(z)/\partial w_i}{A(z)} V(z) \right\rangle$ , que devido ao lema (3.9) se anula no mínimo global. Chegamos então a

$$\left\langle \zeta_k(z), \frac{\partial}{\partial w_i} V(z) \Big|_{\mathbf{w}_0} \right\rangle = \left\langle \zeta_k(z), \frac{\partial \bar{A}(z)/\partial w_i}{A(z)} \Big|_{\mathbf{w}_0} \right\rangle. \quad (3.61)$$

Reunindo (3.44), (3.54), (3.56), (3.59) e (3.61) temos que a Hessiana no mínimo global

$\mathbf{w}_0$  é dada por

$$\mathbf{H}_w(\mathbf{w}_0) = \begin{bmatrix} \delta_{11} & \delta_{12} & \cdots & \delta_{1M} \\ \delta_{21} & & & \\ \vdots & & & \vdots \\ \delta_{M1} & \cdots & & \delta_{MM} \end{bmatrix}, \quad (3.62)$$

onde

$$\begin{aligned} \delta_{ij} &= \sum_{k=1}^M \sigma_k^2 \frac{\partial}{\partial w_i} \frac{\partial}{\partial w_j} \langle \zeta_k(z), V(z) \rangle^2 \Big|_{\mathbf{w}_0} \\ &= \sum_{k=1}^M 2\sigma_k^2 \left\langle \zeta_k(z), \frac{\partial \bar{A}(z)/\partial w_i}{A(z)} \Big|_{\mathbf{w}_0} \right\rangle \left\langle \zeta_k(z), \frac{\partial \bar{A}(z)/\partial w_j}{A(z)} \Big|_{\mathbf{w}_0} \right\rangle. \end{aligned} \quad (3.63)$$

Definindo agora

$$\mathbf{d}_k \triangleq \left[ \left\langle \zeta_k(z), \frac{\partial \bar{A}(z)/\partial w_1}{A(z)} \Big|_{\mathbf{w}_0} \right\rangle, \dots, \left\langle \zeta_k(z), \frac{\partial \bar{A}(z)/\partial w_M}{A(z)} \Big|_{\mathbf{w}_0} \right\rangle \right]^\top, \quad (3.64)$$

$$\mathbf{D} \triangleq [\mathbf{d}_1 \ \mathbf{d}_2 \ \dots \ \mathbf{d}_M], \quad (3.65)$$

e a matriz diagonal

$$\Sigma \triangleq \begin{bmatrix} 2\sigma_1^2 & 0 & \cdots & 0 \\ 0 & 2\sigma_2^2 & 0 & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 2\sigma_M^2 \end{bmatrix}, \quad (3.66)$$

podemos estabelecer a seguinte identidade:

**Identidade 3.1** No ponto de mínimo global  $\hat{H}(z) = H(z) = C(z)/D(z)$  a matriz Hessiana da superfície de erro reduzida é dada por

$$\mathbf{H}_w(\mathbf{w}_0) = \mathbf{D}\Sigma\mathbf{D}^\top, \quad (3.67)$$

onde  $\mathbf{D}$  é dada por (3.64) e (3.65) e  $\Sigma$  é formada a partir dos valores singulares de  $H(z)$ , conforme (3.66).

Podemos ver que nesta forma de escrever a matriz Hessiana há uma separação parcial dos efeitos do sistema sendo identificado e do tipo de realização do filtro adaptativo: a matriz  $\Sigma$

só depende do sistema sendo identificado; nos termos  $\mathbf{d}_k$  da matriz  $\mathbf{D}$  as funções  $\zeta_k(z)$  são obtidas de uma realização balanceada de  $H(z)$  e portanto só dependem do sistema sendo identificado, ao passo que os termos  $\left. \frac{\partial \bar{A}(z)/\partial w_i}{A(z)} \right|_{\mathbf{w}_0}$  dependem também do tipo de realização do filtro adaptativo. Como será visto no Item a seguir esta forma de escrever a matriz Hessiana sugere uma maneira de eliminar a parcela do espalhamento dos autovalores da mesma associada à forma de implementação do filtro adaptativo.

## 3.5 Diagonalização da Hessiana

### 3.5.1 Realização direta

Consideremos inicialmente o caso de uma realização direta do filtro adaptativo  $\hat{H}(z)$ , de modo que os parâmetros  $\mathbf{w}$  considerados no item anterior são os próprios coeficientes de  $A(z)$ . No cálculo dos vetores  $\mathbf{d}_k$  em (3.64) as derivadas utilizadas são então

$$\begin{aligned} \left. \frac{\partial \bar{A}(z)/\partial w_i}{A(z)} \right|_{\mathbf{w}_0} &= \left. \frac{\partial \bar{A}(z)/\partial a_i}{A(z)} \right|_{\mathbf{w}_0} = \left. \frac{\frac{\partial}{\partial a_i}(z^M + a_1 z^{M-1} + \dots + a_M)}{A(z)} \right|_{\mathbf{w}_0} \\ &= \frac{z^{M-i}}{D(z)}, \quad i = 1, 2, \dots, M. \end{aligned} \quad (3.68)$$

Por outro lado, no mínimo global, o vetor de controlabilidade da implementação direta (ver (3.9)) é

$$\mathcal{C}(z, \mathbf{w}_0) = \left[ \frac{1}{D(z)} \quad \frac{z}{D(z)} \cdots \frac{z^{M-1}}{D(z)} \right]^T. \quad (3.69)$$

Se definirmos  $\bar{\mathcal{C}}(z)$  como o vetor obtido invertendo a ordem das linhas de  $\mathcal{C}(z)$ , temos

$$\bar{\mathcal{C}}(z, \mathbf{w}_0) = \left[ \frac{z^{M-1}}{D(z)} \quad \frac{z^{M-2}}{D(z)} \cdots \frac{1}{D(z)} \right]^T \quad (3.70)$$

e

$$\left. \frac{\partial \bar{A}(z)/\partial a_i}{A(z)} \right|_{\mathbf{w}_0} = \bar{\mathcal{C}}_i(z, \mathbf{w}_0), \quad (3.71)$$

onde  $\bar{\mathcal{C}}_i(z) = \mathbf{e}_i \bar{\mathcal{C}}(z)$  é a  $i$ -ésima linha de  $\bar{\mathcal{C}}(z)$  ( $\mathbf{e}_i$  tem a forma dada por (3.46)). Reunindo a expressão acima com (3.47), resulta em que os produtos internos que compõe os vetores  $\mathbf{d}_k$

em (3.64) podem então ser escritos como

$$\left\langle \zeta_k(z), \frac{\partial \bar{A}(z)/\partial a_i}{A(z)} \Big|_{\mathbf{w}_0} \right\rangle = \sigma_k^{-1/2} \langle \mathcal{C}_{b,k}(z), \bar{\mathcal{C}}_i(z, \mathbf{w}_0) \rangle. \quad (3.72)$$

A expressão acima pode ser usada para o cálculo numérico da matriz  $\mathbf{D}$ . Além disso ela é útil no sentido de sugerir procedimentos para a diagonalização da matriz Hessiana da superfície de erro reduzida, como visto a seguir.

### 3.5.2 Uma realização hipotética ideal

O fato das derivadas  $\frac{\partial \bar{A}(z)/\partial w_i}{A(z)}$  serem as próprias linhas do vetor de controlabilidade da realização sendo empregada (conforme (3.71)) é uma particularidade da realização direta que não necessariamente ocorre para outras realizações. Suponhamos porém que fosse possível parametrizar uma realização balanceada de tal modo que isto ocorresse. No mínimo global  $\mathbf{w}_0$  teríamos então,

$$\bar{\mathcal{C}}_i(z, \mathbf{w}_0) = \bar{\mathcal{C}}_{b,i}(z), \quad (3.73)$$

onde  $\bar{\mathcal{C}}_{b,i}(z)$  é o vetor obtido invertendo a ordem das linhas do vetor de controlabilidade  $\mathcal{C}_{b,i}(z)$  de uma realização balanceada do sistema  $H(z)$  sendo identificado dado por (3.48). Devido à ortonormalidade das funções de  $\mathcal{C}_b(z)$  teríamos

$$\left\langle \zeta_k(z), \frac{\partial \bar{A}(z)/\partial w_i}{A(z)} \Big|_{\mathbf{w}_0} \right\rangle = \sigma_k^{-1/2} \langle \mathcal{C}_{b,k}(z), \bar{\mathcal{C}}_{b,i}(z) \rangle = \begin{cases} \sigma_k^{-1/2}, & k = M - i + 1 \\ 0, & k \neq M - i + 1 \end{cases}. \quad (3.74)$$

A matriz  $\mathbf{D}$  seria então anti-diagonal:

$$\mathbf{D} \triangleq \begin{bmatrix} 0 & 0 & \dots & \sigma_M^{-1/2} \\ 0 & & \dots & \vdots \\ \vdots & \sigma_2^{-1/2} & & 0 \\ \sigma_1^{-1/2} & 0 & \dots & 0 \end{bmatrix}, \quad (3.75)$$

e a Hessiana seria diagonal, determinada apenas pelos valores singulares de  $H(z)$ :

$$\mathbf{H}_g = \mathbf{D}\Sigma\mathbf{D}^\top = \begin{bmatrix} 2\sigma_M & 0 & \dots & 0 \\ 0 & 2\sigma_{M-1} & 0 & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & 2\sigma_1 \end{bmatrix}. \quad (3.76)$$

Ou seja, para esta realização hipotética o espalhamento dos autovalores da matriz Hessiana no mínimo global dependeria apenas do espalhamento dos valores singulares do sistema sendo identificado. A parametrização de uma realização balanceada de um filtro, porém, não é um problema resolvido na teoria de síntese de redes [1], e, portanto, a existência da realização hipotética considerada aqui não pode ser aqui afirmada. Mesmo assim, ela é útil para fins de comparação, isto é, como uma forma de realização ideal em relação à qual outras formas de realização se aproximam mais ou menos.

### 3.6 Sistemas com polos próximos ao círculo unitário

Neste item mostra-se que quando o sistema sendo identificado possui polos próximos ao círculo unitário a superfície de erro reduzida é fortemente não quadrática: longe do mínimo global apresenta regiões planas porém de erro elevado, e, em consequência, em torno do mínimo global, há uma decréscimo brusco do erro. Esta característica é bastante desfavorável para algoritmos de passo constante, pois o limite para o valor do passo de adaptação imposto pelo comportamento da superfície de erro próximo ao mínimo global implica em uma convergência muito lenta nas regiões planas.

Se o sistema sendo identificado e o filtro adaptativo são dados por

$$H(z) = \frac{C(z)}{D(z)} \quad e \quad \hat{H}(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1z + \dots + b_Mz^M}{1 + a_1z + \dots + a_Mz^M}, \quad (3.77)$$

quando os zeros de  $\hat{H}(z)$  são otimizados, temos, conforme (3.22),

$$H(z) - \hat{H}(z) = g(z)V(z), \quad (3.78)$$

com

$$g(z) = [V(z^{-1})H(z)]_+, \quad (3.79)$$

onde  $V(z)$  é o passa-tudo dado por

$$V(z) = \frac{z^M A(z^{-1})}{A(z)}. \quad (3.80)$$

Consideremos para maior simplicidade o caso em que  $H(z)$  não tem polos com multiplicidade maior que 1 e que  $\deg[C(z)] \leq \deg[D(z)] = M$ . Nestas condições  $H(z)$  sempre pode ser expandido em resíduos como

$$H(z) = K_0 + \sum_{i=1}^M \frac{r_i}{z - \nu_i}. \quad (3.81)$$

Usando esta expressão e notando que  $V(z^{-1})$  é anti-causal temos que  $g(z)$  pode ser escrita como

$$\begin{aligned} g(z) &= \left[ K_0 V(z^{-1}) + V(z^{-1}) \sum_{i=1}^M \frac{r_i}{z - \nu_i} \right]_+ \\ &= \left[ V(z^{-1}) \sum_{i=1}^M \frac{r_i}{z - \nu_i} \right]_+, \end{aligned} \quad (3.82)$$

pois o termo  $K_0 V(z^{-1})$  não tem potências positivas de  $z$ .

Por sua vez,  $V(z^{-1})$  pode ser escrito como

$$V(z^{-1}) = \frac{z^{-M} A(z)}{A(z^{-1})} = \frac{A(z)}{z^M A(z^{-1})}, \quad (3.83)$$

o que mostra que possui exatamente  $M$  polos dentro do círculo unitário. O termo

$$V(z^{-1}) \sum_{i=1}^M \frac{r_i}{z - \nu_i} \quad (3.84)$$

pode então ser expandido em resíduos como

$$V(z^{-1}) \sum_{i=1}^M \frac{r_i}{z - \nu_i} = \sum_{i=1}^M \frac{s_i}{z - q_i} + \sum_{i=1}^M \frac{V(\nu_i^{-1}) r_i}{z - \nu_i} + r_0, \quad (3.85)$$

onde  $q_i$  são os polos de  $V(z^{-1})$ . Uma vez que os mesmos estão dentro do círculo unitário, o



termo

$$\sum_{i=1}^M \frac{s_i}{z - q_i} \quad (3.86)$$

é anti-causal. Introduzindo a expressão acima em (3.82) temos

$$\begin{aligned} g(z) &= \left[ \sum_{i=1}^M \frac{s_i}{z - q_i} + \sum_{i=1}^M \frac{V(\nu_i^{-1})r_i}{z - \nu_i} + r_0 \right]_+ \\ &= \left[ \sum_{i=1}^M \frac{V(\nu_i^{-1})r_i}{z - \nu_i} \right]_+. \end{aligned} \quad (3.87)$$

Para eliminar agora o operador projeção causal da expressão acima fazemos

$$\begin{aligned} g(z) &= \left[ \sum_{i=1}^M \frac{V(\nu_i^{-1})r_i}{z - \nu_i} \right]_+ = \sum_{i=1}^M \left[ \frac{V(\nu_i^{-1})r_i}{z - \nu_i} \right]_+ \\ &= \sum_{i=1}^M \left[ -\frac{V(\nu_i^{-1})r_i}{\nu_i} + \frac{V(\nu_i^{-1})r_i}{\nu_i} \frac{z}{z - \nu_i} \right]_+ \\ &= \sum_{i=1}^M \frac{V(\nu_i^{-1})r_i}{\nu_i} \frac{z}{z - \nu_i} \\ &= \sum_{i=1}^M R_i \frac{z}{z - \nu_i}, \end{aligned} \quad (3.88)$$

com

$$R_i = \frac{V(\nu_i^{-1})r_i}{\nu_i}. \quad (3.89)$$

Esta expressão para  $g(z)$  pode agora ser usada para calcular  $\|g(z)\|^2$ . Escrevendo explicitamente o produto interno envolvido temos:

$$\begin{aligned} \|g(z)\|^2 &= \langle g(z), g(z) \rangle = \frac{1}{2\pi} \oint_{|z|=1} g(z)g(z^{-1}) \frac{dz}{z} \\ &= \frac{1}{2\pi} \oint_{|z|=1} \sum_{i=1}^M R_i \frac{z}{z - \nu_i} \sum_{j=1}^M R_j \frac{z^{-1}}{z^{-1} - \nu_j} \frac{dz}{z} \\ &= \frac{1}{2\pi} \oint_{|z|=1} \sum_{i=1}^M \frac{R_i}{z - \nu_i} \sum_{j=1}^M R_j \frac{1}{1 - z\nu_j} dz \end{aligned}$$

$$= \frac{1}{2\pi} \oint_{|z|=1} \sum_{i=1}^M \frac{R_i}{z - \nu_i} \sum_{j=1}^M \frac{-R_j/\nu_j}{z - 1/\nu_j} dz. \quad (3.90)$$

Só os polos  $1/\nu_j$  estão dentro do círculo unitário, e, portanto, pelo teorema dos resíduos temos

$$\begin{aligned} \|g(z)\|^2 &= - \sum_{j=1}^M \sum_{i=1}^M \frac{R_i}{1/\nu_j - \nu_i} \frac{R_j}{\nu_j} \\ &= \sum_{j=1}^M \sum_{i=1}^M \frac{R_i R_j}{\nu_i \nu_j - 1}. \end{aligned} \quad (3.91)$$

Podemos ainda agrupar os termos  $R_i = R_j^*$  conjugados e escrever

$$\|g(z)\|^2 = \sum_{i=1}^M \frac{|R_i|^2}{|\nu_i|^2 - 1} + \sum_{\substack{j=1 \\ R_i \neq R_j^*}}^M \sum_{i=1}^M \frac{R_i R_j}{\nu_i \nu_j - 1}.$$

A suposição de polos próximos à circunferência unitária pode agora ser introduzida. Suponhamos que existam  $m$  polos com  $|\nu_i| \simeq 1$  e  $|R_i| \neq 0$ , de modo que a expressão acima seja dominada pelos  $m$  termos relativos a este polos:

$$\|g(z)\|^2 \simeq \sum_{i=1}^m \frac{|R_i|^2}{|\nu_i|^2 - 1}, \quad (3.92)$$

onde o termo  $|R_i|$ , conforme (3.89), é dado por

$$|R_i| = |V(\nu_i^{-1})| \left| \frac{r_i}{\nu_i} \right|. \quad (3.93)$$

Agora, conforme (3.83) os polos de  $\hat{H}(z)$  são zeros de  $V(z^{-1})$ . Portanto se os polos de  $\hat{H}(z)$  estiverem suficientemente longe dos polos  $\nu_i$  de  $H(z)$ , então  $|V(\nu_i^{-1})|$  não será próximo de zero, e, mais do que isso, uma vez que  $V(z)$  é passa-tudo e, por suposição  $|\nu_i^{-1}| \simeq 1$ , temos, nestas condições,  $|V(\nu_i^{-1})| \simeq 1$ , para quaisquer variações dos coeficientes de  $\hat{H}(z)$ . Ou seja, quando o sistema sendo identificado tem polos próximos à circunferência unitária e os polos do filtro adaptativo estão longe destes polos do sistema, o erro na superfície reduzida varia pouco com variações dos coeficientes de  $\hat{H}(z)$ .

Para exemplificar este comportamento a superfície de erro reduzida foi computada para

o caso das funções  $H_1(z)$ ,  $H_4(z)$ ,  $H_9(z)$  e  $H_{12}(z)$ , definidas na Tabela 2.2 do Item 2.6.8. Os gráficos obtidos estão nas Figuras 3-1 e 3-2, onde pode ser bem observado o aumento da extensão das regiões planas da superfície de erro reduzida quando algum polo de  $H(z)$  se aproxima da circunferência unitária.

### 3.7 Exemplos da simulação de algoritmos de adaptação

Neste item são apresentados resultados práticos relativos à convergência global dos algoritmos de adaptação RG, SMM e PLR considerados no Capítulo 2. O algoritmo SHARF ideal, isto é, com  $P(z) = D(z)$ , considerado no Capítulo 2, não foi testado, uma vez que não há sentido prático em supor  $P(z) = D(z)$  e iniciar a adaptação com  $A(z)$  afastado de  $D(z)$ .

Assim como foi feito no Item 2.7 os algoritmos foram testados na identificação das funções  $H_1(z)$ ,  $H_4(z)$ ,  $H_6(z)$ ,  $H_7(z)$ ,  $H_9(z)$ ,  $H_{14}(z)$ ,  $H_{15}(z)$  definidas no Item 2.6 (a função  $H_{12}(z)$  não foi testada pois, já no caso local, os tempos de convergência eram excessivamente longos). Todos os procedimentos relativos à execução dos algoritmos mencionados no Item 2.7 permanecem válidos, com a exceção de que foi tomado como ponto inicial da adaptação a origem, isto é,  $\mathbf{w}(0) = \mathbf{0}$ .

Como está sendo considerada agora a convergência global dos algoritmos de adaptação, são apresentados também os resultados do algoritmo SMM, que, ao contrário do que ocorre localmente em torno do mínimo global, não são necessariamente similares aos do algoritmo RG.

Os valores de  $\mu$  utilizados e o tempo de convergência  $n_{60}$  obtidos (ver Item 2.6 para a metodologia envolvendo estas grandezas), estão apresentados na Tabela 3.1. São apresentados também gráficos com a adaptação dos coeficientes e a evolução do erro quadrático  $(y(n) - \hat{y}(n))^2$ . O número da figura para cada caso está na Tabela 3.2.

Pode-se verificar pelos resultados, que em regiões (afastadas do mínimo global) onde, devido ao fato do sistema  $H(z)$  ter polos próximos à circunferência unitária, a superfície de erro reduzida é excessivamente plana, a convergência dos algoritmos de passo constante testados é lenta, o que vale mesmo para os algoritmos SMM e PLR que não são, a rigor, baseados no gradiente da superfície de erro.

Função	Algoritmo					
	RG		PLR		SMM	
	$\mu$	$n_{60}$	$\mu$	$n_{60}$	$\mu$	$n_{60}$
$H_1(z)$	$1,5 \times 10^{-3}$	$2,5 \times 10^4$	$1,5 \times 10^{-2}$	$7,2 \times 10^3$	$2 \times 10^{-3}$	$1,4 \times 10^4$
$H_4(z)$	$1,5 \times 10^{-5}$	(*)	$3 \times 10^{-4}$	(*)	$1,5 \times 10^{-5}$	(*)
$H_6(z)$	$8 \times 10^{-3}$	$1,9 \times 10^4$	$3 \times 10^{-2}$	$1,2 \times 10^4$	$1 \times 10^{-2}$	$4,1 \times 10^3$
$H_7(z)$	$1,5 \times 10^{-4}$	(*)	$2 \times 10^{-3}$	(*)	$1,5 \times 10^{-4}$	(*)
$H_9(z)$	$1,5 \times 10^{-3}$	(*)	$5,0 \times 10^{-2}$	$4,2 \times 10^4$	$1,5 \times 10^{-3}$	$2,4 \times 10^5$
$H_{14}(z)$	$7,0 \times 10^{-3}$	$1,8 \times 10^4$	$3 \times 10^{-2}$	$8,2 \times 10^3$	$8 \times 10^{-3}$	$4,4 \times 10^3$
$H_{15}(z)$	$1,0 \times 10^{-4}$	(*)	$8 \times 10^{-3}$	(*)	$1,0 \times 10^{-4}$	(*)

Tabela 3.1: Resultados da execução dos algoritmos RG, PLR e SMM com início na origem;  $\mu$ : passo de adaptação máximo;  $n_{60}$ : número de iterações para erro quadrático de saída ficar 60 dB abaixo da potência de saída; (\*): execução não completada devido ao longo tempo de convergência ( $> 10^6$ );

Função	Número da figura		
	RG	PLR	SMM
$H_1(z)$	3-3	3-4	3-5
$H_4(z)$	3-6	3-7	3-8
$H_6(z)$	3-9	3-10	3-11
$H_7(z)$	3-12	3-13	3-14
$H_9(z)$	3-15	3-16	3-17
$H_{14}(z)$	3-18	3-19	3-20
$H_{15}(z)$	3-21	3-22	3-23

Tabela 3.2: Número da figura com gráficos da execução dos algoritmos para cada caso de  $H(z)$

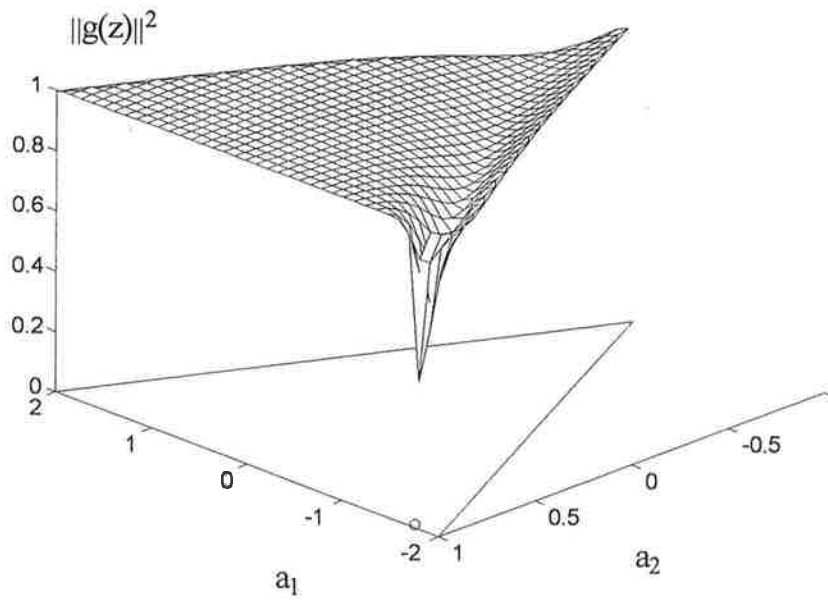
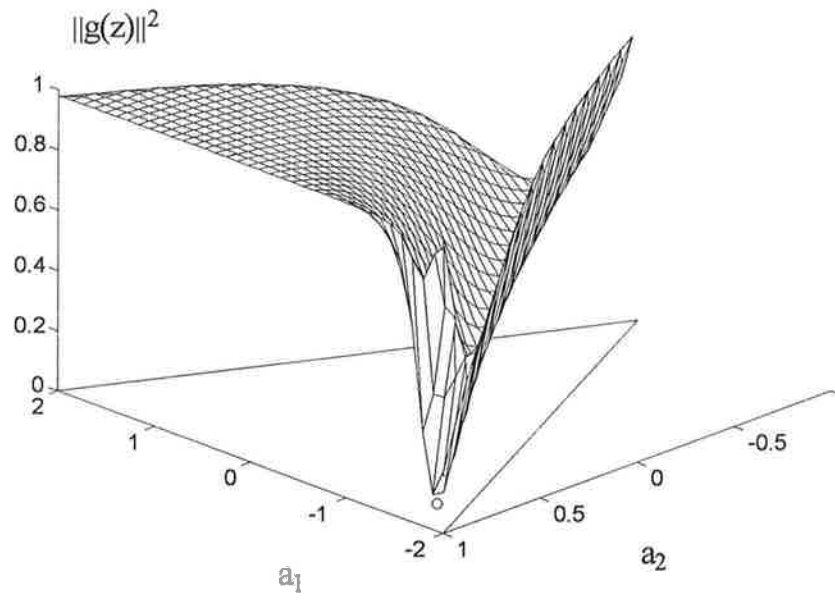


Figura 3-1: Superfície de erro reduzida para  $H_1(z)$  (em cima) e  $H_4(z)$  (em baixo)

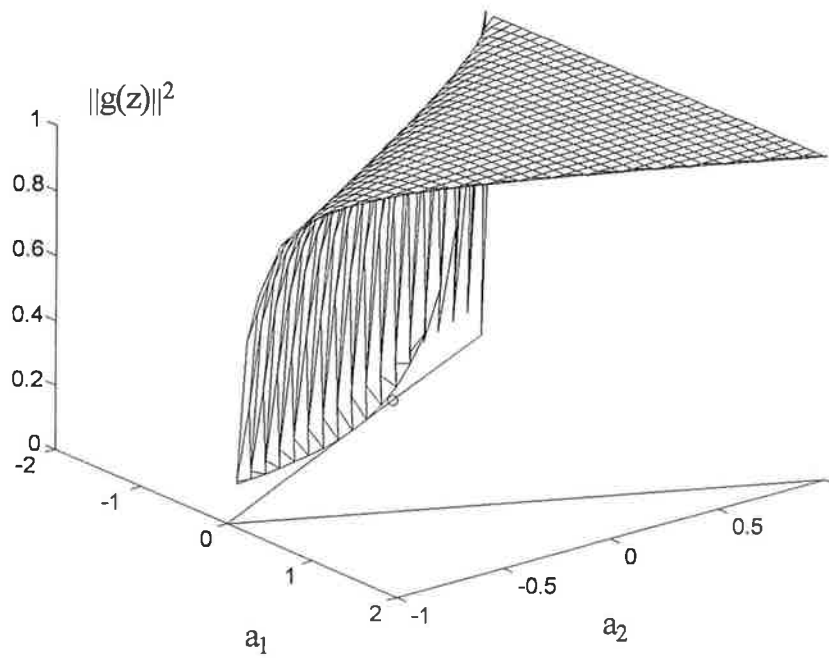
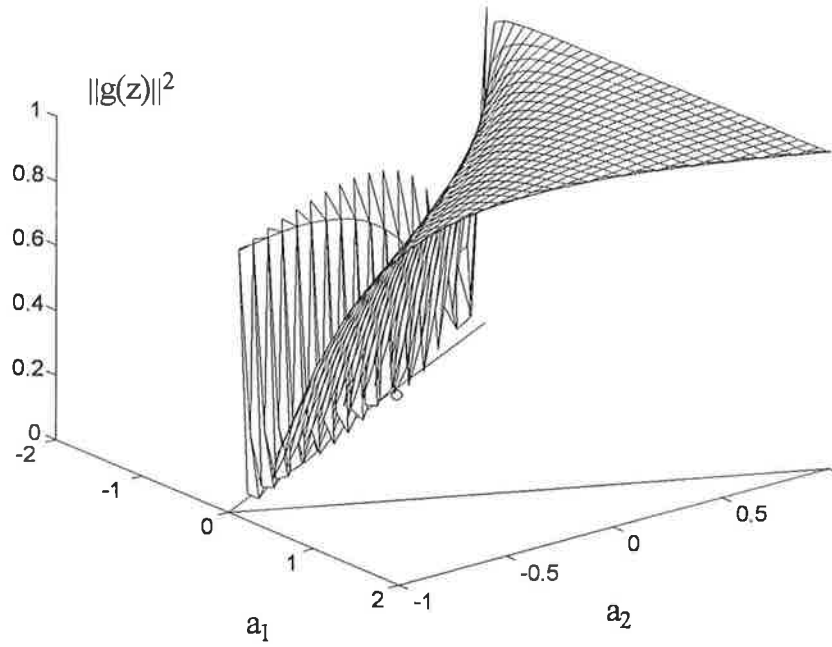


Figura 3-2: Superfície de erro reduzida para  $H_9(z)$  (em cima) e  $H_{12}(z)$  (em baixo)

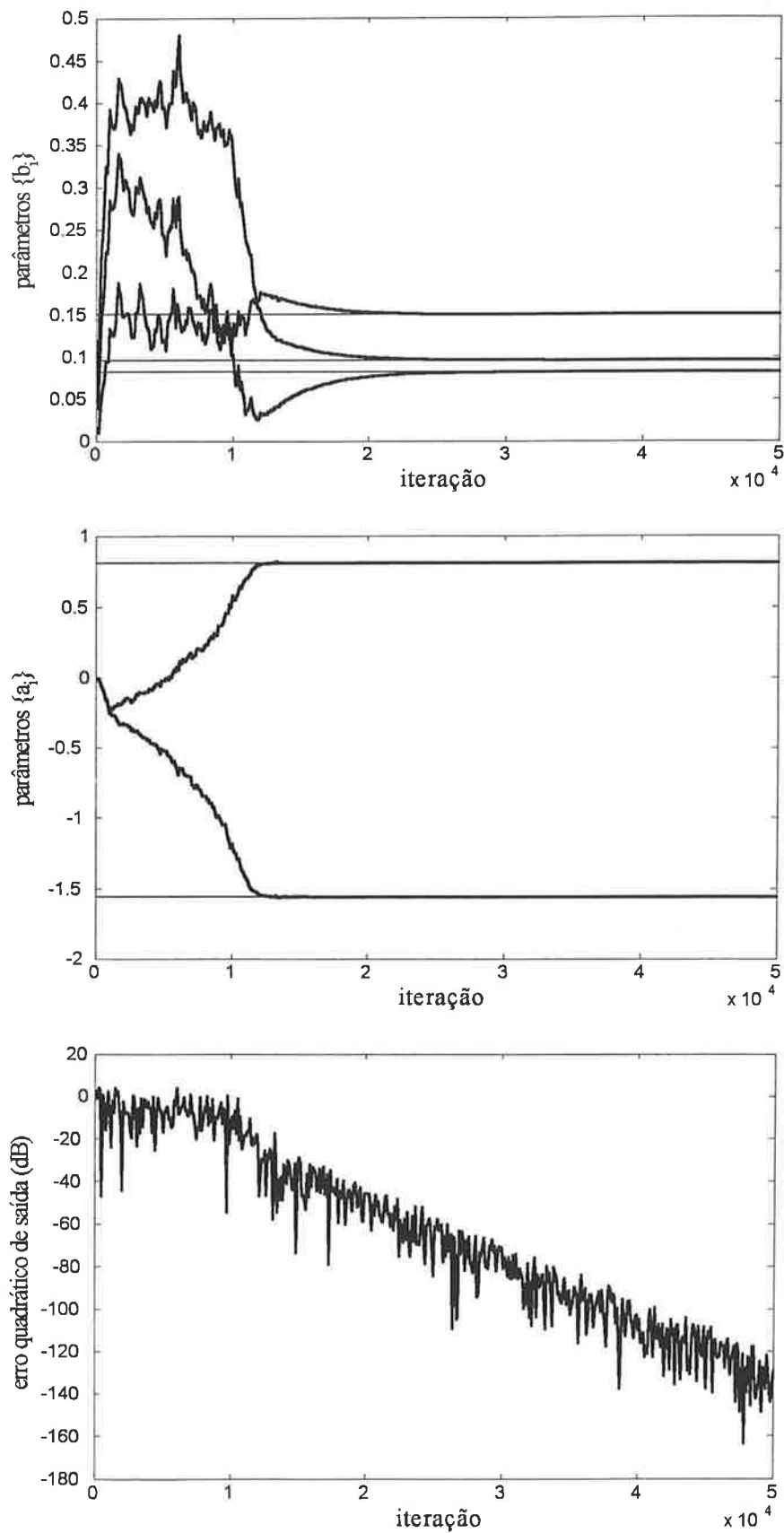


Figura 3-3: Algoritmo RG, sistema  $H_1(z)$  (parâmetros em traços claros)

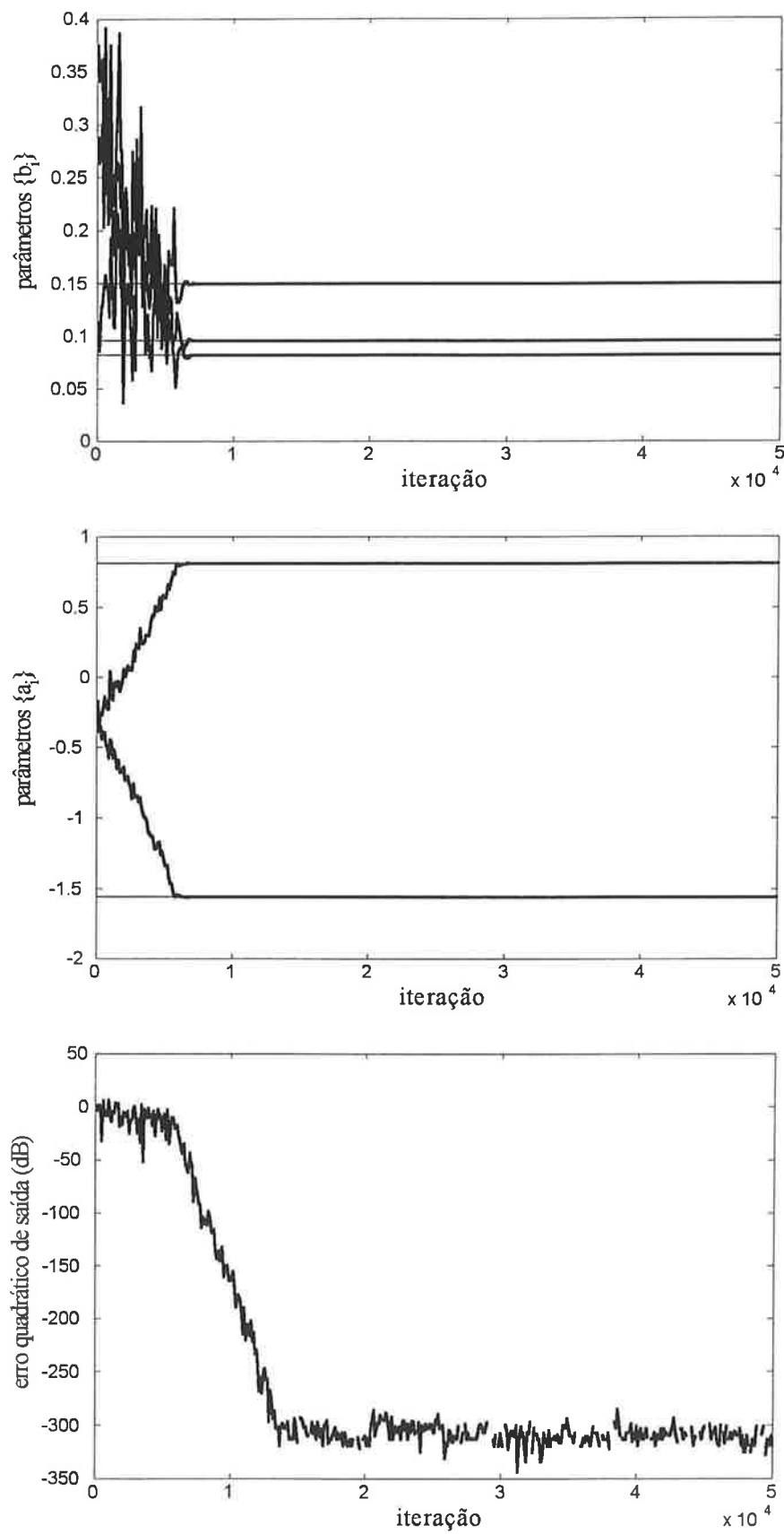


Figura 3-4: Algoritmo PLR, sistema  $H_1(z)$  (parâmetros em traços claros)



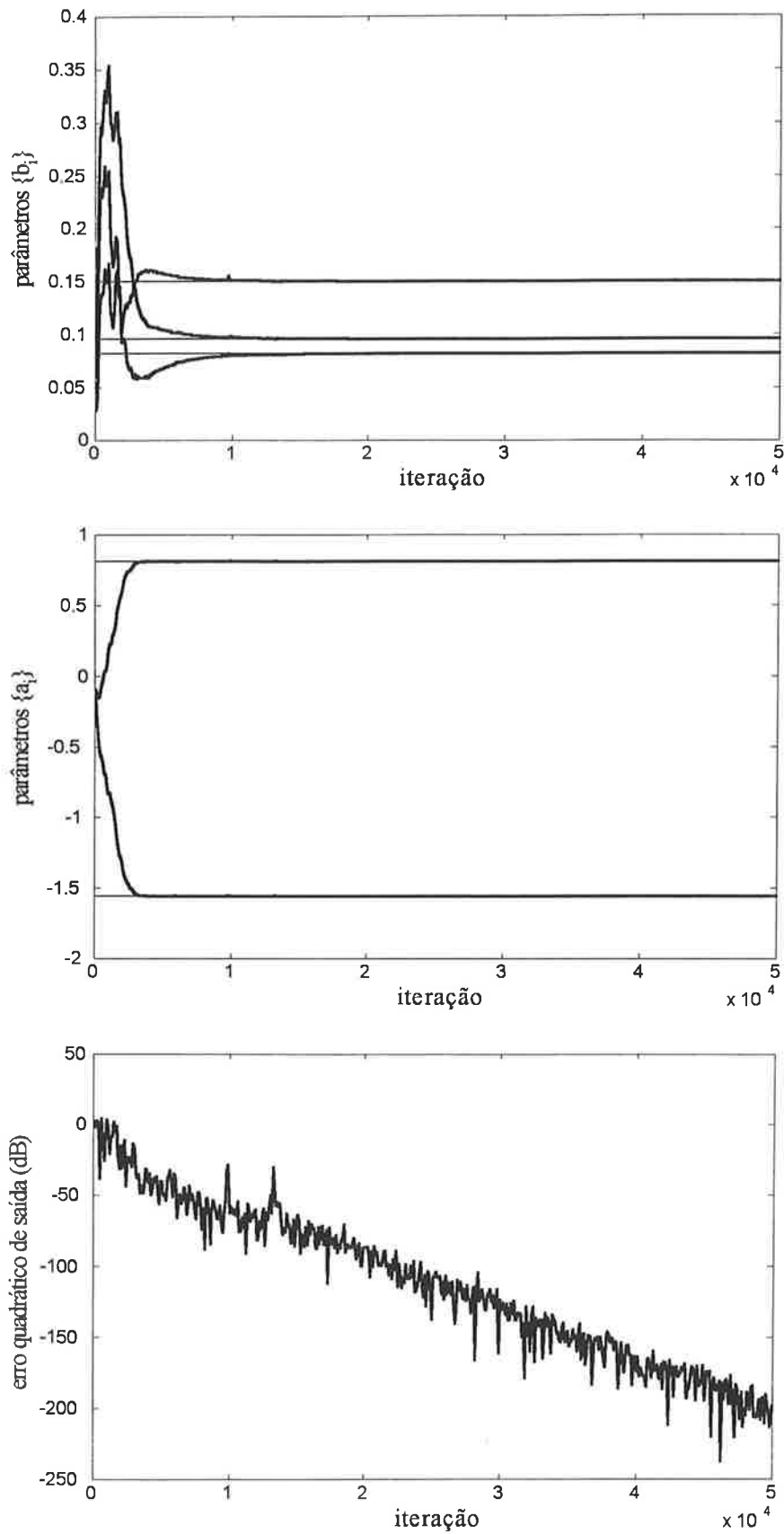


Figura 3-5: Algoritmo SMM, sistema  $H_1(z)$  (parâmetros em traços claros)

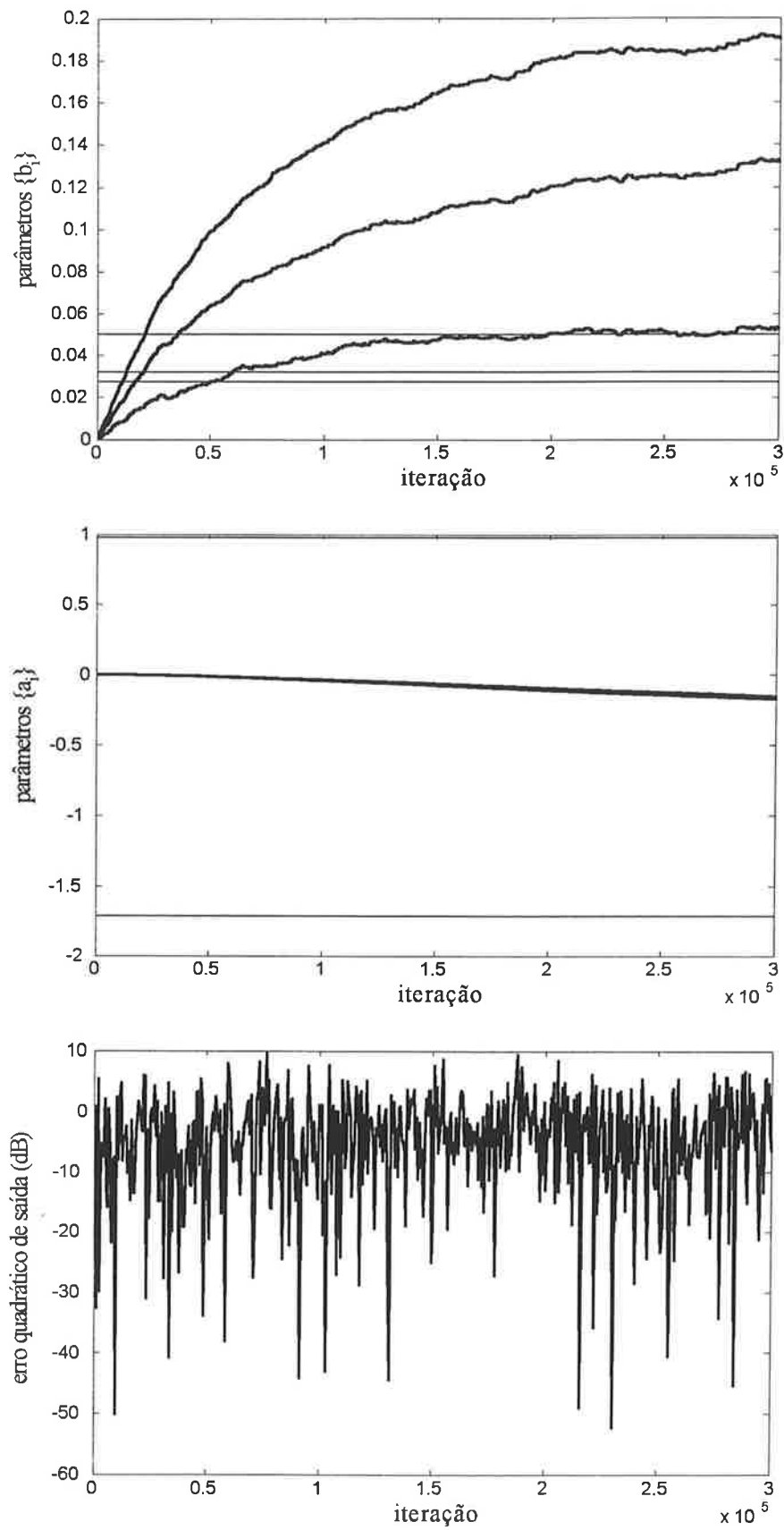


Figura 3-6: Algoritmo RG, sistema  $H_4(z)$  (parâmetros em traços claros)

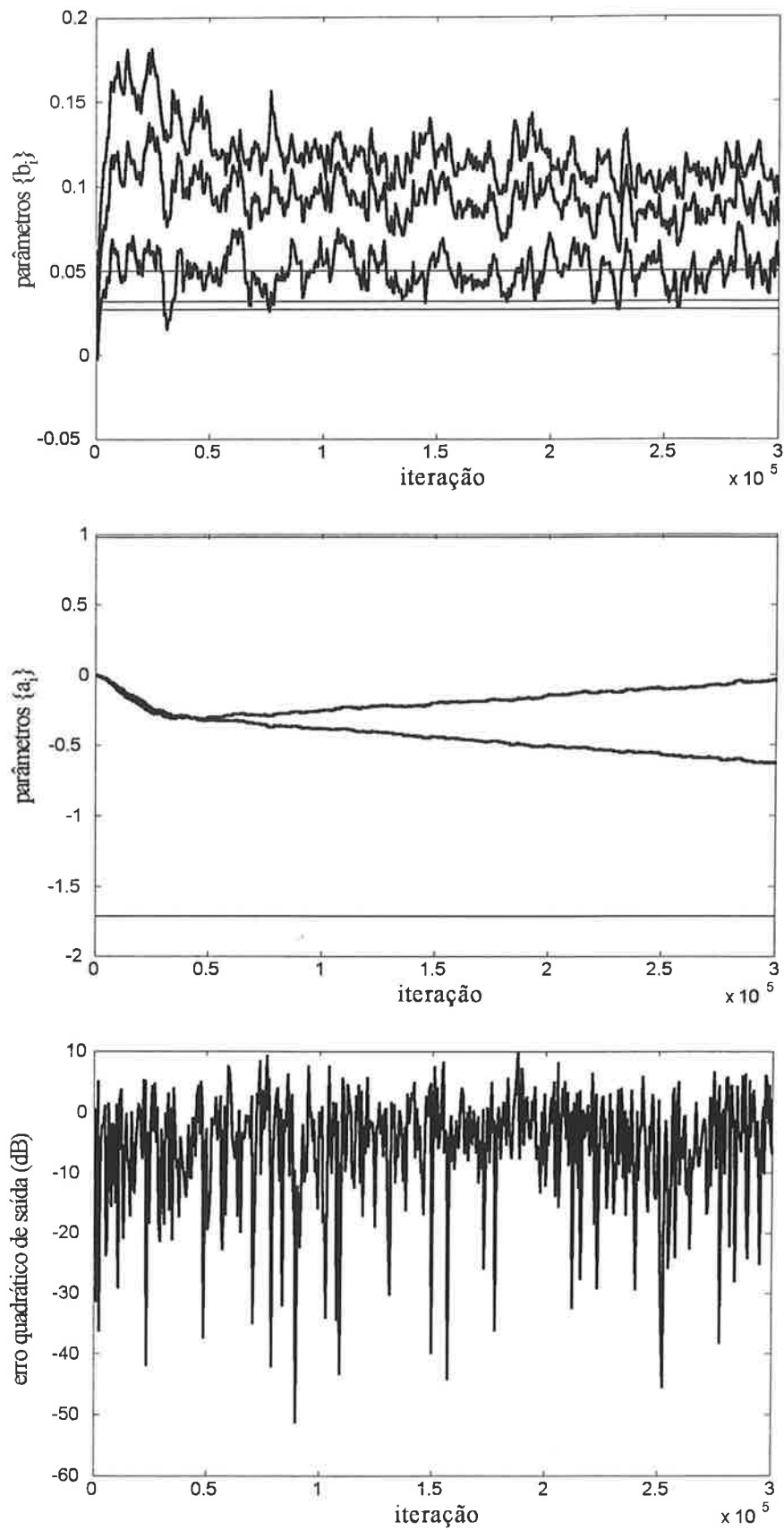


Figura 3-7: Algoritmo PLR, sistema  $H_4(z)$  (parâmetros em traços claros)

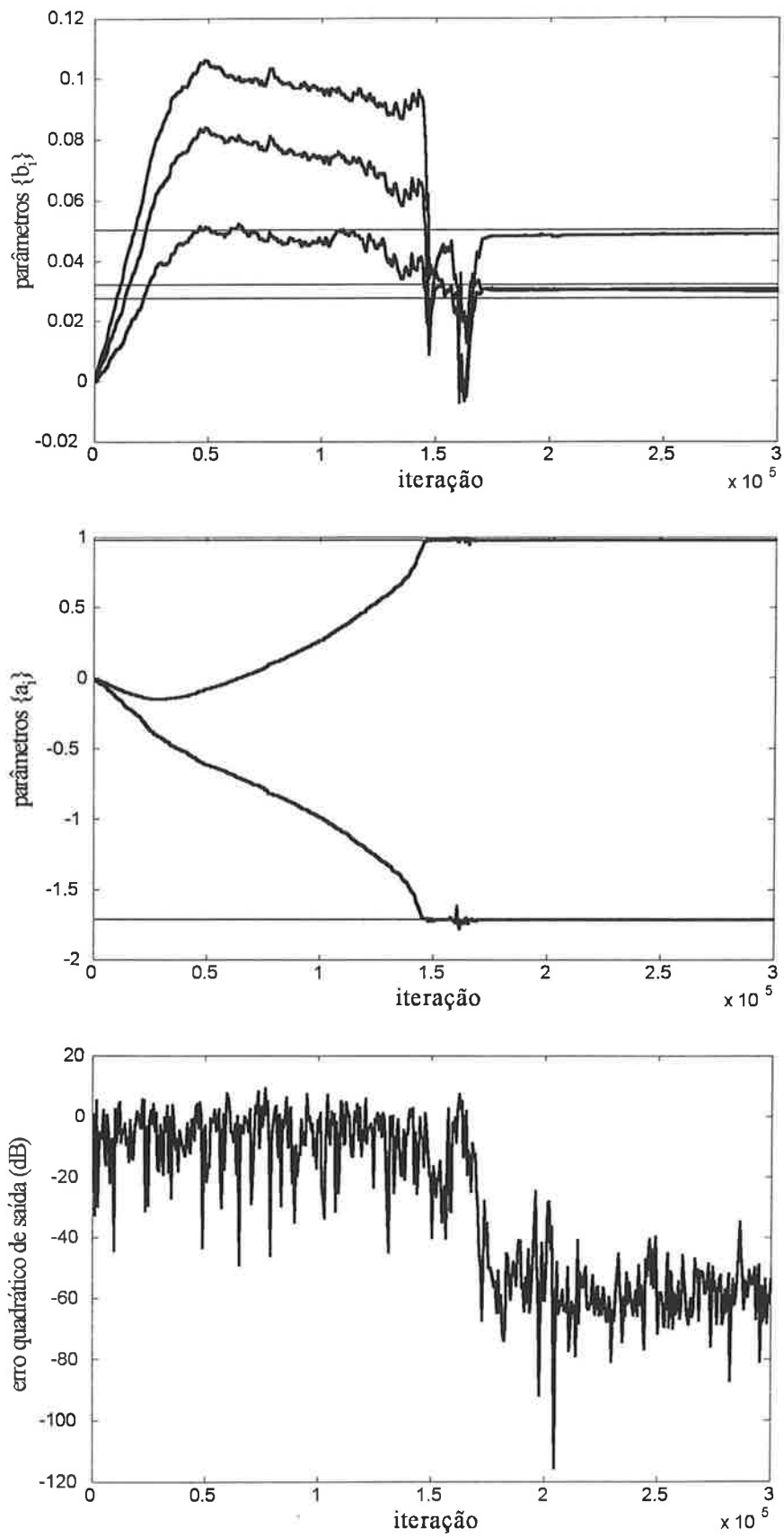


Figura 3-8: Algoritmo SMM, sistema  $H_4(z)$  (parâmetros em traços claros)

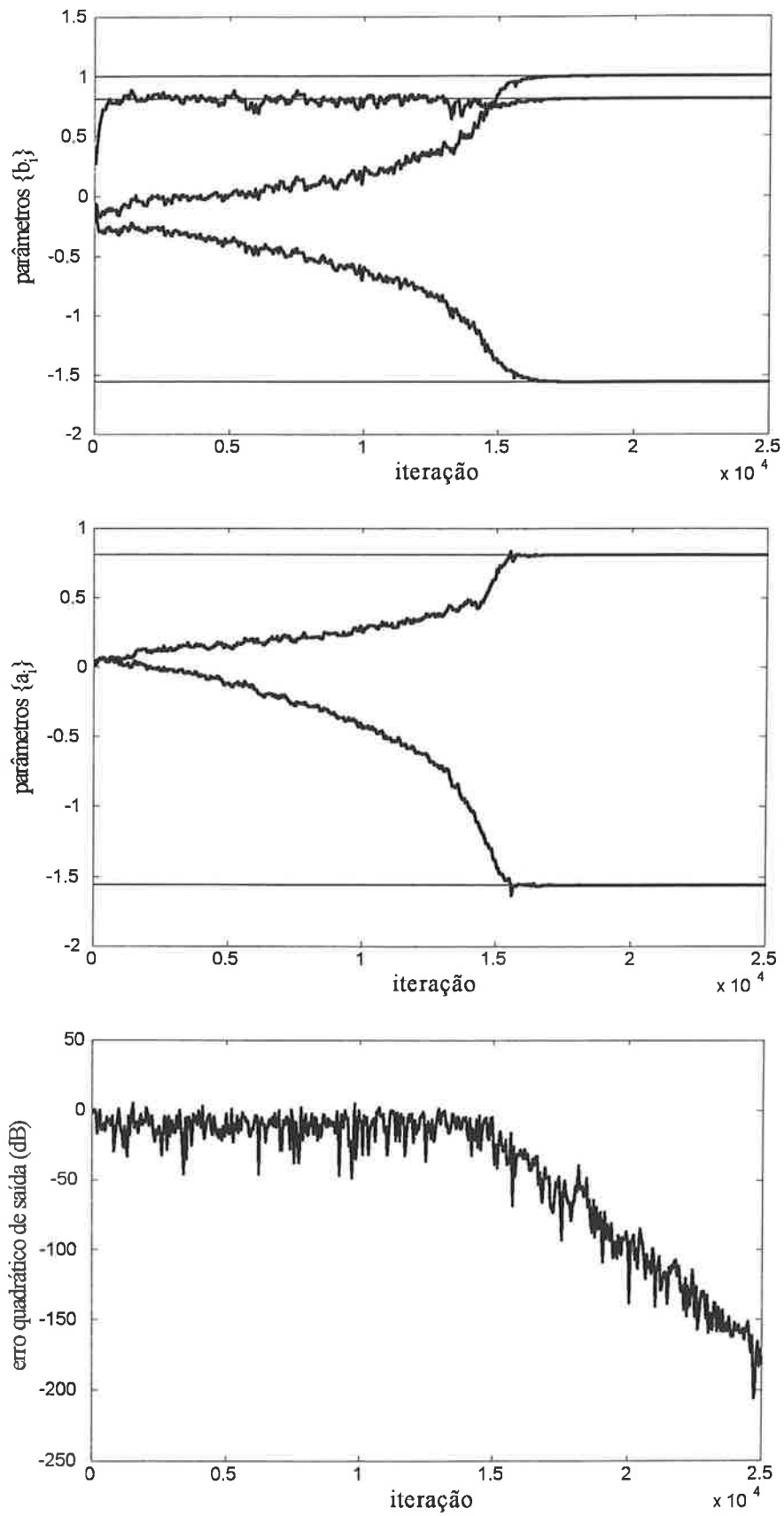


Figura 3-9: Algoritmo RG, sistema  $H_6(z)$  (parâmetros em traços claros)

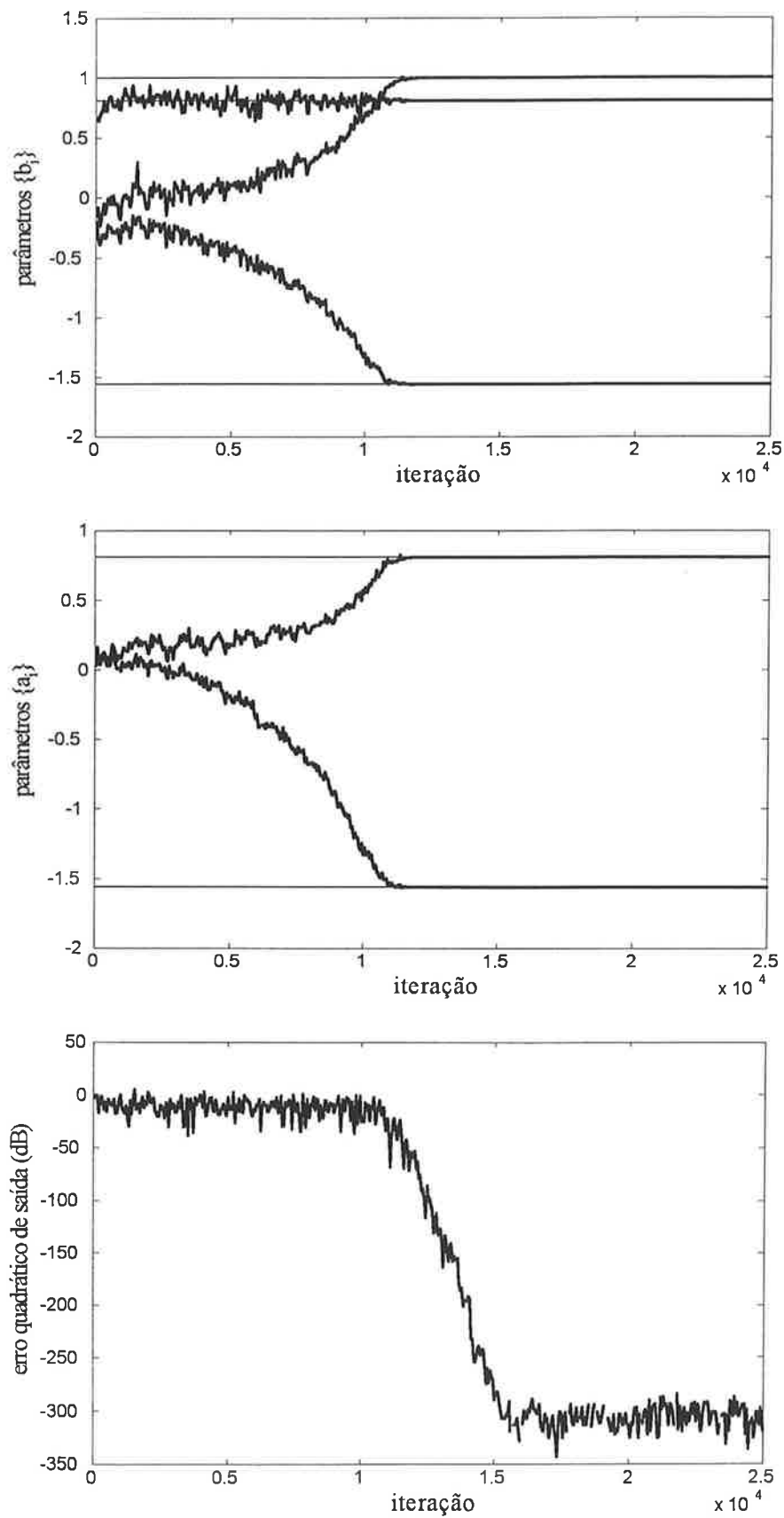


Figura 3-10: Algoritmo PLR, sistema  $H_6(z)$  (parâmetros em traços claros)

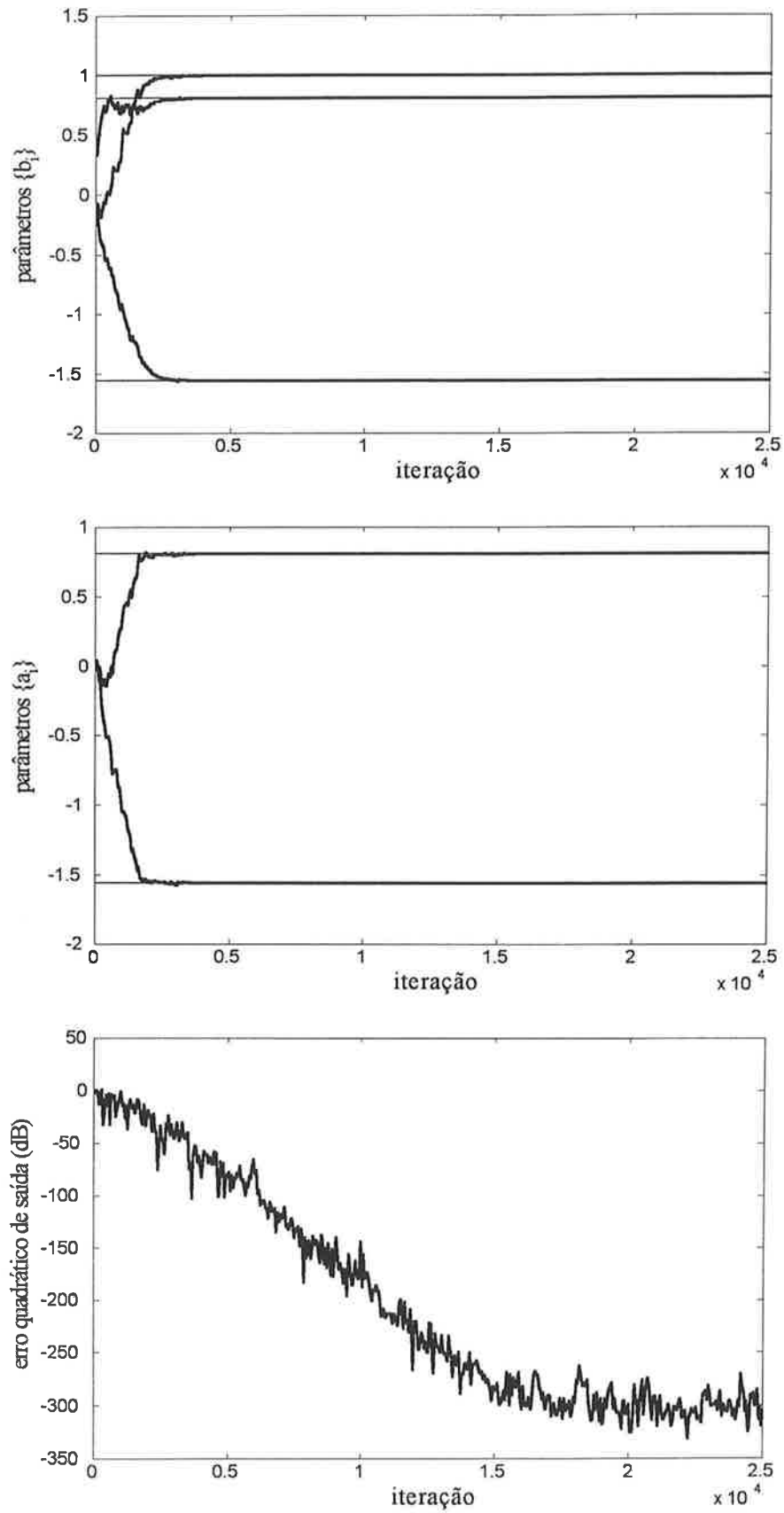


Figura 3-11: Algoritmo SMM, sistema  $H_6(z)$  (parâmetros em traços claros)

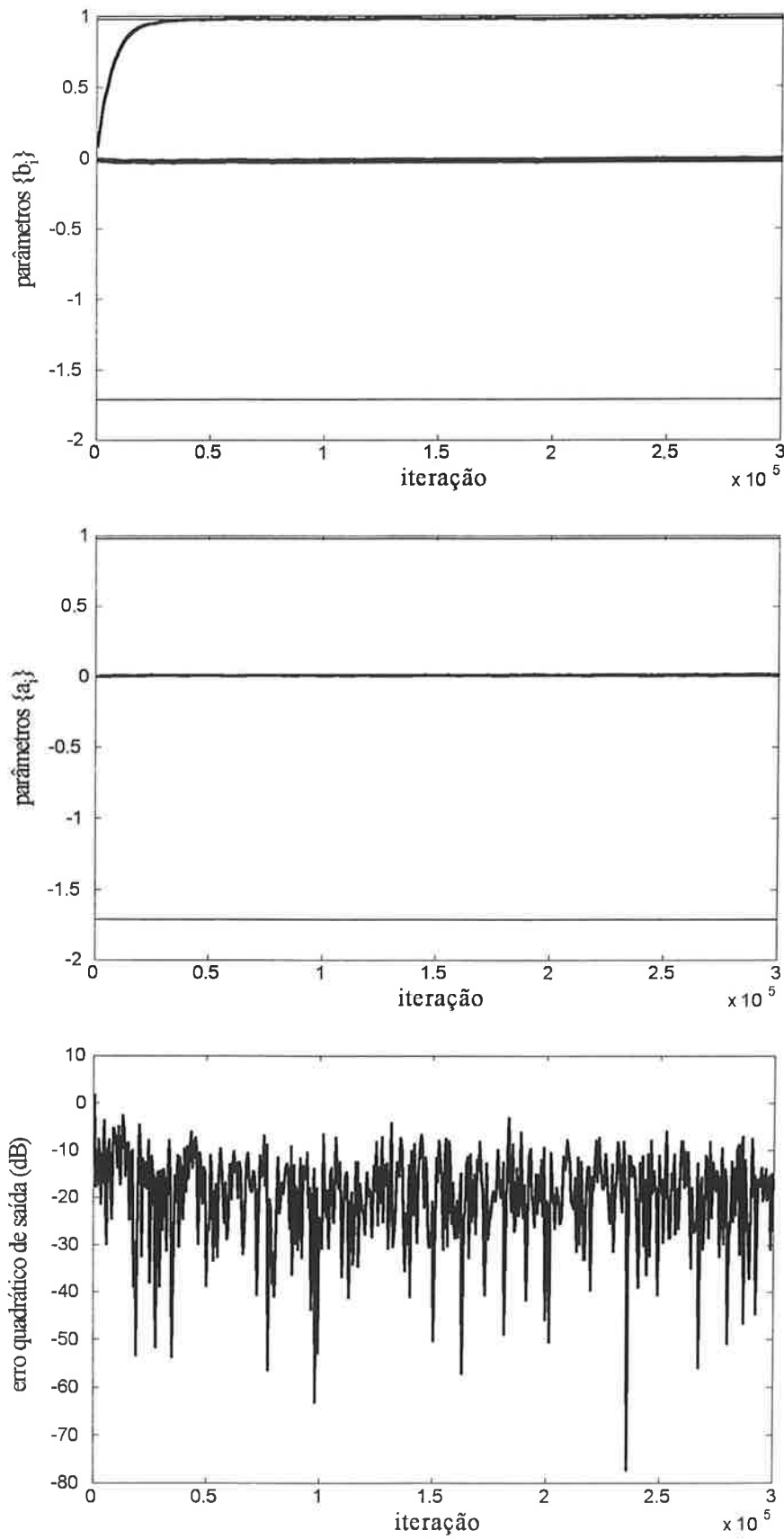


Figura 3-12: Algoritmo RG, sistema  $H_7(z)$  (parâmetros em traços claros)



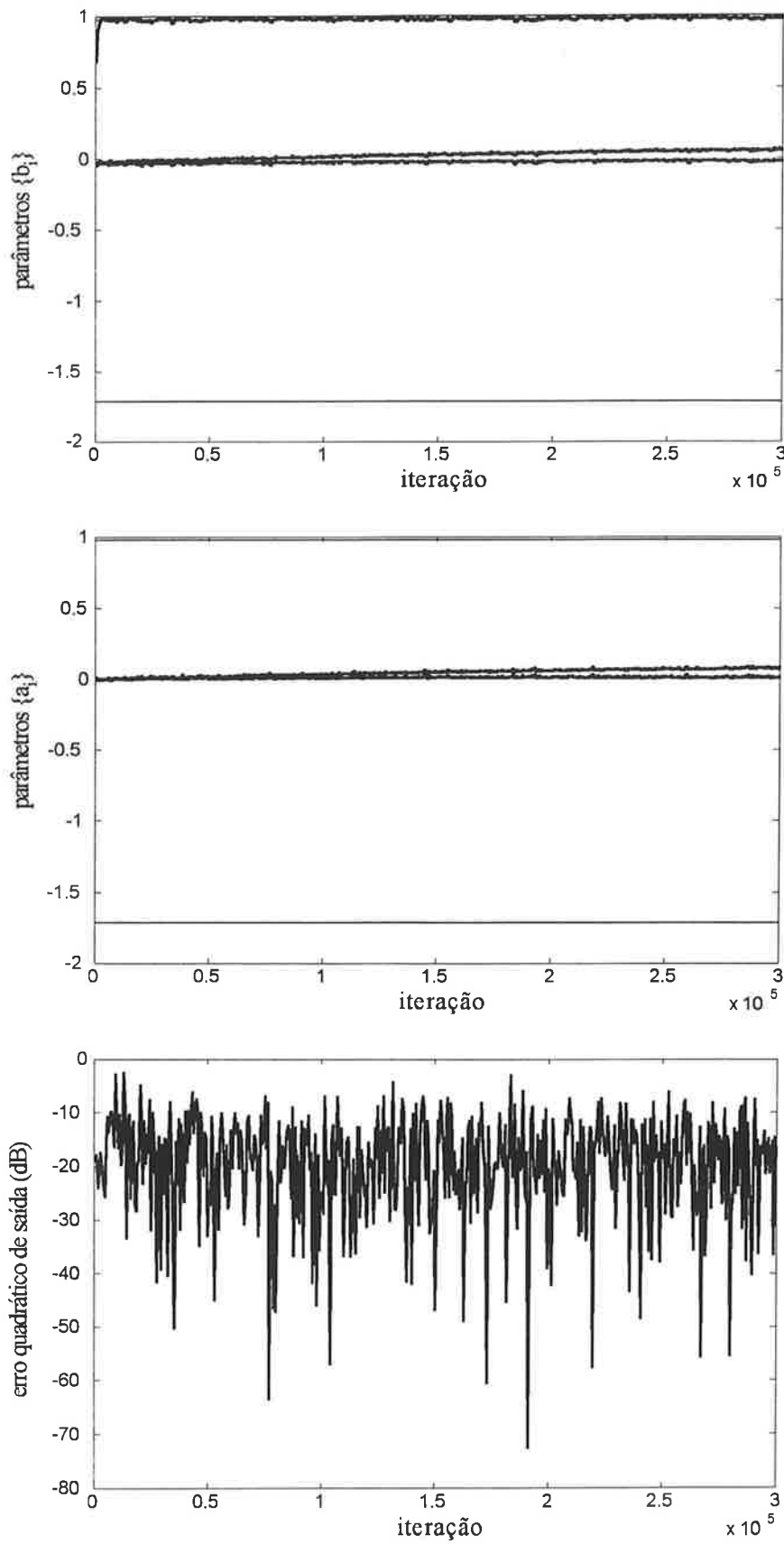


Figura 3-13: Algoritmo PLR, sistema  $H_7(z)$  (parâmetros em traços claros)

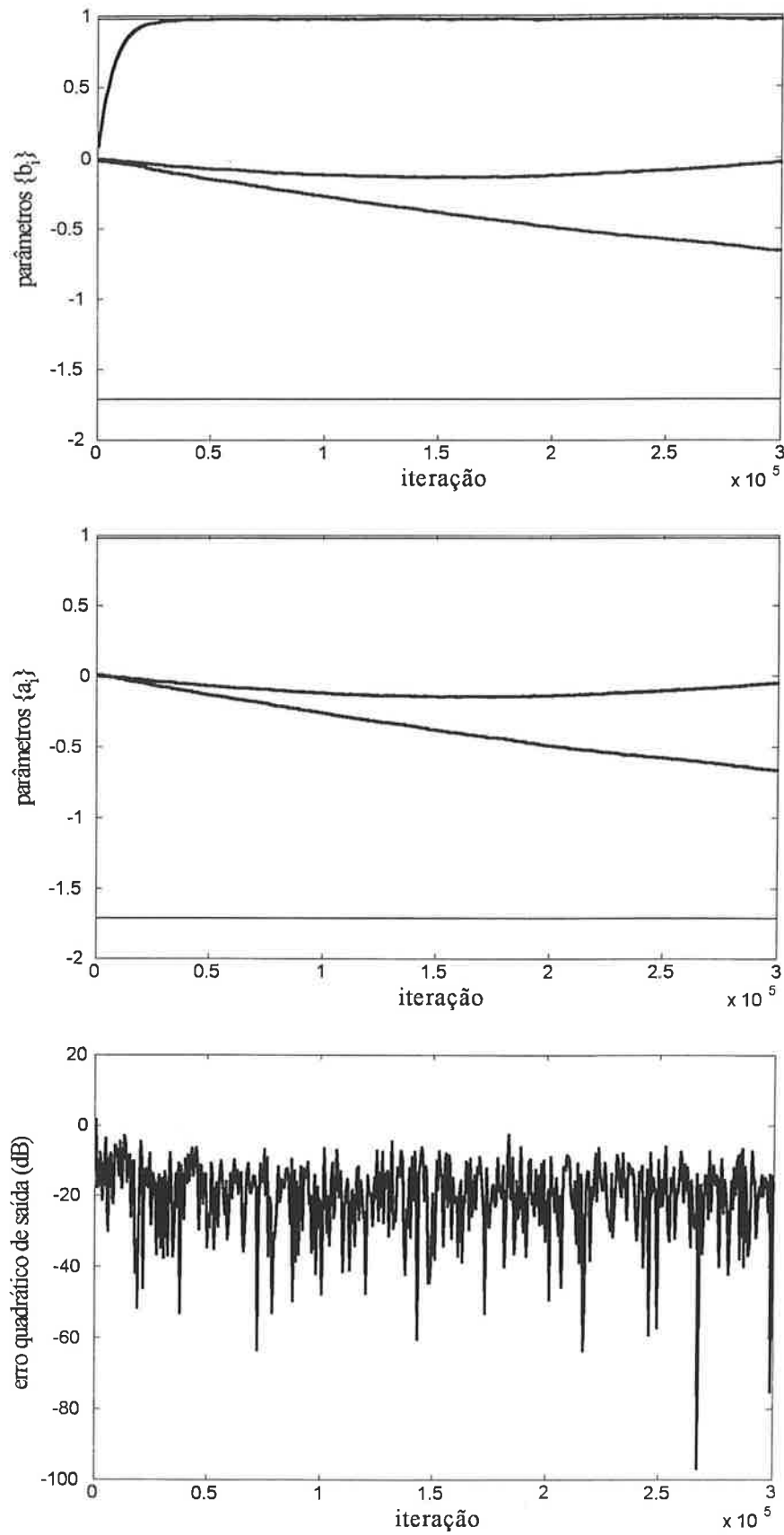


Figura 3-14: Algoritmo SMM, sistema  $H_7(z)$  (parâmetros em traços claros)

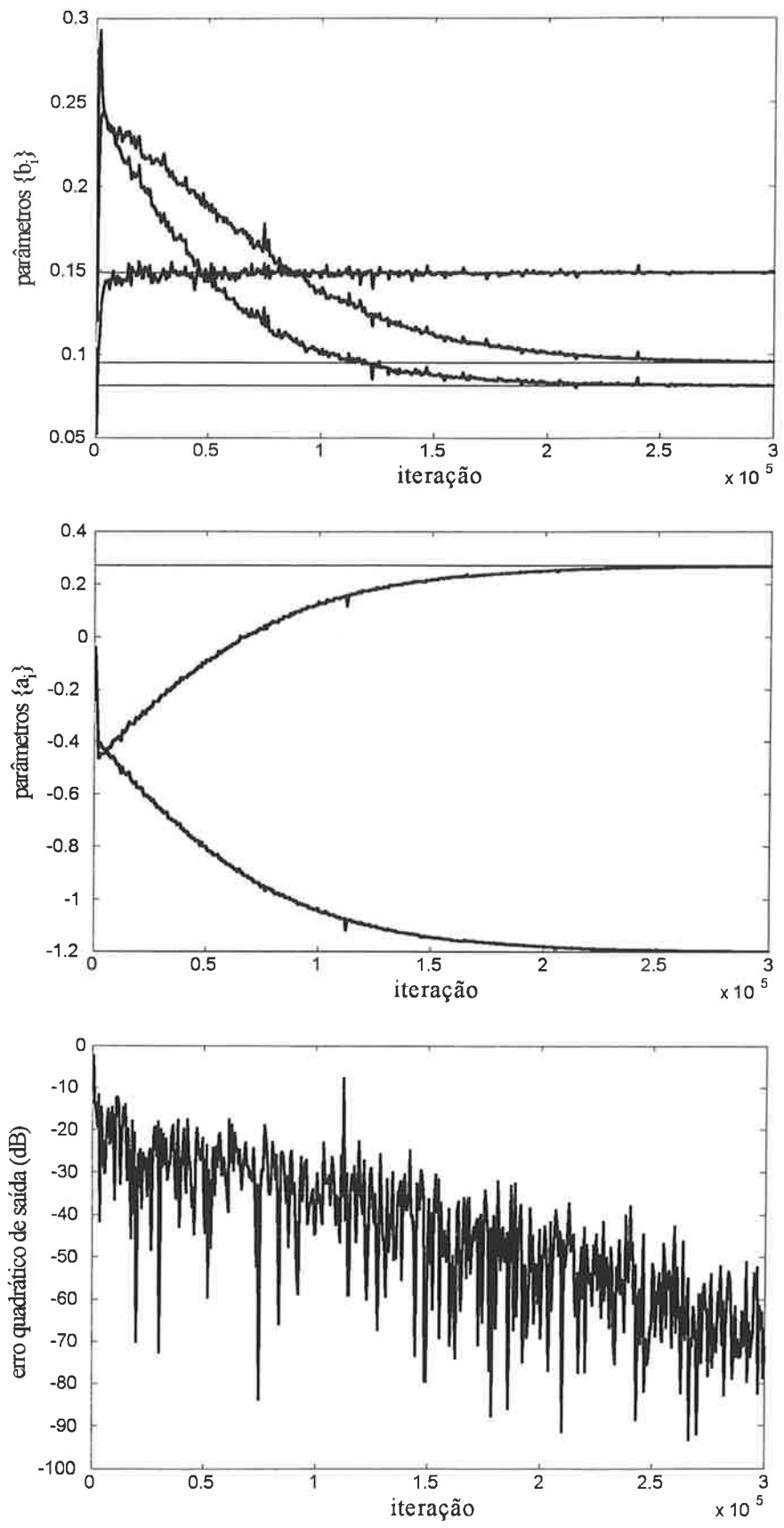


Figura 3-15: Algoritmo RG; sistema  $H_9(z)$  (parâmetros em traços claros)

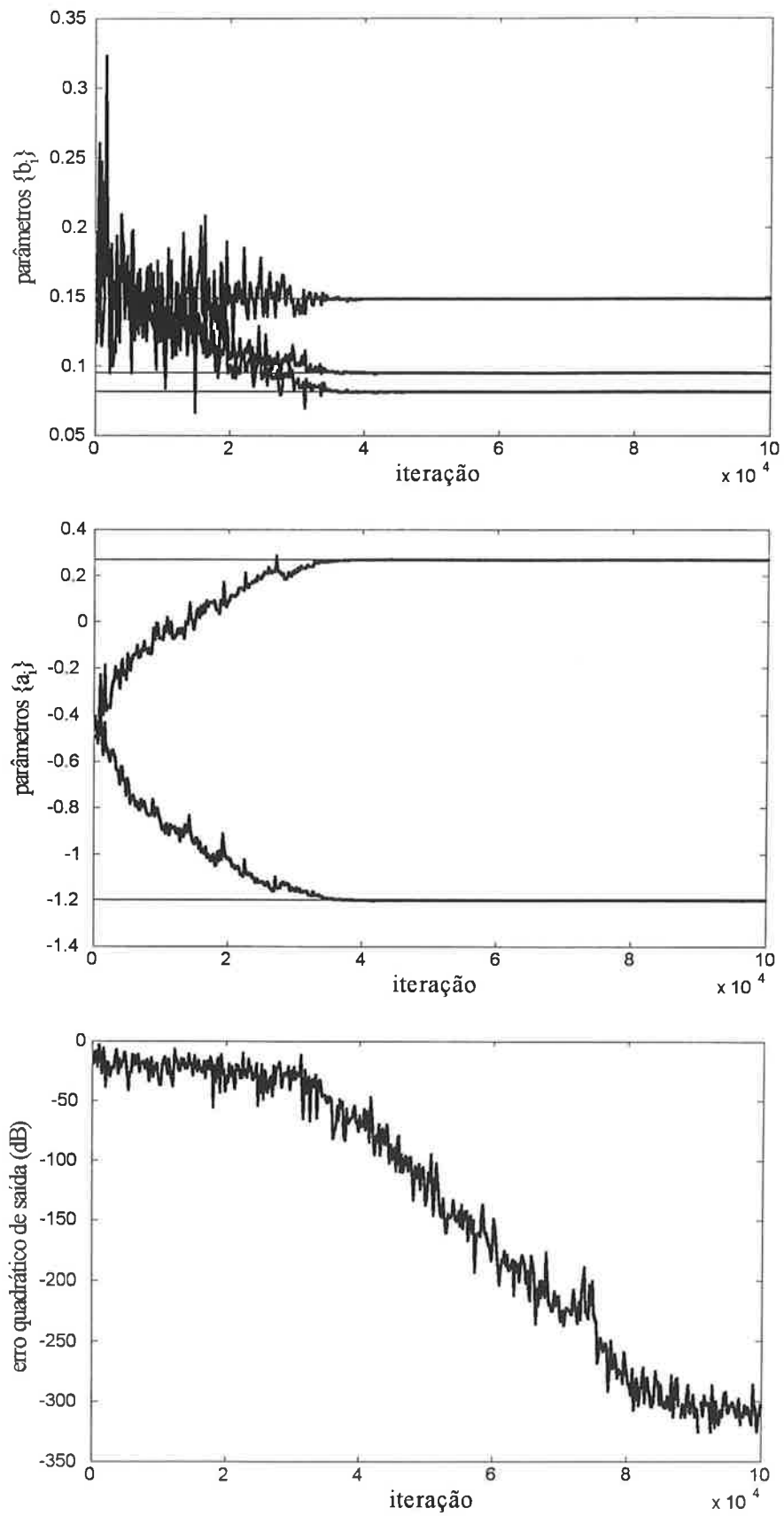


Figura 3-16: Algoritmo PLR; sistema  $H_9(z)$  (parâmetros em traços claros)

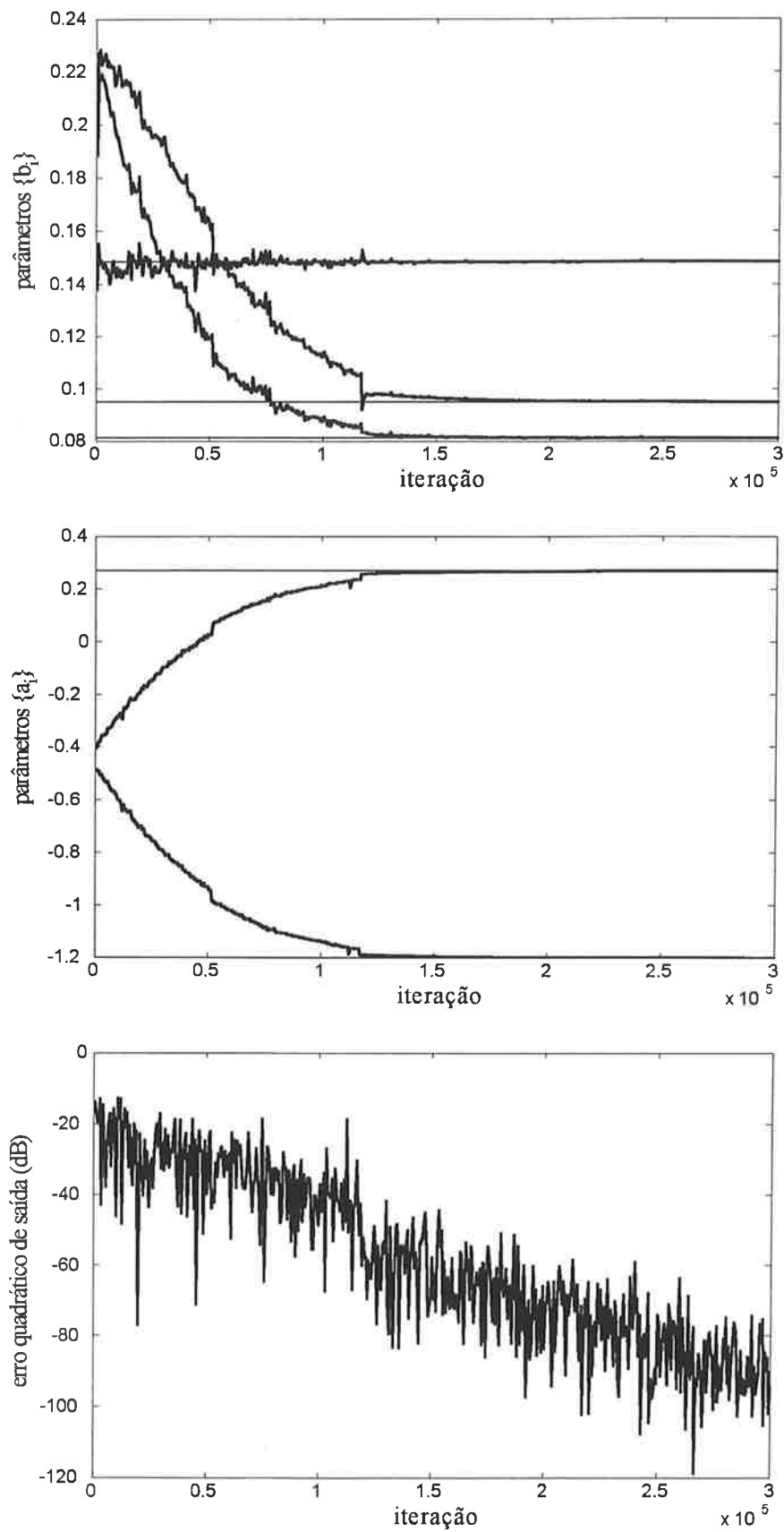


Figura 3-17: Algoritmo SMM; sistema  $H_9(z)$  (parâmetros em traços claros)

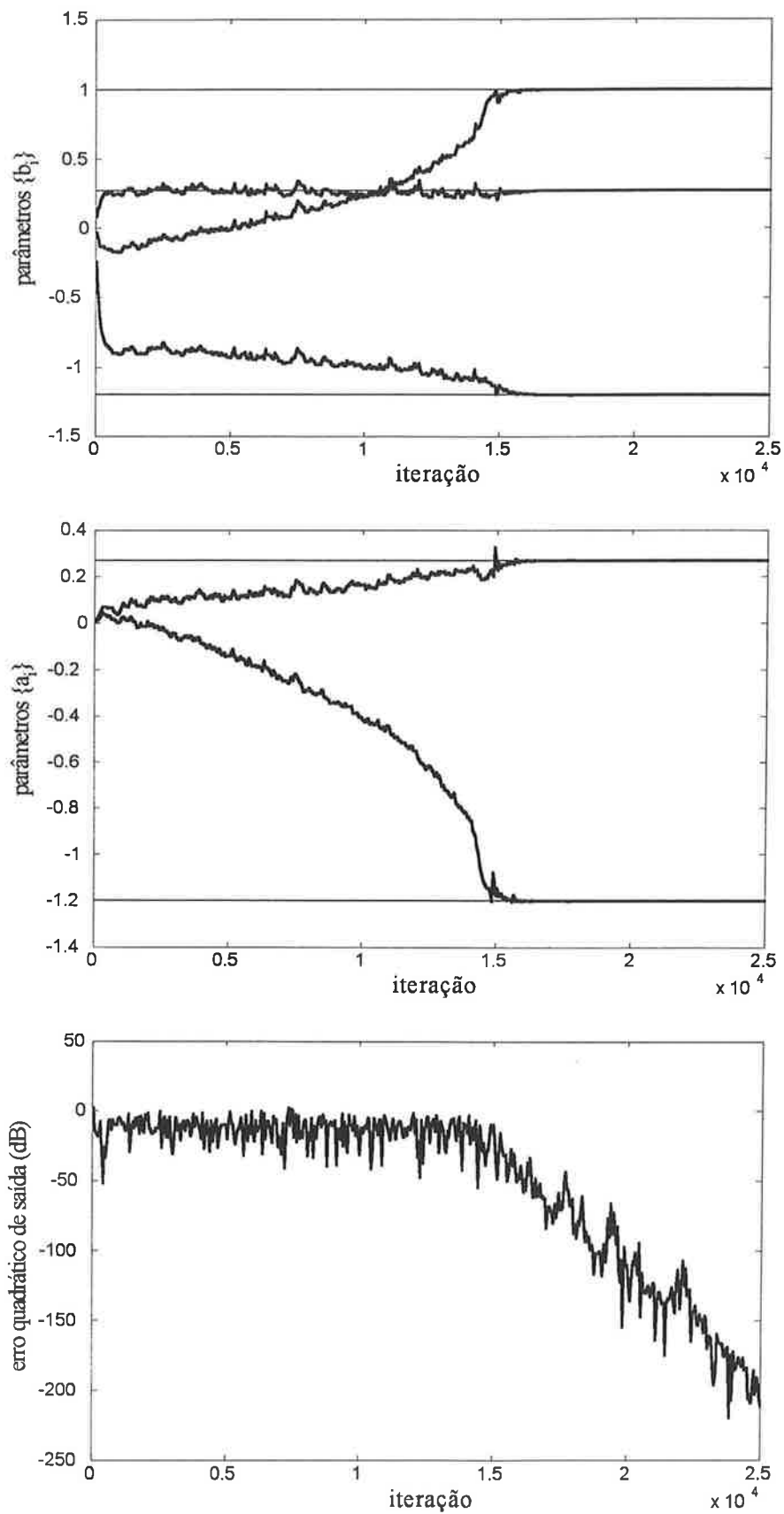


Figura 3-18: Algoritmo RG; sistema  $H_{14}(z)$  (parâmetros em traços claros)

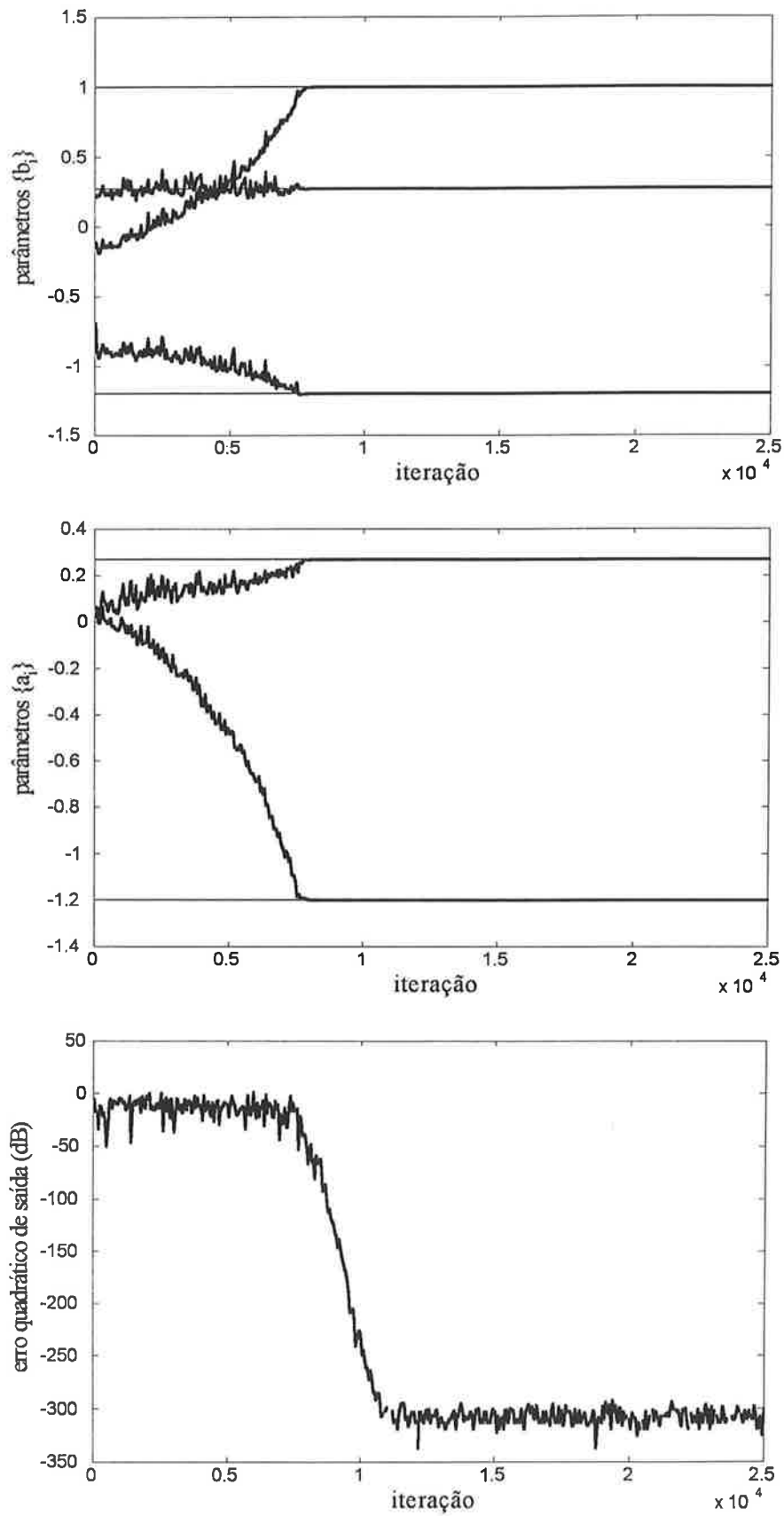


Figura 3-19: Algoritmo PLR; sistema  $H_{14}(z)$  (parâmetros em traços claros)

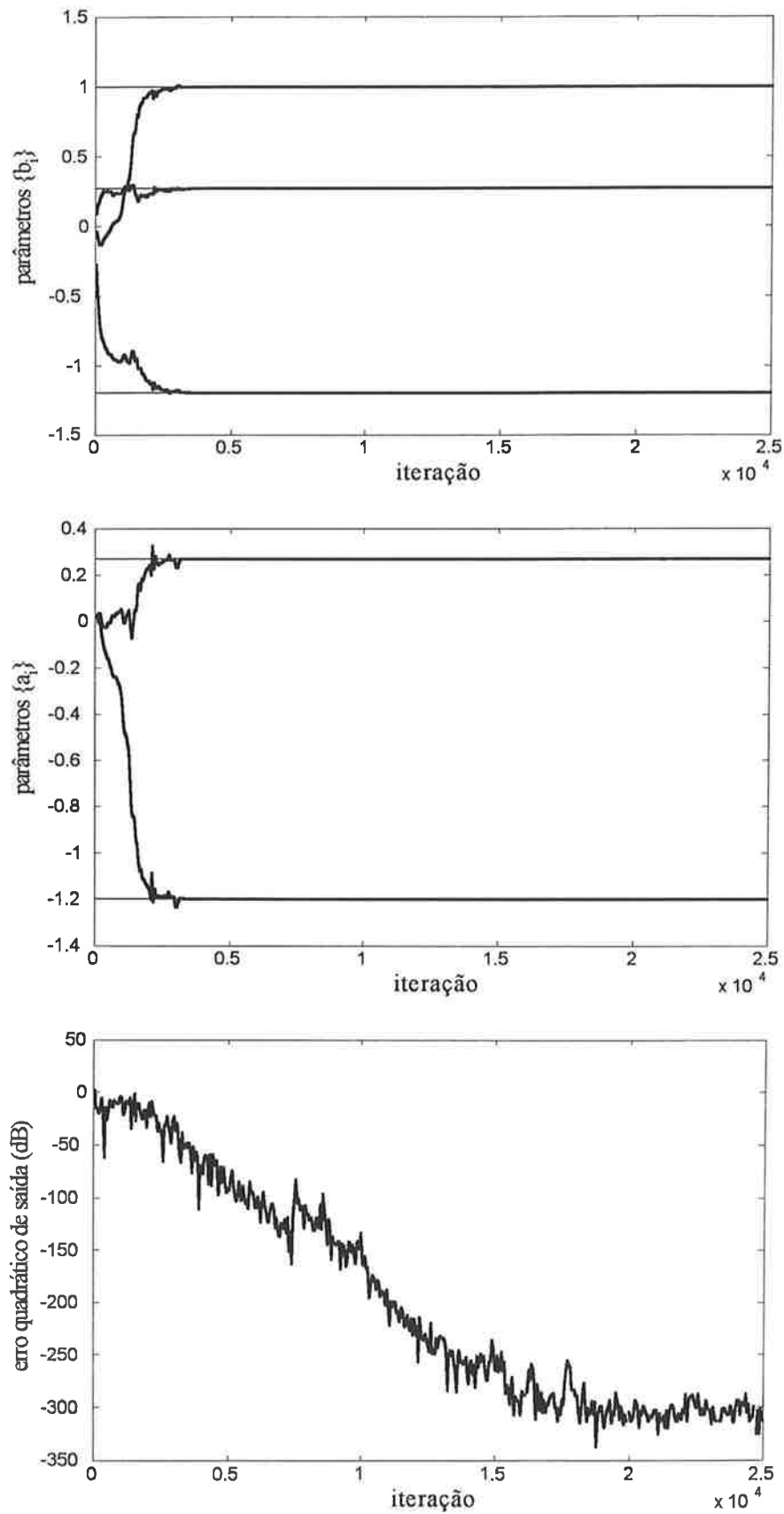


Figura 3-20: Algoritmo SMM; sistema  $H_{14}(z)$  (parâmetros em traços claros)



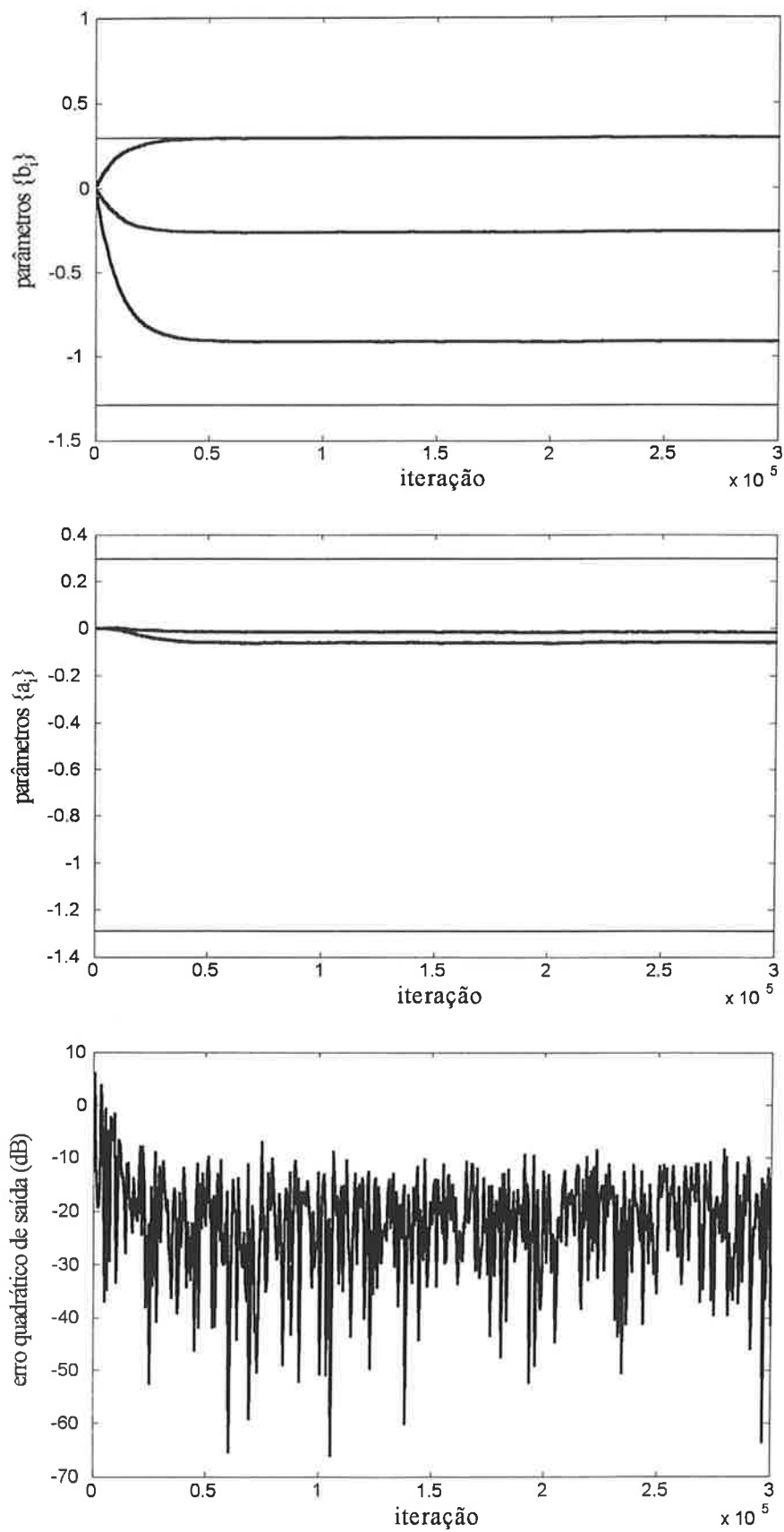


Figura 3-21: Algoritmo RG; sistema  $H_{15}(z)$  (parâmetros em traços claros)

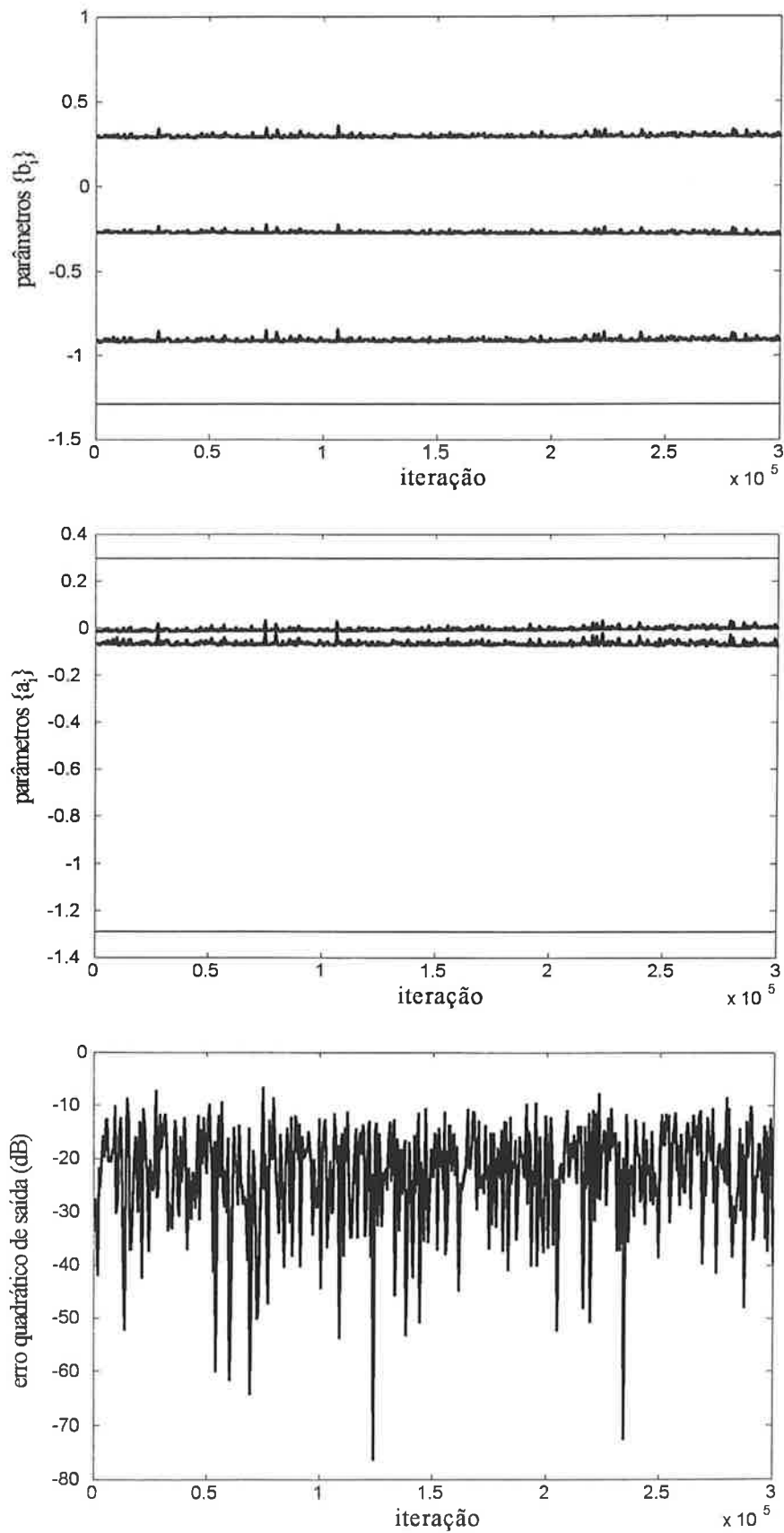


Figura 3-22: Algoritmo PLR; sistema  $H_{15}(z)$  (parâmetros em traços claros)

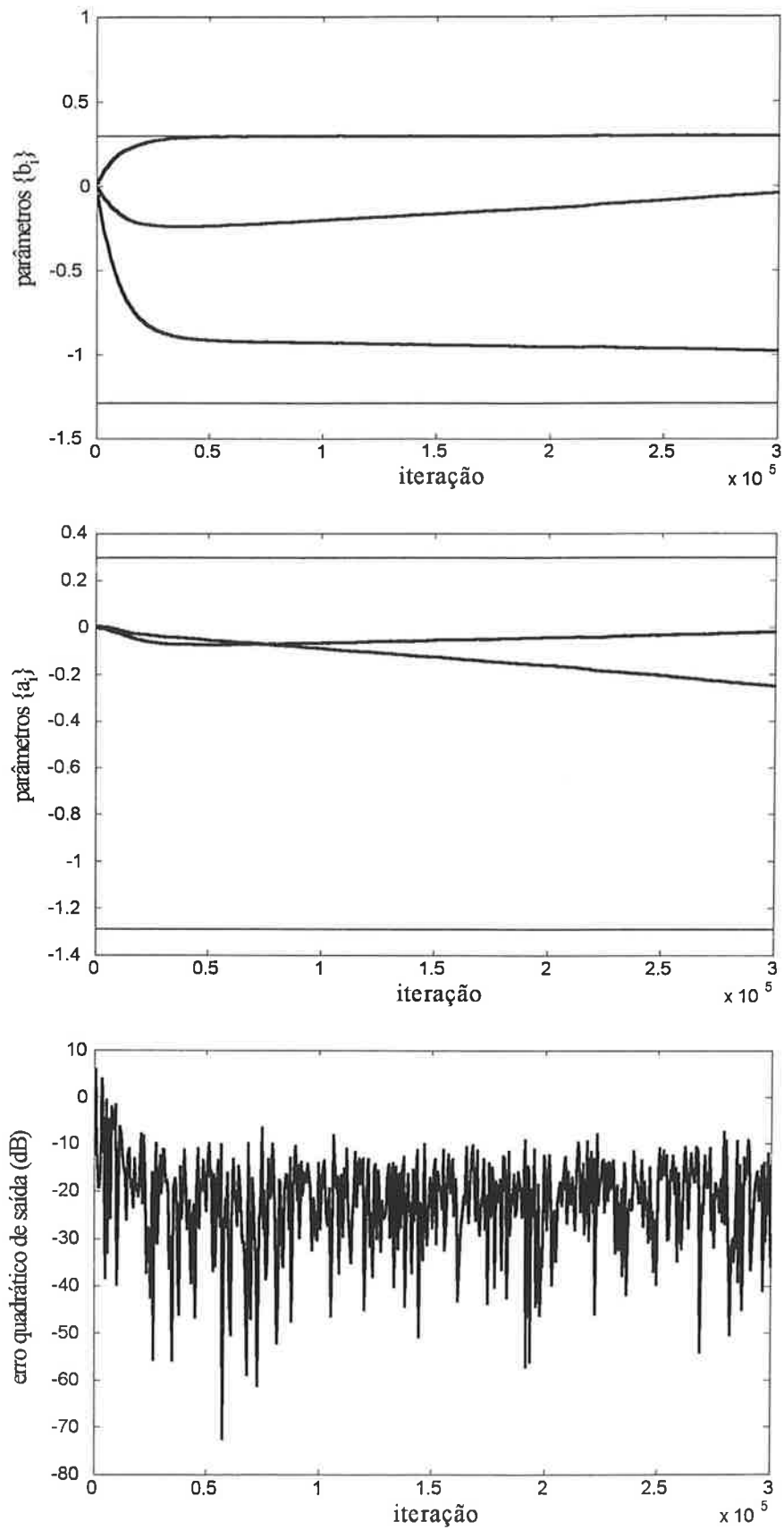


Figura 3-23: Algoritmo SMM; sistema  $H_{15}(z)$  (parâmetros em traços claros)

## Capítulo 4

# Um filtro IIR polifásico adaptativo

Neste capítulo apresenta-se uma estrutura polifásica para filtragem IIR adaptativa e seu desempenho é comparado com o de filtros IIR adaptativos na forma direta [27].

No Item 4.1 a expansão polifásica de uma função racional é formulada e é analisada a questão de sua unicidade. A motivação para a utilização de uma estrutura polifásica em filtragem IIR adaptativa será evidenciada a partir do Item 4.3.

No Item 4.2 discute-se inicialmente duas abordagens possíveis para filtragem IIR polifásica adaptativa e sua relação com a questão da unicidade da expansão polifásica. Considera-se, a seguir, a aplicação à estrutura polifásica dos algoritmos de adaptação de passo constante vistos anteriormente para a estrutura direta. Por fim, é calculada a complexidade computacional de filtros polifásicos adaptativos com algoritmos de passo constante, comparando-a com a de filtros adaptativos na forma direta com algoritmos de passo constante e algoritmos do tipo Newton.

No Item 4.3 a análise da convergência local de algoritmos de passo constante para filtros IIR adaptativos na forma direta, realizada nos Itens 2.2 a 2.6, é estendida para filtros IIR adaptativos na forma polifásica. Mostra-se como as condições de positividade relacionadas à convergência dos algoritmos podem ser menos restritivas no caso polifásico. A análise permite também uma compreensão do possível ganho da estrutura polifásica quanto ao condicionamento da matriz de estados associada a cada algoritmo, o que, como visto, tem um efeito importante na velocidade de convergência local.

Em seguida passa-se a analisar a superfície de erro reduzida para o caso da estrutura polifásica. No Item 4.4 demonstra-se, para a estrutura polifásica, que, com entrada branca e modelamento suficiente, todos os pontos estacionários da superfície de erro são mínimos

globais da mesma. Deve-se notar que esta propriedade não decorre diretamente de propriedades análogas relativas à estrutura direta, já bastante conhecidas, e trata-se portanto de um resultado novo. No Item 4.5, a partir da expressão para a superfície de erro reduzida obtida no item anterior, mostra-se a possível melhoria da mesma em relação ao caso da estrutura direta. Para um dado sistema sendo modelado  $H(z)$ , a planicidade da superfície de erro reduzida em regiões distantes do mínimo global, resultante da existência de polos de  $H(z)$  próximos à circunferência unitária, tende a ser menos pronunciada do que no caso da estrutura direta, o que resulta em uma maior velocidade de convergência global de algoritmos de passo constante.

Finalmente, no Item 4.6 são mostrados exemplos práticos da aplicação da estrutura polifásica, e seu desempenho é comparado com o da estrutura direta.

## 4.1 Expansão polifásica de uma função racional

Neste item é introduzida a expansão polifásica de uma função racional e é analisada a questão da sua unicidade. A motivação para a utilização de uma estrutura polifásica em filtragem IIR adaptativa será evidenciada a partir do Item 4.3.4.

### 4.1.1 Determinação da expansão polifásica pela descrição no espaço de estados

Consideremos a função racional  $H(z)$  dada por

$$H(z) = \frac{C(z)}{D(z)} = \frac{c_0 + c_1z + \dots + c_{M_c} z^{M_c}}{1 + d_1z + \dots + d_{M_d} z^{M_d}}, \quad (4.1)$$

onde  $C(z) = c_0 + c_1z + \dots + c_{M_c} z^{M_c}$ ,  $c_{M_c} \neq 0$ , e  $D(z) = 1 + d_1z + \dots + d_{M_d} z^{M_d}$ ,  $d_{M_d} \neq 0$ , não tem raízes comuns. Considerando  $M = \deg[H(z)]$  como o menor número de retardos necessários para implementar  $H(z)$ , temos então que  $M = \deg[H(z)] = \max(M_c, M_d)$ .

A expansão polifásica de ordem  $p$  de  $H(z)$  é obtida agrupando em  $p$  blocos os termos da resposta ao pulso unitário de  $H(z)$ :

$$H(z) = \sum_{k=-\infty}^{\infty} h_k z^k = \sum_{k=-\infty}^{\infty} \sum_{l=0}^{p-1} h_{pk+l} z^{pk+l} = \sum_{l=0}^{p-1} z^l \sum_{k=-\infty}^{\infty} h_{pk+l} z^{pk} = \sum_{l=0}^{p-1} z^l H_l(z^p), \quad (4.2)$$

onde  $H_l(z) = \sum_{k=-\infty}^{\infty} h_{pk+l} z^k$ . Determinamos a seguir polinômios  $C_l(z)$  e  $D_l(z)$  tais que  $C_l(z)/D_l(z) = H_l(z)$  (ao que consta para o autor, a dedução a seguir não existe na literatura sobre filtros polifásicos).

A partir de uma descrição mínima  $(\mathbf{A}, \mathbf{b}, \mathbf{c}, d)$  de  $H(z)$  no espaço de estados, sua resposta ao pulso unitário pode ser expressa como

$$h_k = \begin{cases} d, & k = 0, \\ \mathbf{c}^\top \mathbf{A}^{k-1} \mathbf{b}, & k > 0. \end{cases} \quad (4.3)$$

Definindo agora

$$\begin{aligned} \mathbf{b}_l &= \mathbf{A}^{p+l-1} \mathbf{b}, \quad l = 0, 1, \dots, p-1, \\ \mathbf{A}_p &= \mathbf{A}^p, \end{aligned} \quad (4.4)$$

temos

$$\mathbf{c}^\top \mathbf{A}^{pk+l-1} \mathbf{b} = \mathbf{c}^\top (\mathbf{A}^p)^{k-1} (\mathbf{A}^{p+l-1} \mathbf{b}) = \mathbf{c}^\top \mathbf{A}_p^{k-1} \mathbf{b}_l, \quad (4.5)$$

e, portanto, podemos exprimir as respostas ao pulso unitário  $h_{l,k} = h_{pk+l}$  de cada fase da expansão polifásica como

$$h_{l,k} = \begin{cases} d_l, & k = 0, \\ \mathbf{c}^\top \mathbf{A}_p^{k-1} \mathbf{b}_l, & k > 0, \end{cases} \quad l = 0, 1, \dots, p-1,$$

onde

$$d_l = \begin{cases} d, & l = 0, \\ \mathbf{c}^\top \mathbf{A}^{l-1} \mathbf{b}, & l = 1, 2, \dots, p-1. \end{cases} \quad (4.6)$$

Portanto,  $(\mathbf{A}_p, \mathbf{b}_l, \mathbf{c}, d_l)$  é uma descrição no espaço de estados de  $H_l(z) = C_l(z)/D_l(z)$ , de onde segue que

$$H_l(z) = d_l + z \mathbf{c}^\top (\mathbf{I} - z \mathbf{A}_p)^{-1} \mathbf{b}_l, \quad l = 0, 1, \dots, p-1. \quad (4.7)$$

Podemos tomar portanto  $C_l(z) = d_l + z \mathbf{c}^\top \text{adj}(\mathbf{I} - z \mathbf{A}_p) \mathbf{b}_l$  e  $D_l(z) = \det(\mathbf{I} - z \mathbf{A}_p)$ . Os denominadores  $D_l(z)$  das funções  $H_l(z)$  são idênticos e para explicitar este fato é usado um mesmo

$D_P(z)$  para representá-los, obtendo então:

$$H(z) = \sum_{l=0}^{p-1} z^l H_l(z^p) = \frac{\sum_{l=0}^{p-1} z^l C_l(z^p)}{D_P(z^p)} = \frac{C_P(z)}{D_P(z^p)}. \quad (4.8)$$

Escrevendo (4.7) como  $H_l(z) = d_l + \mathbf{c}^\top (z^{-1}\mathbf{I} - \mathbf{A}_P)^{-1} \mathbf{b}_l$  vemos que as raízes de  $D_P(z)$  correspondem ao inverso dos auto-valores de  $\mathbf{A}_P$  que, por sua vez, são os auto-valores de  $\mathbf{A}$  elevados à  $p$ -ésima potência (ver (4.4)). Os polos de  $H_l(z)$ , portanto, são os polos de  $H(z)$  elevados à  $p$ -ésima potência. Cada polo  $\nu_j$  de  $H(z)$  resulta então, em  $H_l(z^p)$ , nos polos  $|\nu_j| \angle [\arg(\nu_j) + 2\pi l/p]$ ,  $l = 0, 1, \dots, p-1$ . Os polos correspondentes a  $l = 1, 2, \dots, p-1$  nesta expressão, que perfazem um total de  $(p-1)M_d$  polos, devem ser cancelados por zeros de  $C_P(z)$  que também deve conter os  $M_c$  zeros de  $H(z)$ . Pode-se destacar então:

- $D_P(z^p)$  tem as  $M_d$  raízes  $\nu_j$  de  $D(z)$  mais as  $(p-1)M_d$  raízes adicionais  $|\nu_j| \angle [\arg(\nu_j) + 2\pi l/p]$ ,  $l = 1, 2, \dots, p-1$
- $C_P(z)$  tem as  $M_c$  raízes de  $C(z)$  mais as mesmas  $(p-1)M_d$  raízes adicionais de  $D_P(z^p)$

Pode ser visto ainda que  $\deg[C_P(z)] \leq pM$  :

$$\begin{aligned} \deg[C_P(z)] &= (p-1)M_d + M_c \\ &\leq (p-1) \max(M_d, M_c) + \max(M_d, M_c) = pM, \end{aligned}$$

e, conseqüentemente, as funções  $C_l(z)/D_P(z) = H_l(z)$  tem a forma geral

$$\frac{C_0(z)}{D_P(z)} = \frac{c_{0,0} + c_{0,1}z + \dots + c_{0,M} z^M}{1 + d_{P,1}z + \dots + d_{P,M} z^M} = H_0(z) \quad (4.9)$$

e

$$\frac{C_l(z)}{D_P(z)} = \frac{c_{l,0} + c_{l,1}z + \dots + c_{l,M-1} z^{M-1}}{1 + d_{P,1}z + \dots + d_{P,M} z^M} = H_l(z), \quad l = 1, 2, \dots, p-1, \quad (4.10)$$

onde  $\deg[C_l(z)] < M$  para  $l > 0$  é necessário para que  $\deg[C_P(z)] \leq pM$ . Notar que para todos os  $l$  temos  $\deg[H_l(z)] \leq \deg[H(z)]$ . Mais explicitamente, nas expressões (4.9) e (4.10) coeficientes correspondentes às potências mais elevadas no numerador (respectivamente, no denominador) poderão ser nulos se  $M_d > M_c$  (respectivamente, se  $M_c > M_d$ ).

## 4.1.2 Possível não-unicidade da expansão polifásica

Inicialmente, consideremos a definição de uma nomenclatura apropriada para a discussão posterior:

**Definição 4.1** *Seja uma função racional  $H(z)$  com  $\deg[H(z)] = M$ . Uma função racional*

$$\frac{C_P(z)}{D_P(z^p)} = \frac{c_{P,0} + c_{P,1}z + \dots + c_{P,pM}z^{pM}}{1 + d_{P,1}z^p + \dots + d_{P,M}z^{pM}} \quad (4.11)$$

tal que  $C_P(z)/D_P(z^p) = H(z)$  será denominada uma expansão polifásica de ordem  $p$  de  $H(z)$ . Se, adicionalmente,  $C_P(z)$  e  $D_P(z)$  forem determinados segundo (4.8), (4.7), (4.6) e (4.4), então  $C_P(z)/D_P(z^p)$  será denominada a expansão polifásica principal de ordem  $p$  de  $H(z)$ .

Em relação à expansão polifásica principal, a seguinte propriedade é válida:

**Propriedade 4.1** *Seja  $H(z)$  com  $\deg[H(z)] = M$  e termo direto  $d$ . Uma expansão polifásica de  $H(z)$  dada por*

$$H(z) = \sum_{l=0}^{p-1} z^l \bar{H}_l(z^p) = \frac{\bar{C}_P(z)}{\bar{D}_P(z^p)}, \quad (4.12)$$

com

$$\bar{H}_l(z) = \bar{d}_l + z \bar{c}^\top (\mathbf{I} - z \bar{\mathbf{A}}_P)^{-1} \bar{\mathbf{b}}_l, \quad (4.13)$$

onde  $\dim[\bar{\mathbf{A}}_P] = (M, M)$ , é a expansão polifásica principal de  $H(z)$  se, e somente se, valer, para um certo  $\bar{\mathbf{b}}$ ,

$$\begin{aligned} \bar{\mathbf{b}}_l &= \bar{\mathbf{A}}^{p+l-1} \bar{\mathbf{b}}, \quad l = 0, 1, \dots, p-1, \\ \bar{\mathbf{A}}_P &= \bar{\mathbf{A}}^p, \\ \bar{d}_l &= \begin{cases} d, & l = 0, \\ \bar{c}^\top \bar{\mathbf{A}}^{l-1} \bar{\mathbf{b}}, & l = 1, 2, \dots, p-1. \end{cases} \end{aligned} \quad (4.14)$$

*Prova:* Consideremos  $(\mathbf{A}, \mathbf{b}, \mathbf{c}, d)$  uma representação no espaço de estados mínima de  $H(z)$ , e sejam  $H_l(z)$  as funções da expansão polifásica principal de  $H(z)$ . De (4.8) e (4.12) temos que, necessariamente,  $\bar{H}_l(z) = H_l(z)$ ,  $l = 0, 1, \dots, p-1$ . Podemos escrever, então,



usando (4.13) e (4.14),

$$\begin{aligned}
\bar{\mathbf{c}}^\top \bar{\mathbf{A}}_P^{k-1} \bar{\mathbf{b}}_l &= \bar{\mathbf{c}}^\top (\bar{\mathbf{A}}^p)^{k-1} (\bar{\mathbf{A}}^{p+l-1} \bar{\mathbf{b}}) \\
&= \bar{\mathbf{c}}^\top \bar{\mathbf{A}}^{pk+l-1} \bar{\mathbf{b}} \\
&= \mathbf{c}^\top \mathbf{A}^{pk+l-1} \mathbf{b}, \quad k > 0, \quad l = 0, 1, \dots, p-1,
\end{aligned} \tag{4.15}$$

e

$$\bar{\mathbf{c}}^\top \bar{\mathbf{A}}^{l-1} \bar{\mathbf{b}} = \mathbf{c}^\top \mathbf{A}^{l-1} \mathbf{b}, \quad l = 1, 2, \dots, p-1. \tag{4.16}$$

Temos então que  $(\bar{\mathbf{A}}, \bar{\mathbf{b}}, \bar{\mathbf{c}}^\top, d)$  e  $(\mathbf{A}, \mathbf{b}, \mathbf{c}^\top, d)$  tem a mesma resposta ao pulso unitário e tem a mesma ordem  $M$ . Uma vez que, por hipótese,  $(\mathbf{A}, \mathbf{b}, \mathbf{c}^\top, d)$  é uma representação mínima de  $H(z)$  segue que  $(\bar{\mathbf{A}}, \bar{\mathbf{b}}, \bar{\mathbf{c}}^\top, d)$  também o é, e, portanto, ambas são relacionadas por uma transformação de similaridade. As raízes de  $\bar{\mathbf{A}}$  e  $\mathbf{A}$ , conseqüentemente, são iguais, de onde resulta que  $\bar{D}_P(z) = D_P(z)$  o que, por sua vez, implica em que  $\bar{C}_P(z) = C_P(z)$ , onde  $C_P(z)/D_P(z^p)$  é a expansão polifásica principal de  $H(z)$ .

No sentido contrário, temos que se  $\bar{C}_P(z)/\bar{D}_P(z^p)$  é a expansão polifásica principal de  $H(z)$  então pela Definição 4.1 as condições (4.14) são atendidas. ■

Uma situação onde existem outras expansões polifásicas de ordem  $p$  além da principal ocorre quando, para dois autovalores  $\nu_i$  e  $\nu_j$  de  $\mathbf{A}$  (polos, portanto, de  $H(z)$ ), com  $\nu_i \neq \nu_j$ , valer

$$\nu_i^p = \nu_j^p. \tag{4.17}$$

Isto pode ser visto como segue. Supondo  $\mathbf{A}$  diagonalizável e escrevendo  $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}$ , onde  $\mathbf{Q}$  é a matriz dos autovetores de  $\mathbf{A}$  e  $\mathbf{\Lambda}$  é a matriz diagonal dos seus autovalores, temos que a resposta ao pulso unitário de cada fase da expansão polifásica principal pode ser escrita como

$$\begin{aligned}
h_{l,k} &= \mathbf{c}^\top \mathbf{A}_P^{k-1} \mathbf{b}_l \\
&= \mathbf{c}^\top \mathbf{Q} (\mathbf{\Lambda}^p)^{k-1} \mathbf{Q}^{-1} \mathbf{b}_l \\
&= f_1 g_{l,1} \nu_1^{p(k-1)} + f_2 g_{l,2} \nu_2^{p(k-1)} + \dots + f_M g_{l,M} \nu_M^{p(k-1)},
\end{aligned} \tag{4.18}$$

$$l = 0, 1, \dots, p-1, \quad k > 0,$$

onde

$$[ f_1 \ f_2 \ \dots \ f_M ] = \mathbf{f}^\top = \mathbf{c}^\top \mathbf{Q} \quad (4.19)$$

e

$$[ g_{l,1} \ g_{l,2} \ \dots \ g_{l,M} ]^\top = \mathbf{g}_l = \mathbf{Q}^{-1} \mathbf{b}_l. \quad (4.20)$$

Consideremos agora que os polos  $\nu_{M-1}$  e  $\nu_M$  sejam complexos conjugados e  $\nu_{M-1}^p = \nu_M^p$ . Podemos então escrever

$$\begin{aligned} h_{l,k} &= f_1 g_{l,1} \nu_1^{p(k-1)} + f_2 g_{l,2} \nu_2^{p(k-1)} + \dots + (f_{M-1} g_{l,M-1} + f_M g_{l,M}) \nu_{M-1}^{p(k-1)} \\ &= \mathbf{f}_d^\top (\Lambda_d^p)^{k-1} \mathbf{g}_{d,l}, \end{aligned} \quad (4.21)$$

$$l = 0, 1, \dots, p-1, \quad k > 0,$$

com

$$\mathbf{f}_d = [ 1 \ 1 \ \dots \ 1 ]^\top, \quad (4.22)$$

$$\mathbf{g}_{d,l} = [ f_1 g_{l,1} \ f_2 g_{l,2} \ \dots \ f_{M-1} g_{l,M-1} + f_M g_{l,M} ]^\top, \quad (4.23)$$

$$\Lambda_d^p = \begin{bmatrix} \nu_1^p & 0 & \dots & 0 \\ 0 & \nu_2^p & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \nu_{M-1}^p \end{bmatrix}. \quad (4.24)$$

Temos então que  $(\Lambda_d^p, \mathbf{g}_{d,l}, \mathbf{f}_d, d_l)$  é uma descrição no espaço de estados de uma função  $H_{d,l}(z)$ , com  $\deg[H_{d,l}(z)] < M$ , para a qual vale

$$H_{d,l}(z) = H_l(z), \quad l = 0, 1, \dots, p-1, \quad (4.25)$$

e para a qual não valem, em geral, relações como (4.14). Da igualdade acima segue diretamente que, no caso considerado,

$$\deg[H_l(z)] < M, \quad l = 0, 1, \dots, p-1, \quad (\nu_i^p = \nu_j^p), \quad (4.26)$$

e, também que, no caso considerado, existem infinitas funções  $\overline{H}_l(z)$  com  $2M + 1$  coeficientes e com a forma

$$\overline{H}_l(z) = H_{d,l}(z) \frac{R(z)}{R(z)} = H_l(z), \quad l = 0, 1, \dots, p-1, \quad (4.27)$$

onde  $R(z)$  é qualquer polinômio de fase mínima e de primeira ordem. Portanto, neste caso considerado, existem infinitas expansões polifásicas de ordem  $p$  de  $H(z)$ .

Uma outra maneira de visualizar este exemplo é em termos de polos e zeros: os polos  $\nu_{M-1}$  e  $\nu_M$  de  $H(z)$  resultam, em  $D_P(z)$ , em duas raízes iguais  $\nu_{M-1}^p$  e  $\nu_M^p$  e, portanto, resultam, em  $D_P(z^p)$ , no mesmo conjunto de raízes  $|\nu_{M-1}| \angle [\arg(\nu_{M-1}) + 2\pi l/p]$ ,  $l = 0, 1, \dots, p-1$ . Portanto,  $D_P(z^p)$  tem uma raiz dupla em cada uma destas posições, sendo que ao menos uma delas deve ser cancelada por uma raiz de  $C_P(z)$ . Podemos então construir  $D_d(z^p)$  de ordem  $p(M-1)$  a partir de  $D_P(z^p)$ , transformando estas raízes duplas de  $D_P(z^p)$  em raízes simples, e construir  $C_d(z)$  de ordem  $p(M-1)$  a partir de  $C_P(z)$ , eliminando as raízes de  $C_P(z)$  que cancelariam as raízes eliminadas de  $D_P(z^p)$ .

## 4.2 Filtro polifásico adaptativo

Neste item discute-se inicialmente a questão da utilização da forma polifásica vista no item anterior como um filtro IIR adaptativo. Em seguida, é considerada a aplicação a este filtro polifásico adaptativo dos algoritmos de passo constante já vistos para a forma direta. Por fim, considera-se a questão da complexidade computacional de filtros polifásicos adaptativos.

### 4.2.1 Abordagens para filtragem IIR polifásica adaptativa

Para realizar um filtro polifásico adaptativo  $\widehat{H}(z)$  poderíamos, em uma primeira abordagem, partir de uma implementação  $(\mathbf{A}_P, \mathbf{b}_l, \mathbf{c}, d_l)$  na forma direta transposta (para ter uma menor complexidade computacional do que em uma implementação genérica no espaço de estados) de cada  $\widehat{H}_l(z)$ , com  $\deg[\widehat{H}_l(z)] \leq M$ , onde fossem válidas as relações

$$\begin{aligned} \mathbf{b}_l &= \mathbf{A}^{p+l-1} \mathbf{b}, \quad l = 0, 1, \dots, p-1, \\ \mathbf{A}_P &= \mathbf{A}^p, \\ d_l &= \begin{cases} d, & l = 0, \\ \mathbf{c}^\top \mathbf{A}^{l-1} \mathbf{b}, & l = 1, 2, \dots, p-1, \end{cases} \end{aligned} \quad (4.28)$$

para algum  $\mathbf{b} = [\beta_1 \dots \beta_M]^\top$ . Os  $(2M + 1)$  parâmetros a serem adaptados seriam os  $M$  coeficientes de  $D_P(z)$ , que correspondem ao negativo da primeira coluna de  $\mathbf{A}_P$ , mais os  $M$  coeficientes de  $\mathbf{b}$ , mais o termo direto  $d$ . Uma vez que as condições dadas por (4.14) seriam desta maneira sempre atendidas, teríamos que, pela Propriedade 4.1,  $\hat{H}(z)$  seria sempre a expansão polifásica principal de alguma função racional de ordem menor ou igual a  $M$ . Esta abordagem, no entanto, tem algumas dificuldades, como por exemplo, a obtenção de  $\mathbf{A}$  a partir de  $\mathbf{A}_P$ , em cada instante.

Uma outra abordagem, que não incorre nestas dificuldades, é simplesmente adaptar sem restrições os  $pq + m + 1$  coeficientes da função racional

$$\hat{H}(z) = \frac{b_{P,0} + b_{P,1}z + \dots + b_{P,pq}z^{pq}}{1 + a_{P,1}z^p + \dots + a_{P,m}z^{pm}} = \frac{B_P(z)}{A_P(z^p)}. \quad (4.29)$$

Como será visto no item a seguir e no Item 4.3, nesta abordagem os algoritmos de adaptação para a estrutura direta vistos anteriormente podem ser prontamente estendidos para a estrutura polifásica, assim como a análise de convergência dos mesmos. Por outro lado, como será visto no Item 4.3, se a condição (4.17) for válida então as matrizes de estados associadas aos algoritmos de adaptação considerados poderão ser singulares. Em termos práticos, resulta que para funções  $H(z)$  que estejam suficientemente próximas de atender à condição (4.17) o condicionamento da matriz de estados será ruim e a velocidade local de convergência será baixa.

Consideremos a limitação que isto introduz para a aplicação desta segunda abordagem. Para sistemas com todos os polos reais basta utilizar  $p$  ímpar para garantir que (4.17) nunca seja válida. Por outro lado, para sistemas com polos complexos cujos ângulos absolutos estão no conjunto  $\{\theta_i\}$ ,  $0 \leq \theta_i \leq \pi$ , uma condição necessária para que (4.17) não seja nunca válida é que

$$p \neq \frac{\pi}{\theta_i}, \quad i = 1, 2, \dots, \quad (4.30)$$

e uma condição suficiente para que (4.17) não seja nunca válida é que

$$p < \frac{\pi}{\max\{\theta_i\}}. \quad (4.31)$$

A imposição desta última condição, em termos práticos, implica em limitar a utilização da segunda abordagem da filtragem polifásica adaptativa a casos nos quais o sistema sendo

modelado tem apenas polos de baixa frequência. Para um dado  $p$  que representasse um compromisso adequado entre o ganho de velocidade de convergência e o aumento da complexidade computacional, deveríamos ter então  $\max \{\theta_i\} < \pi/p$ . Esta limitação não eliminaria, no entanto, o interesse prático pela segunda abordagem. É comum em aplicações práticas que os polos do sistema de tempo discreto sendo modelados sejam de baixa frequência [26], o que ocorre quando o período de amostragem é muito mais curto do que as constantes de tempo do sistema de tempo contínuo envolvido. Por exemplo, em cancelamento de eco acústico, para um padrão CD de qualidade de áudio teríamos um período de amostragem de aproximadamente  $22 \mu\text{s}$  ao passo que as constantes de tempo do eco acústico são da ordem de dezenas de milissegundos ou ainda maiores.

Com base nestas considerações e em sua maior simplicidade, adotaremos então esta segunda abordagem à filtragem adaptativa polifásica no restante deste trabalho. Uma investigação da primeira abordagem é deixada para um trabalho futuro.

#### 4.2.2 Forma de $\hat{H}(z)$

Consideremos agora a questão da escolha dos valores de  $q$  e  $m$  em (4.29). Para tanto é necessário, antes, estender a análise realizada no Item 1.5.4 para o caso da estrutura polifásica. Consideremos então os polinômios  $C(z)$  e  $D(z)$ , com  $D(z)$  de fase mínima e  $\deg[D(z)] = M_d$  e  $\deg[C(z)] = M_c$ , e o sinal gerado como

$$y(n) = \frac{C(z)}{D(z)}u(n), \quad (4.32)$$

onde  $u(n)$  é uma excitação persistente de ordem  $\max(pM_d + M_c + 1, pM_c + M_d + 1)$  ou maior. Um vetor polifásico de regressão  $\mathbf{r}_P(n)$  associado a  $C(z)$  e  $D(z)$  é:

$$\mathbf{r}_P(n) = \begin{bmatrix} -y(n-p) \\ -y(n-2p) \\ \vdots \\ -y(n-pM_d) \\ u(n) \\ u(n-1) \\ \vdots \\ u(n-pM_c) \end{bmatrix}. \quad (4.33)$$

Repartindo os elementos de um vetor  $\mathbf{e}_P$  como

$$\mathbf{e}_P = \left[ r_{P,0} \ r_{P,1} \ \cdots \ r_{P,M_d-1} \ q_{P,0} \ q_{P,1} \ \cdots \ q_{P,pM_c} \right]^\top, \quad (4.34)$$

o produto interno  $\mathbf{e}_P^\top \mathbf{r}_P(n)$  pode ser escrito como

$$\mathbf{e}_P^\top \mathbf{r}_P(n) = \left[ Q_P(z) - z^p R_P(z^p) \frac{C(z)}{D(z)} \right] u(n), \quad (4.35)$$

onde

$$R_P(z) = r_{P,0} + r_{P,1}z + \dots + r_{P,M_d-1}z^{M_d-1} \quad (4.36)$$

e

$$Q_P(z) = q_{P,0} + q_{P,1}z + \dots + q_{P,pM_c}z^{pM_c} \quad (4.37)$$

. A seguinte propriedade é válida:

**Propriedade 4.2** Com  $\mathbf{r}_P(n)$  dado por (4.33) e (4.32), existe um vetor  $\mathbf{e}_P$  tal que  $\mathbf{e}_P^\top \mathbf{r}_P(n) \equiv 0$  somente se  $C(z)$  e  $D(z)$  tem uma outra expansão polifásica além da principal.

*Prova:* Em (4.35),  $Q_P(z) - z^p R_P(z^p) C(z)/D(z)$  tem, no máximo,  $\max(pM_d + M_c, pM_c + M_d)$  zeros. Uma vez que  $u(n)$  tem excitação persistente de ordem maior que este número máximo de zeros, temos que  $\mathbf{e}_P^\top \mathbf{r}_P(n) \equiv 0$  se, e somente se,  $Q_P(z) - z^p R_P(z^p) C(z)/D(z) \equiv 0$ , ou seja,

$$\frac{Q_P(z)}{z^p R_P(z^p)} = \frac{C(z)}{D(z)}. \quad (4.38)$$

Por hipótese,  $D(z)$  é de fase mínima e, portanto, não pode ter  $z = 0$  como raiz. O termo  $z^p$  do denominador tem que ser cancelado então por  $Q_P(z) = z^p Q'_P(z)$  resultando em

$$\frac{Q'_P(z)}{R_P(z^p)} = \frac{C(z)}{D(z)}, \quad (4.39)$$

onde  $\deg[Q'_P(z)] = p(M_c - 1)$  e  $\deg[R_P(z^p)] = p(M_d - 1)$ . A função  $Q'_P(z)/R_P(z^p)$  é uma expansão polifásica de  $H(z)$  mas não é a sua expansão polifásica principal  $C_P(z)/D_P(z^p)$ , pois, como visto no Item 4.1,  $\deg[C_P(z)] = pM_c$  e  $\deg[D_P(z^p)] = pM_d$ . Portanto se  $\mathbf{e}_P^\top \mathbf{r}_P(n) \equiv 0$  existe uma expansão polifásica de  $H(z)$  além da principal. ■

Analogamente ao que foi visto para o caso da estrutura direta, será visto adiante que o fato de  $v(n) = \mathbf{e}_P^\top \mathbf{r}_P(n)$  não poder ser identicamente nulo é necessário para que o valor correto dos parâmetros seja um ponto de convergência dos algoritmos de adaptação. Pela propriedade acima, se  $H(z)$  só admitir a expansão polifásica principal de ordem  $p$ , então  $v(n)$  não poderá ser identicamente nulo. Portanto, a mesma discussão relativa às restrições necessárias para impedir a existência de outras expansões polifásicas, feita no item anterior, é aplicável aqui.

Além disso, como no caso da estrutura direta (ver Item 2.3.4), a possível nulidade de  $v(n)$  está relacionada com a questão do remodelamento, o que nos traz à questão inicial da escolha dos valores de  $q$  e  $m$  em (4.29). A forma de  $\mathbf{r}_P(n)$  em (4.33) implica em que, para aplicar a Propriedade 1.3 na análise da convergência dos algoritmos de adaptação, tenhamos  $q = M_c$  e  $m = M_d$ . Se, diferentemente, valer  $m > M_d$ , ou seja, se houver um remodelamento do número de polos de  $H(z)$  então a Propriedade 1.3 não será aplicável, e mesmo que a expansão polifásica principal seja única poderá ocorrer que  $v(n) = \mathbf{e}_P^\top \mathbf{r}_P(n)$  seja identicamente nulo.

Como para a estrutura direta, a questão do remodelamento não será abordada na análise da estrutura polifásica, tendo sido dado prioridade à questão do efeito da proximidade dos polos à circunferência unitária. Considera-se então, no que segue, que  $m = M_d$ , e, também, para maior simplicidade,  $q = m$  e  $M_c \leq M_d$ , o que resulta em  $q = m = \max(M_c, M_d) = M$ , e na seguinte forma para  $\hat{H}(z)$ :

$$\hat{H}(z) = \frac{b_{P,0} + b_{P,1}z + \dots + b_{P,pM}z^{pM}}{1 + a_{P,1}z^p + \dots + a_{P,M}z^{pM}} = \frac{B_P(z)}{A_P(z^p)}. \quad (4.40)$$

### 4.2.3 Algoritmos de passo constante

A seguir apresenta-se a aplicação à estrutura polifásica dos algoritmos de passo constante RG, SMM, PLR e SHARF, vistos nos Itens 2.3, 2.4, 2.5 e 2.5.3, respectivamente, destinados originalmente à estrutura direta de implementação. Os algoritmos para a estrutura polifásica são então enquadrados em uma formulação genérica como (2.7), que servirá para a análise de sua convergência.

#### Algoritmo do gradiente recursivo (RG)

Segue diretamente da obtenção do algoritmo RG para a forma direta, apresentada no Item 2.3, que, no caso do filtro polifásico com fator de expansão  $p$ , considerando a forma de  $\hat{H}(z)$  dada por (4.40), o vetor de adaptação  $\psi_P(n)$  é dado por

$$\psi_P(n) = \begin{bmatrix} -\nabla_{P,a}(n-p) \\ -\nabla_{P,a}(n-2p) \\ \vdots \\ -\nabla_{P,a}(n-pM) \\ \nabla_{P,b}(n) \\ \nabla_{P,b}(n-1) \\ \vdots \\ \nabla_{P,b}(n-pM) \end{bmatrix}, \quad (4.41)$$

com

$$\nabla_{P,a}(n) = \frac{\hat{y}(n)}{A_P(z^p, n)} \quad \text{e} \quad \nabla_{P,b}(n) = \frac{u(n)}{A_P(z^p, n)}. \quad (4.42)$$

A adaptação dos coeficientes do numerador e do denominador é dada então por

$$\begin{aligned} a_{P,i}(n+1) &= a_i(n) - \mu \nabla_{P,a}(n-pi)e(n), \quad i = 1, 2, \dots, M, \\ b_{P,i}(n+1) &= b_i(n) + \mu \nabla_{P,b}(n-i)e(n), \quad i = 0, 1, 2, \dots, pM, \end{aligned} \quad (4.43)$$

onde  $e(n)$  é o erro de saída

$$e(n) = y(n) - \hat{H}(z)u(n). \quad (4.44)$$



### Algoritmo Steiglitz-McBride (SMM)

Como visto no Item 2.4, as fórmulas de adaptação dos coeficientes segundo o algoritmo SMM resultam de aproximações das derivadas

$$\frac{\partial}{\partial a_{i,k+1}} \frac{A_{k+1}(z)}{A_k(z)} \quad \text{e} \quad \frac{\partial}{\partial b_{i,k+1}} \frac{B_{k+1}(z)}{A_k(z)}, \quad (4.45)$$

onde o índice  $k$  indica a iteração na identificação *off-line* do sistema. Se for assumido agora que os polinômios acima tem a forma polifásica, resulta em um vetor de adaptação  $\psi_P(n)$  com a mesma forma de (4.41) e expressões de adaptação dos coeficientes com a mesma forma de (4.43), mas agora com

$$\nabla_{P,a}(n) = \frac{y(n)}{A_P(z^p, n)} \quad \text{e} \quad \nabla_{P,b}(n) = \frac{u(n)}{A_P(z^p, n)}. \quad (4.46)$$

### Algoritmos PLR e SHARF

Da mesma maneira como feito no Item 2.5 para a forma direta, uma aproximação adicional do gradiente do erro quadrático de saída pode ser obtida eliminando a filtragem por  $1/A_P(z, n)$  no algoritmo RG e obtendo, para o algoritmo PLR,

$$\psi_P(n) = \begin{bmatrix} -\hat{y}(n-p) \\ -\hat{y}(n-2p) \\ \vdots \\ -\hat{y}(n-pM) \\ u(n) \\ u(n-1) \\ \vdots \\ u(n-pM) \end{bmatrix}, \quad (4.47)$$

e

$$\begin{aligned} a_{P,i}(n+1) &= a_{P,i}(n) - \mu \hat{y}(n-pi)e(n), \quad i = 1, 2, \dots, M, \\ b_{P,i}(n+1) &= b_{P,i}(n) + \mu u(n-i)e(n), \quad i = 0, 1, 2, \dots, pM. \end{aligned} \quad (4.48)$$

Para o algoritmo SHARF as expressões de adaptação são obtidas das expressões acima,

apenas trocando o erro de saída  $e(n)$  pelo erro de saída filtrado

$$\epsilon_P(n) = P_P(z^p)e(n), \quad (4.49)$$

onde a razão para adotar-se um filtro de saída com a forma  $P_P(z^p)$  ficará mais clara adiante.

### Forma genérica dos algoritmos de adaptação

Pelo exposto acima podemos ver que os algoritmos de passo constante aplicados à estrutura polifásica podem ser colocados na seguinte formulação comum:

$$\mathbf{w}_P(n+1) = \mathbf{w}_P(n) + \mu \boldsymbol{\psi}_P(n) \epsilon_P(n), \quad (4.50)$$

onde o vetor de parâmetros  $\mathbf{w}_P$  agora tem dimensão  $(p+1)M+1$  e é dado por

$$\mathbf{w}_P = \begin{bmatrix} a_{P,1} \\ a_{P,2} \\ \vdots \\ a_{P,M} \\ b_{P,0} \\ b_{P,1} \\ \vdots \\ b_{P,pM} \end{bmatrix}, \quad (4.51)$$

$\epsilon_P(n) = P_P(z^p)e(n)$  é o erro de saída filtrado,  $P_P(z^p)$  e o vetor  $\boldsymbol{\psi}_P(n)$  dependem do algoritmo específico e  $\mu$  é o passo de adaptação.

#### 4.2.4 Complexidade computacional de filtros adaptativos polifásicos

Neste Item são comparadas as complexidades computacionais do algoritmo RG aplicado à estrutura polifásica e à estrutura direta e da versão Newton do algoritmo RG aplicado à estrutura direta. Quando são considerados os demais algoritmos (SMM, PLR e SHARF) os resultados desta comparação são essencialmente os mesmos do que os relativos ao algoritmo RG.

Passo	+	×
$\nabla_b(n) = u(n) - \sum_{i=1}^M a_i(n) \nabla_b(n-i)$	$M$	$M$
$\hat{y}(n) = \sum_{i=0}^M b_i(n) \nabla_b(n-i)$	$M$	$M+1$
$\nabla_a(n) = \hat{y}(n) - \sum_{i=1}^M a_i(n) \nabla_a(n-i)$	$M$	$M$
$\mu e(n) = \mu[y(n) - \hat{y}(n)]$	1	1
$b_i(n+1) = b_i(n) + \mu e(n) \nabla_b(n-i)$ $i = 0, 1, \dots, M$	$M+1$	$M+1$
$a_i(n+1) = a_i(n) - \mu e(n) \nabla_a(n-i)$ $i = 1, 2, \dots, M$	$M$	$M$
<b>Total</b>	$5M+2$	$5M+3$

Tabela 4.1: Somas e multiplicações necessárias para algoritmo RG aplicado à estrutura direta;  $M$ : ordem do filtro adaptativo.

Passo	+	×
$\nabla_b(n) = u(n) - \sum_{i=1}^M a_{P,i}(n) \nabla_b(n-ip)$	$M$	$M$
$\hat{y}(n) = \sum_{i=0}^{pM} b_{P,i}(n) \nabla_b(n-i)$	$pM$	$pM+1$
$\nabla_a(n) = \hat{y}(n) - \sum_{i=1}^M a_{P,i}(n) \nabla_a(n-ip)$	$M$	$M$
$\mu e(n) = \mu[y(n) - \hat{y}(n)]$	1	1
$b_{P,i}(n+1) = b_{P,i}(n) + \mu e(n) \nabla_b(n-i)$ $i = 0, 1, \dots, pM$	$pM+1$	$pM+1$
$a_{P,i}(n+1) = a_{P,i}(n) - \mu e(n) \nabla_a(n-ip)$ $i = 1, 2, \dots, M$	$M$	$M$
<b>Total</b>	$(2p+3)M+2$	$(2p+3)M+3$

Tabela 4.2: Somas e multiplicações necessárias para algoritmo RG aplicado à estrutura polifásica;  $M$ : ordem do filtro adaptativo;  $p$ : fator de expansão polifásica.

Na Tabela 4.1 temos, para um filtro IIR adaptativo de ordem  $M$ , o número necessário de somas e de multiplicações para cada passo do algoritmo RG aplicado à estrutura direta, e na Tabela 4.2, à estrutura polifásica com fator de expansão  $p$ .

Para o caso do algoritmo do tipo Newton, temos, inicialmente, a forma genérica de adaptação dada por

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \mathbf{P}(n+1) \boldsymbol{\psi}(n) e(n) \quad (4.52)$$

e

$$\mathbf{P}(n+1) = \frac{1}{\lambda} \left[ \mathbf{P}(n) - \frac{\mathbf{P}(n) \boldsymbol{\psi}(n) \boldsymbol{\psi}^\top(n)}{\lambda/\mu + \boldsymbol{\psi}^\top(n) \mathbf{P}(n) \boldsymbol{\psi}(n)} \right], \quad (4.53)$$

Passo	+	×	÷
$\nabla_b(n) = u(n) - \sum_{i=1}^M a_i(n) \nabla_b(n-i)$	$M$	$M$	0
$\hat{y}(n) = \sum_{i=0}^M b_i(n) \nabla_b(n-i)$	$M$	$M+1$	0
$\nabla_a(n) = \hat{y}(n) - \sum_{i=1}^M a_i(n) \nabla_a(n-i)$	$M$	$M$	0
$e(n) = y(n) - \hat{y}(n)$	1	0	0
$\mathbf{q} = \mathbf{P}\boldsymbol{\psi}$	$N(N-1)$	$N^2$	0
$s = \lambda/\mu + \boldsymbol{\psi}^\top \mathbf{q}$	$N$	$N$	0
$\mathbf{k} = (1/s)\mathbf{q}$	0	$N$	1
$\mathbf{w}(n+1) = \mathbf{w}(n) + \mathbf{k}e(n)$	$N$	$N$	0
$\mathbf{M} = (1/s)\mathbf{q}\mathbf{q}^\top$ (*)	0	$N^2 + N$	0
$[\mathbf{P} - \mathbf{M}](1/\lambda)$	$(N^2 + N)/2$	$(N^2 + N)/2$	0
<b>Total</b>	$\frac{3}{2}(N^2 + N) + N + M$	$\frac{5}{2}(N^2 + N) + 3N + M$	1

Tabela 4.3: Somas e multiplicações necessárias para algoritmo Newton/RG aplicado à estrutura direta;  $M$ : ordem do filtro adaptativo,  $N = 2M + 1$ ; (\*): este passo é necessário ao invés de  $\mathbf{M} = \mathbf{L}\mathbf{Q}^\top$  devido a questões de estabilidade numérica.

Algoritmo/Estrutura	$M = 2$		$M = 3$		$M = 4$	
	+	×	+	×	+	×
RG/direta	12	13	17	18	22	23
RG/polifásica ( $p = 4$ )	24	25	35	36	46	47
Newton RG/direta	52	92	94	164	148	256

Tabela 4.4: Somas e multiplicações necessárias para os algoritmos indicados;  $M$ : ordem do filtro adaptativo.

onde  $\lambda = 1 - \mu$  é o fator de esquecimento e

$$\mu \mathbf{P}(n+1) \boldsymbol{\psi}(n) e(n) = \frac{\mathbf{P}(n) \boldsymbol{\psi}(n)}{\lambda/\mu + \boldsymbol{\psi}^\top(n) \mathbf{P}(n) \boldsymbol{\psi}(n)}. \quad (4.54)$$

Os passos do algoritmo e o número necessário de somas, multiplicações e divisões estão na Tabela 4.3, onde para maior conveniência usou-se  $N = 2M + 1$ . Não foram consideradas versões rápidas do algoritmo pois para os baixos valores de  $M$  utilizados sua complexidade computacional é, na realidade, maior.

Com as expressões obtidas acima, foi calculado o número de operações necessárias para cada caso, considerando  $M = 2, 3, 4$  e  $p = 4$ . Os resultados estão na Tabela 4.4.

Como pode ser visto, a utilização da estrutura polifásica, para valores não muito elevados de  $p$ , resulta em uma complexidade computacional bastante menor do que a de um algoritmo do tipo Newton. Ao mesmo tempo, como será visto nos itens a seguir, pode

proporcionar um grande aumento na velocidade de convergência em relação à estrutura direta. O compromisso entre velocidade de convergência e complexidade computacional da estrutura polifásica, pode, portanto, ser bastante vantajoso em relação às outras duas alternativas.

### 4.3 Análise da convergência e do condicionamento da matriz de estados

Neste item, a análise da convergência de algoritmos de passo constante para filtros IIR adaptativos na forma direta, apresentada no Item 2.2, é estendida para filtros IIR polifásicos adaptativos. Analisa-se, em seguida, de que modo são afetadas as condições de convergência de cada algoritmo de adaptação, vistas nos Itens 2.3 a 2.5.3. Finalmente, é estendida para o caso polifásico a análise, feita no Item 2.6, do efeito da proximidade à circunferência unitária dos polos do sistema sendo modelado sobre o condicionamento das matrizes de estados dos algoritmos de adaptação.

#### 4.3.1 Matrizes de estado da ODE no caso polifásico

Assim como foi visto para a estrutura direta, o processo de adaptação dado por (4.50) também pode ser aproximado pela solução de uma equação diferencial como (2.8). Para reformular agora esta nova equação diferencial em termos do erro de parâmetros, assim como feito para a estrutura direta, escrevemos inicialmente  $H(z)$  em sua forma polifásica:

$$H(z) = \frac{C_P(z)}{D_P(z^p)} = \frac{c_{P,0} + c_{P,1}z + \dots + c_{P,pM}z^{pM}}{1 + d_{P,1}z^p + \dots + d_{P,M}z^{pM}}, \quad (4.55)$$

com  $\deg[H(z)] = M$ . Temos então, para a saída do sistema antes da adição do ruído de medida  $\zeta(n)$ ,

$$y'(n) = y(n) - \zeta(n) = \frac{C_P(z)}{D_P(z^p)}u(n) \quad (4.56)$$

e portanto

$$y'(n)D_P(z^p) = C_P(z)u(n). \quad (4.57)$$

Resulta a seguinte identidade:

$$\begin{aligned}
& [C_P(z) - B_P(z)] \frac{u(n)}{A_P(z^p)} - [D_P(z^p) - A_P(z^p)] \frac{y'(n)}{A_P(z^p)} \\
&= \frac{C_P(z)}{A_P(z^p)} u(n) - \hat{y}(n) - \frac{D_P(z^p)}{A_P(z^p)} y'(n) + y'(n) \\
&= y'(n) - \hat{y}(n) \\
&= e(n, \mathbf{w}_P) - \zeta(n).
\end{aligned} \tag{4.58}$$

Portanto, o erro de saída filtrado pode ser escrito em forma matricial como

$$\epsilon_P(n, \mathbf{w}_P) = P_P(z^p) e(n, \mathbf{w}_P) = \boldsymbol{\phi}_{P,1}^\top(n, \mathbf{w}_P) \tilde{\mathbf{w}}_P + P_P(z^p) \zeta(n), \tag{4.59}$$

onde

$$\boldsymbol{\phi}_{P,1}(n, \mathbf{w}_P) = \begin{bmatrix} -y'(n-p)P_P(z^p)/A_P(z^p) \\ -y'(n-2p)P_P(z^p)/A_P(z^p) \\ \vdots \\ -y'(n-pM)P_P(z^p)/A_P(z^p) \\ u(n)P_P(z^p)/A_P(z^p) \\ u(n-1)P_P(z^p)/A_P(z^p) \\ \vdots \\ u(n-pM)P_P(z^p)/A_P(z^p) \end{bmatrix} \tag{4.60}$$

e

$$\tilde{\mathbf{w}}_P = \begin{bmatrix} d_{P,1} - a_{P,1} \\ d_{P,2} - a_{P,2} \\ \vdots \\ d_{P,M} - a_{P,M} \\ c_{P,0} - b_{P,0} \\ c_{P,1} - b_{P,1} \\ \vdots \\ c_{P,pM} - b_{P,pM} \end{bmatrix}. \tag{4.61}$$

Podemos perceber que a razão de ter sido adotado em (4.49) a forma  $P_P(z^p)$  para  $P(z)$  é fazer com que  $P(z)$  possa, idealmente, cancelar o termo  $A_P(z^p)$  em (4.60), analogamente ao que ocorre no caso da estrutura direta.

Assim como visto para a forma direta, o erro de saída também pode ser escrito como

$$\epsilon(n, \mathbf{w}) = P_P(z^p)e(n, \mathbf{w}_P) = \boldsymbol{\phi}_{P,2}^\top(n, \mathbf{w}_P)\tilde{\mathbf{w}}_P + P_P(z^p)\zeta(n), \quad (4.62)$$

onde

$$\boldsymbol{\phi}_{P,2}(n, \mathbf{w}_P) = \begin{bmatrix} -\hat{y}(n-p)P_P(z^p)/D_P(z^p) \\ -\hat{y}(n-2p)P_P(z^p)/D_P(z^p) \\ \vdots \\ -\hat{y}(n-pM)P_P(z^p)/D_P(z^p) \\ u(n)P_P(z^p)/D_P(z^p) \\ u(n-1)P_P(z^p)/D_P(z^p) \\ \vdots \\ u(n-pM)P_P(z^p)/D_P(z^p) \end{bmatrix}. \quad (4.63)$$

Para a estrutura polifásica temos então que as matrizes

$$\mathbf{R}_P(\mathbf{w}_P) = E [\boldsymbol{\psi}_P(n, \mathbf{w}_P)\boldsymbol{\phi}_{P,1}^\top(n, \mathbf{w}_P)] \quad (4.64)$$

e

$$\mathbf{R}'_P(\mathbf{w}_P) = E [\boldsymbol{\psi}_P(n, \mathbf{w}_P)\boldsymbol{\phi}_{P,2}^\top(n, \mathbf{w}_P)] \quad (4.65)$$

desempenham, respectivamente, o mesmo papel que as matrizes  $\mathbf{R}(\mathbf{w})$  e  $\mathbf{R}'(\mathbf{w})$  no caso da estrutura direta. Portanto, toda a análise feita no Item 2.2 (que, especificamente, abordou os aspectos: condição para convergência local; domínio de atração; limites para o passo de adaptação; aproximação local; velocidade de convergência local) é válida para o caso polifásico, apenas devendo-se trocar  $\mathbf{w}$  por  $\mathbf{w}_P$ ,  $\boldsymbol{\psi}$  por  $\boldsymbol{\psi}_P$ ,  $\mathbf{R}(\mathbf{w})$  por  $\mathbf{R}_P(\mathbf{w}_P)$ ,  $\mathbf{R}'(\mathbf{w})$  por  $\mathbf{R}'_P(\mathbf{w}_P)$ , e proceder analogamente para os demais vetores e matrizes utilizados na análise. No próximo item isto é feito para analisar a questão das condições de convergência e do domínio de atração para o filtro polifásico adaptativo, e, no item subsequente, para analisar o espalhamento dos valores singulares de  $\mathbf{R}_P(\mathbf{w}_P^*)$ .

### 4.3.2 Condições para convergência e domínio de atração

#### Algoritmo RG

Com base na discussão do item anterior sobre o papel de  $\mathbf{R}'_P(\mathbf{w}_P)$ , deseja-se que, analogamente ao caso da estrutura direta,  $\mathbf{R}'_P(\mathbf{w}_P)$  seja positiva definida. Usando as definições (4.34), (4.36) e (4.37) e a forma de  $\hat{H}(z)$  dada por (4.40), devemos ter então

$$\begin{aligned}
 \mathbf{e}_P^\top \mathbf{R}'_P(\mathbf{w}_P) \mathbf{e}_P &= E[\mathbf{e}_P^\top \boldsymbol{\psi}_P(n, \mathbf{w}_P) \boldsymbol{\phi}_{P,2}^\top(n, \mathbf{w}_P) \mathbf{e}_P] \\
 &= E \left[ \frac{v_P(n)}{A_P(z^p)} \frac{v_P(n)}{D_P(z^p)} \right] \\
 &= \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{A_P(e^{j\omega p})} \frac{1}{D_P(e^{-j\omega p})} \mathcal{S}_{v,P}(e^{j\omega}) d\omega \\
 &> 0
 \end{aligned} \tag{4.66}$$

onde

$$v_P(n) = [Q_P(z) - z^p R_P(z^p) \frac{B_P(z)}{A_P(z^p)}] u(n). \tag{4.67}$$

Escrevendo agora

$$\begin{aligned}
 \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{A_P(e^{j\omega})} \frac{\mathcal{S}_{v,P}(e^{j\omega})}{D_P(e^{-j\omega p})} d\omega &= \frac{1}{2\pi} \int_0^{2\pi} \operatorname{Re} \left[ \frac{A_P(e^{j\omega p})}{D_P(e^{j\omega p})} \right] \left| \frac{1}{A_P(e^{j\omega p})} \right|^2 \mathcal{S}_{v,P}(e^{j\omega}) d\omega \\
 &= \frac{1}{2\pi p} \int_0^{2\pi p} \operatorname{Re} \left[ \frac{A_P(e^{j\omega})}{D_P(e^{j\omega})} \right] \left| \frac{1}{A_P(e^{j\omega})} \right|^2 \mathcal{S}_{v,P}(e^{j\omega/p}) d\omega,
 \end{aligned} \tag{4.68}$$

segue a propriedade:

**Propriedade 4.3** *Uma condição suficiente para que, no caso do algoritmo RG aplicado à estrutura polifásica,  $\mathbf{R}'_P(\mathbf{w}_P)$  seja positiva definida é que*

$$\operatorname{Re} \left[ \frac{A_P(e^{j\omega})}{D_P(e^{j\omega})} \right] > 0, \quad -\pi < \omega \leq \pi, \tag{4.69}$$



e também que

$$\mathcal{S}_{v,P}(e^{j\omega}) > 0. \quad (4.70)$$

Pode-se notar que (4.67) tem a mesma forma que (4.35). Portanto, pela Propriedade 4.2 e pela discussão que segue esta propriedade no Item 4.2.2, temos que se  $H(z)$  só tem a expansão polifásica principal, com parâmetros dados por  $\mathbf{w}_P^*$ , então, analogamente ao caso da estrutura direta, sempre existirá uma vizinhança de  $\mathbf{w}_P^*$  na qual (4.69) e (4.70) serão atendidas para qualquer sinal  $u(n)$  que seja uma excitação persistente de ordem  $2pM + 1$  ou maior.

Analogamente ao discutido no Item 2.3, a condição de positividade 4.69, como a condição (2.108) para o caso da estrutura direta, pode ser usada para caracterizar parcialmente o domínio de atração de  $\mathbf{w}_P^*$  para o filtro polifásico. Também foi visto no Item 2.3 que quando  $H(z)$  tinha polos próximos à circunferência unitária a extensão do domínio de atração garantida pelo atendimento à condição de positividade (2.108) era pequena. Temos, quanto a este aspecto, então, uma possível melhora devido ao uso da estrutura polifásica: as raízes de  $D_P(z)$  em (4.69), conforme visto no Item 4.1, são os polos de  $H(z)$  elevados à  $p$ -ésima potência, e, portanto, estão mais afastadas da circunferência unitária. Sendo assim, para um dado sistema  $H(z)$ , para poder-se garantir a convergência ao valor correto dos parâmetros, o erro tolerado na posição inicial dos polos de  $\hat{H}(z)$  poderá ser maior no caso da estrutura polifásica do que no caso da estrutura direta.

Este ganho pode ser exemplificado para os mesmos casos de  $D(z^{-1})$  com raízes em  $0,7 \angle \pm 15^\circ$  e  $0,9 \angle \pm 15^\circ$  usados no caso da estrutura direta. Para estes casos e um fator  $p = 4$ , a região para as raízes de  $A_P(z^p)$  (considerando raízes conjugadas e, das  $pM$  raízes de  $A_P(z^p)$ , apenas aquelas que deveriam corresponder às raízes de  $D(z)$ ) na qual a condição de positividade (4.69) é atendida e a parcela que pode ser garantida do domínio de atração de  $\mathbf{w}_P^*$  estão representados na Figura 4-1 (comparar com a Figura 2-1). Notar que, com  $\theta$  denotando o ângulo dos polos de  $\hat{H}(z)$ , foi considerada apenas a região  $\theta < \pi/p$ , uma vez que, conforme discutido no Item 4.1, a aplicação desta restrição aos polos de  $H(z)$  é uma condição suficiente para a existência apenas da expansão polifásica principal.

### Algoritmo SMM

Como no caso da estrutura direta, seguindo um procedimento análogo ao acima, pode-se ver que, para valer a positividade de  $\mathbf{R}_P(\mathbf{w}_P)$ , (4.70) deve ser obedecida e que nenhuma restrição

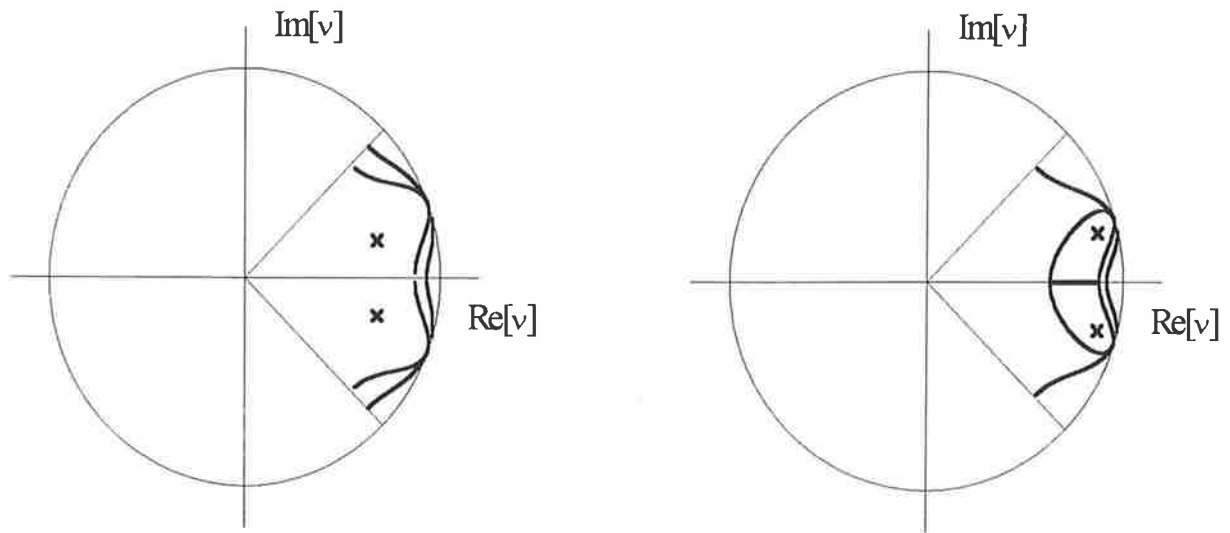


Figura 4-1: Algoritmo RG,  $p = 4$ , sistemas de segunda ordem (polos de  $H(z^{-1})$  indicados com  $\times$ ): região das raízes de  $A(z^{-1})$  na qual a condição de positividade é obedecida e a imagem da maior bola no espaço de parâmetros contida nesta região

aos valores de  $\mathbf{w}_P$  é colocada.

### Algoritmos PLR e SHARF

Seguindo um procedimento análogo ao usado para o algoritmo RG, temos que, para que  $\mathbf{R}'_P(\mathbf{w}_P)$  seja positiva definida, (4.70) deve ser obedecida e a condição de positividade a seguir deve ser atendida:

$$\operatorname{Re} \left[ \frac{1}{D_P(e^{j\omega})} \right] > 0, \quad -\pi < \omega \leq \pi. \quad (4.71)$$

Assim como no caso do algoritmo RG, há então uma possível melhora quanto à restrição imposta por esta condição de positividade, isto é, a região para os polos de  $H(z)$  para a qual pode-se garantir a convergência do filtro adaptativo pode ser menos restritiva no caso da estrutura polifásica do que no caso da estrutura direta.

Para  $H(z)$  de segunda ordem, isto está exemplificado na Figura 4-2, onde está indicada a região para os polos de  $H(z)$  (considerando polos conjugados) na qual a condição acima é atendida, para fatores  $p = 2$  e  $p = 4$  (comparar com a Figura 2-3). Notar, como discutido para o algoritmo RG, que foi considerada apenas a região  $\theta < \pi/p$ . Para  $H(z)$  de quarta ordem, na Figura 4-3 temos a região na qual deve estar um dos pares de polos conjugados

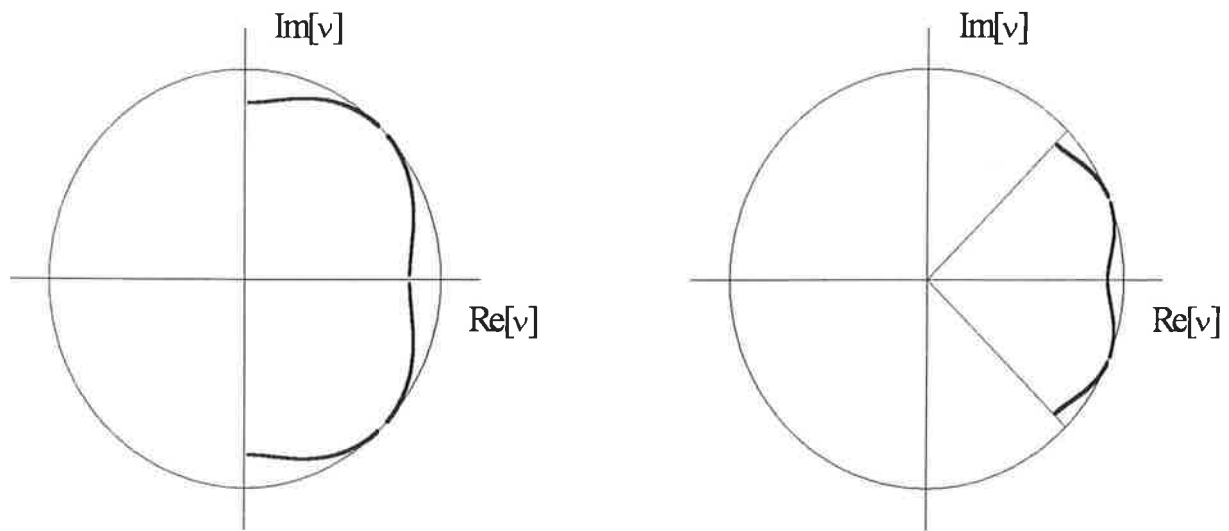


Figura 4-2: Algoritmo PLR, sistema de segunda ordem: região dos polos de  $H(z^{-1})$  onde a condição de positividade é obedecida; à esquerda:  $p = 2$ , à direita:  $p = 4$

de  $H(z)$  dado que o outro par está em  $0,7\angle \pm 30^\circ$  ou em  $0,95\angle \pm 30^\circ$ , para  $p = 4$  (comparar com a Figura 2-4).

Para o algoritmo SHARF, a expressão (4.71) se transforma em

$$\operatorname{Re} \left[ \frac{P_P(e^{j\omega})}{D_P(e^{j\omega})} \right] > 0, \quad -\pi < \omega \leq \pi. \quad (4.72)$$

Como no caso da estrutura direta, o atendimento desta condição está relacionada, no caso do algoritmo SHARF, com a convergência global do algoritmo, e, analogamente ao que foi visto acima, pode ser menos restritiva do que no caso da estrutura direta.

Isto pode ser visto tomando os mesmos exemplos adotados para o caso RG. Como a expressão acima tem a mesma forma que (4.69), temos que a região para as raízes de  $P_P(z^p)$  (considerando raízes conjugadas e, das  $pM$  raízes de  $P_P(z^p)$ , apenas aquelas que deveriam corresponder às raízes de  $D(z)$ ) na qual a condição acima é atendida é dada também pelos gráficos da Figura (4-1).

#### 4.3.4 Expressões para a análise dos valores singulares de $\mathbf{R}_P(\mathbf{w}_P^*)$

Como visto nos Itens 2.2, 2.6 e 2.7, o espalhamento dos valores singulares da matriz de estados  $\mathbf{R}(\mathbf{w}^*)$  tem um efeito importante na convergência local dos algoritmos de adaptação. Para

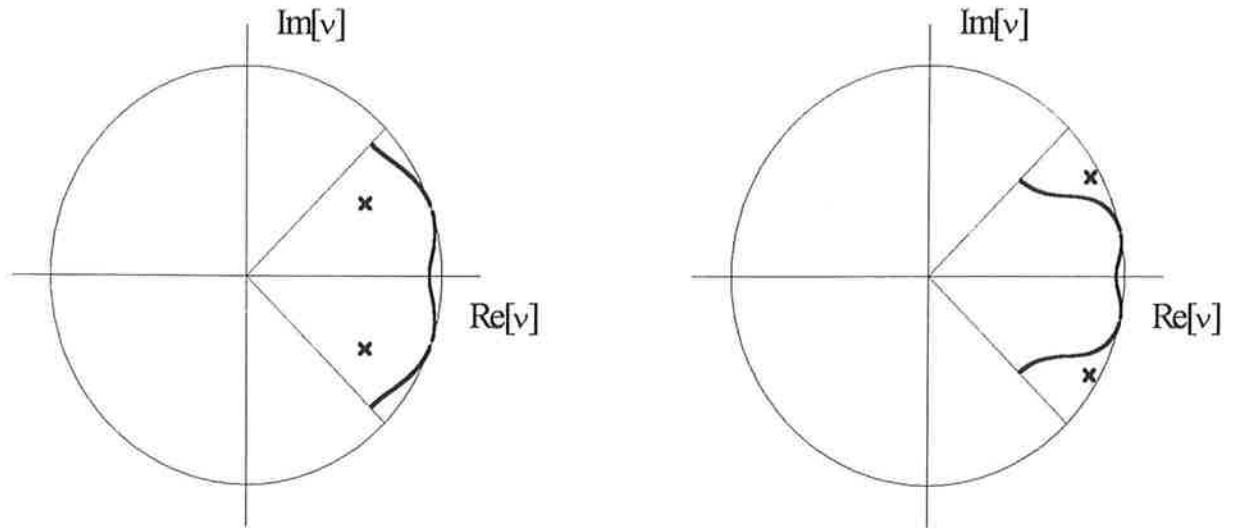


Figura 4-3: Algoritmo PLR,  $p = 4$ , sistemas de quarta ordem (um par de polos de  $H(z^{-1})$  indicado com  $\times$ ): região do outro par de polos na qual a condição de positividade é obedecida

o caso polifásico pode-se estender a análise feita no Item 2.6 definindo, inicialmente,

$$\mathbf{e}_P = [r_0 \ r_1 \ \cdots \ r_{M-1} \ q_0 \ q_1 \ \cdots \ q_{pM}]^T, \quad (4.73)$$

$$R_P(z^p) = r_0 + r_1 z^p + \cdots + r_{M-1} z^{p(M-1)}, \quad (4.74)$$

$$\begin{aligned} Q_P(z) &= q_0 + q_1 z + \cdots + q_{pM} z^{pM} \\ &= \sum_{l=0}^{p-1} z^l Q_l(z^p), \end{aligned} \quad (4.75)$$

onde

$$\deg[Q_0(z)] \leq M, \quad \deg[Q_l(z)] < M, \quad l = 1, 2, \dots, p-1. \quad (4.76)$$

Podemos então definir  $F_P(z)$  como

$$\begin{aligned} F_P(z) &= Q_P(z) - z^p R_P(z^p) H(z) \\ &= \sum_{l=0}^{p-1} z^l Q_l(z^p) - z^p R_P(z^p) \sum_{l=0}^{p-1} z^l H_l(z^p) \\ &= \sum_{l=0}^{p-1} z^l [Q_l(z^p) - z^p R_P(z^p) H_l(z^p)] \end{aligned}$$

$$\begin{aligned}
&= \sum_{l=0}^{p-1} z^l [Q_l(z^p) - z^p R_P(z^p) H_l(z^p)] \\
&= \sum_{l=0}^{p-1} z^l F_l(z^p), \tag{4.77}
\end{aligned}$$

onde

$$F_l(z) = Q_l(z) - z R_P(z) H_l(z). \tag{4.78}$$

Além das definições acima, a seguinte propriedade será utilizada:

**Propriedade 4.4** *Sejam duas funções  $f(z)$  e  $g(z)$  e respectivas expansões polifásicas  $\sum_{l=0}^{p-1} z^l f_l(z^p)$  e  $\sum_{l=0}^{p-1} z^l g_l(z^p)$ . A seguinte igualdade é válida*

$$\langle f(z), g(z) \rangle = \sum_{l=0}^{p-1} \langle f_l(z), g_l(z) \rangle. \tag{4.79}$$

*Prova:* Passando o produto interno para o domínio do tempo e usando (4.2) temos:

$$\begin{aligned}
\langle f(z), g(z) \rangle &= \sum_{k=-\infty}^{\infty} f_k g_k \\
&= \sum_{l=0}^{p-1} \sum_{k=-\infty}^{\infty} f_{pk+l} g_{pk+l} \\
&= \sum_{l=0}^{p-1} \langle f_l(z), g_l(z) \rangle. \quad \blacksquare \tag{4.80}
\end{aligned}$$

Podemos agora obter expressões convenientes para as formas quadráticas usadas na análise dos valores singulares extremos de  $\mathbf{R}_P(\mathbf{w}_P^*)$ , onde  $\mathbf{w}_P^*$  são os parâmetros de uma expansão polifásica de  $H(z)$ . De (4-23), (4-25), (4.64), usando a Propriedade 4.4 e considerando, com base nos mesmos motivos discutidos no Item 2.6,  $u(n)$  branco de potência unitária, temos, para os algoritmos RG e SMM

$$\mathbf{e}_P^\top \mathbf{R}_P(\mathbf{w}_P^*) \mathbf{e}_P = \left\langle \frac{F_P(z)}{D_P(z^p)}, \frac{F_P(z)}{D_P(z^p)} \right\rangle$$

$$= \sum_{l=0}^{p-1} \left\langle \frac{F_l(z)}{D_P(z)}, \frac{F_l(z)}{D_P(z)} \right\rangle, \quad (4.81)$$

enquanto que para o algoritmo SHARF,

$$\mathbf{e}_P^T \mathbf{R}_P(\mathbf{w}_P^*) \mathbf{e}_P = \langle F_P(z), F_P(z) \rangle \quad (4.82)$$

$$(4.83)$$

$$= \sum_{l=0}^{p-1} \langle F_l(z), F_l(z) \rangle. \quad (4.84)$$

Para o algoritmo PLR temos, de maneira análoga ao caso da estrutura direta (ver obtenção de (2.148)),

$$\mathbf{e}_P^T \mathbf{R}_P^T(\mathbf{w}_P^*) \mathbf{R}_P(\mathbf{w}_P^*) \mathbf{e}_P = \sum_{i=1}^M \left\langle \frac{F_P(z)}{D_P(z^p)}, z^{pi} H(z) \right\rangle^2 + \sum_{i=0}^{pM} \left\langle \frac{F_P(z)}{D_P(z^p)}, z^i \right\rangle^2. \quad (4.85)$$

Pela Propriedade 1.1, os valores singulares extremos de  $\mathbf{R}_P(\mathbf{w}_P^*)$  são dados pelos valores máximo e mínimo das formas quadráticas em (4.81), (4.82) e (4.85), com a restrição  $\|\mathbf{e}_P\| = 1$ . Notar que no caso polifásico a restrição  $\|\mathbf{e}_P\| = 1$  significa, usando a Propriedade 4.4, que

$$\|Q_P(z)\|^2 + \|R_P(z)\|^2 = \sum_{l=0}^{p-1} \|Q_l(z)\|^2 + \|R_P(z)\|^2 = 1. \quad (4.86)$$

Nos itens a seguir é analisado o comportamento dos valores singulares extremos de  $\mathbf{R}_P(\mathbf{w}_P^*)$  quando o sistema sendo modelado  $H(z)$  tem polos próximos à circunferência unitária. São considerados os mesmos algoritmos e restrições à função  $H(z)$  que na análise feita no Item 2.6 para a estrutura direta. Na análise apresentada a seguir, o aspecto de interesse investigado é se a estrutura polifásica, em cada caso considerado, resulta, em relação à estrutura direta, em uma matriz de estados com menor valor singular máximo e/ou um maior valor singular mínimo.

Antes de proceder a esta análise, porém, é necessário considerar o efeito da possível não unicidade da representação polifásica de  $H(z)$  no condicionamento de  $\mathbf{R}_P(\mathbf{w}_P^*)$ .

## Efeito da não unicidade da representação polifásica

Como visto no Item 4.1, considerando  $\deg[H(z)] = M$ , se  $\nu_i$  e  $\nu_j$ , com  $\nu_i \neq \nu_j$ , são polos de  $H(z)$  e valer  $\nu_i^p = \nu_j^p$ , então na expansão polifásica de ordem  $p$  de  $H(z)$  temos que  $\deg[H_l(z)] < M$ ,  $l = 0, 1, \dots, p-1$ . Deste fato e de (4.76) podemos escrever

$$H_0(z) = \frac{C_0(z)}{D_P(z)} = \frac{C'_0(z)}{D'_P(z)}, \quad \deg[C'_0(z)] < M \quad (4.87)$$

e

$$H_l(z) = \frac{C_l(z)}{D_P(z)} = \frac{C'_l(z)}{D'_P(z)}, \quad \deg[C'_l(z)] < M - 1. \quad (4.88)$$

Das expressões acima e de (4.76) resulta que sempre podemos escolher  $Q_l(z)$  em (4.78) de modo que

$$Q_l(z) = zC_l(z), \quad l = 0, 1, \dots, p-1. \quad (4.89)$$

Se fizermos então

$$R_P(z) = \kappa D_P(z), \quad Q_l(z) = z\kappa C_l(z) \quad (4.90)$$

onde  $\kappa$  é tal que  $\|\mathbf{e}_P\| = 1$  (ver (4.86)), teremos

$$F_l(z) \equiv 0, \quad l = 0, 1, \dots, p-1. \quad (4.91)$$

De (4.78), (4.81), (4.82), (4.85), (2.141) e (2.143) resulta portanto que, nesta situação, para todos os algoritmos considerados, o valor singular mínimo de  $\mathbf{R}_P(\mathbf{w}_P^*)$  é zero. Em termos práticos, se  $\nu_i^p \simeq \nu_j^p$  resulta que o valor singular mínimo de  $\mathbf{R}_P(\mathbf{w}_P^*)$  é próximo de zero. Com base na discussão do Item 4.2.1, a análise a seguir restringe-se, quando não indicado o contrário, aos casos em que a situação acima não ocorre, e nos quais vale, portanto,  $\deg[H_l(z)] = M$ ,  $l = 0, 1, \dots, p-1$ .

### 4.3.4 Valores singulares extremos de $\mathbf{R}_P(\mathbf{w}_P^*)$ para os algoritmos RG e SMM

**Observação:** Para não sobrecarregar a notação com índices adicionais será usado o símbolo  $\rho$  para indicar os valores singulares no caso polifásico. Assim, por exemplo, o valor singular máximo no caso do algoritmo RG é denotado por  $\rho_{\max, RG}$ .

Fazendo  $R_P(z) \equiv 1$  e  $Q_P(z) \equiv 0$  em (4.77), temos  $F_P(z) = -H(z)$  e de (4.81) resulta para os algoritmos RG e SMM

$$\rho_{\max, RG} \geq \left\langle \frac{H(z)}{D_P(z^p)}, \frac{H(z)}{D_P(z^p)} \right\rangle \triangleq \rho_1. \quad (4.92)$$

O limite inferior dado por  $\rho_1$ , pelos mesmos motivos que no caso da estrutura direta (ver Item 2.6.4), é uma boa aproximação da ordem de grandeza de  $\rho_{\max, RG}$ , e, portanto, pode-se comparar os valores singulares máximos das matrizes de estados para as estruturas direta e polifásica por meio da comparação destes limites inferiores.

### Valor singular máximo de $R_P(\mathbf{w}_P^*)$ para as Restrições 1 e 2

Considerando as restrições a  $H(z)$  definidas no Item 2.6, para manter uma notação análoga à utilizada para a estrutura direta, o limite inferior para  $\rho_{\max, RG}$  definido como  $\rho_1$  seria dado pelo produto interno em (4.92) para o caso da Restrição 1, e o limite definido com  $\rho_4$  seria dado pela mesma expressão, mas para o caso da Restrição 2. Como na análise a seguir não há distinção entre estas duas restrições, para maior simplicidade faz-se referência apenas ao limite  $\rho_1$ .

Comparando as expressões (4.92) e (2.152), podemos ver que a diferença entre as duas está na troca do fator  $D(z)$  pelo fator  $D_P(z^p)$ . O efeito desta troca é melhor visualizado no domínio da frequência, como segue. Notando que para quaisquer  $F(z)$  e  $G(z)$  temos  $\langle F(z), G(z) \rangle = \langle F(z^{-1}), G(z^{-1}) \rangle$ , podemos escrever inicialmente

$$\frac{1}{D(z^{-1})} = \frac{z^M}{(z - r_1 e^{j\theta_1}) \dots (z - r_M e^{j\theta_M})}, \quad (4.93)$$

e, lembrando que as raízes de  $D_P(z^{-p})$  são dadas por  $|r_i| \angle(\theta_i + 2\pi l/p)$ ,  $l = 0, 1, \dots, p-1$ ,  $i = 1, 2, \dots, M$  (ver Item 4.1),

$$\frac{1}{D_P(z^{-p})} = \frac{z^{pM}}{(z - r_1 e^{j\theta_1}) \dots (z - r_1 e^{j(\theta_1 + 2\pi \frac{p-1}{p})}) \dots (z - r_M e^{j\theta_M}) \dots (z - r_M e^{j(\theta_M + 2\pi \frac{p-1}{p})})}. \quad (4.94)$$

Escrevendo agora o produto interno em (4.92) como

$$\left\langle \frac{H(z)}{D_P(z^p)}, \frac{H(z)}{D_P(z^p)} \right\rangle = \frac{1}{2\pi} \int_0^{2\pi} \left| \frac{H(e^{j\omega})}{D_P(e^{j\omega p})} \right|^2 d\omega, \quad (4.95)$$



e analogamente para (2.152), vemos que, para o caso das Restrições 1 e 2, se  $H(z^{-1})$  tem polos em  $r_j e^{j\theta_j}$ ,  $j = 1, 2, \dots, m$ , próximos à circunferência unitária, isto é, com  $r_j = r \simeq 1$ , então, devido à seletividade de  $H(z)$ , os produtos internos em (4.92) e (2.152) dependem essencialmente dos valores absolutos em  $e^{j\theta_j}$ ,  $j = 1, 2, \dots, m$ , das funções envolvidas, isto é,

$$\left\langle \frac{H(z)}{D_P(z^p)}, \frac{H(z)}{D_P(z^p)} \right\rangle \simeq \sum_{j=1}^m \frac{1}{2\pi} \int_{\theta_j - \varepsilon}^{\theta_j + \varepsilon} \left| \frac{H(e^{j\omega})}{D_P(e^{j\omega p})} \right|^2 d\omega \quad (4.96)$$

e

$$\left\langle \frac{H(z)}{D(z)}, \frac{H(z)}{D(z)} \right\rangle \simeq \sum_{j=1}^m \frac{1}{2\pi} \int_{\theta_j - \varepsilon}^{\theta_j + \varepsilon} \left| \frac{H(e^{j\omega})}{D(e^{j\omega})} \right|^2 d\omega, \quad (4.97)$$

onde  $\varepsilon > 0$  tem um valor pequeno. Nestes pontos, como pode ser visto em (4.93) e (4.94),  $D(z^{-1})$  e  $D_P(z^{-p})$  tem em comum os termos  $(e^{j\theta_j} - r_i e^{j\theta_i})$ ,  $i = 1, 2, \dots, M$ , porém  $D_P(z^{-p})$  tem ainda todos os termos adicionais  $(e^{j\theta_j} - r_i e^{j(\theta_i + 2\pi \frac{l}{p})})$ ,  $i = 1, 2, \dots, M$ ,  $l = 1, 2, \dots, p - 1$ , resultantes da expansão polifásica. Estes termos adicionais, como visto a seguir, podem fazer com que, em  $e^{j\theta_j}$ ,  $j = 1, 2, \dots, m$ , os valores absolutos de  $D_P(z^{-p})$  seja maiores do que os de  $D(z^{-1})$  e, conseqüentemente,  $\rho_1 < \sigma_1$ . Como um exemplo disto, consideremos a função  $H_5(z)$  definida no Item 2.6.8: na Figura 4-4 temos, para três valores de  $p$ , os valores absolutos de  $H(e^{j\omega})$ ,  $1/D(e^{j\omega})$  e  $1/D_P(e^{j\omega p})$ .

O valor exato da relação  $\sigma_1/\rho_1$  depende do número de polos de  $H(z)$ , dos seus ângulos e do fator de expansão polifásica, sendo que, em geral, não é possível obter uma expressão simples para  $\sigma_1/\rho_1$ . Porém, no caso limite em que  $H(z)$  tem  $M$  polos em 1, ou seja, nas expressões acima,  $r_i = 1$  e  $\theta_i = 0$ ,  $i = 1, 2, \dots, M$ , a relação  $\sigma_1/\rho_1$  tem uma expressão simples, como visto a seguir.

De início, temos que a seguinte identidade é válida

$$\prod_{l=1}^{p-1} (1 - e^{j2\pi l/p}) = p. \quad (4.98)$$

Podemos ver que a produtória acima corresponde, no caso específico sendo considerado, ao produto dos valores absolutos, em  $e^{j\theta_j}$ ,  $j = 1, 2, \dots, M$ , dos termos adicionais de  $D_P(z^{-p})$  em relação a  $D(z^{-1})$ , referentes a um polo de  $H(z)$ . Como  $H(z)$  tem  $M$  polos, e, ainda, a

$p$	$H_a(z), \sigma_{\max, RG} = 5,0 \times 10^5$			$H_b(z), \sigma_{\max, RG} = 1,1 \times 10^{10}$			$H_c(z), \sigma_{\max, RG} = 8,4 \times 10^4$		
	$\rho_{\max, RG}$	$\frac{\sigma_{\max, RG}}{\rho_{\max, RG}}$	$p^{2M}$	$\rho_{\max, RG}$	$\frac{\sigma_{\max, RG}}{\rho_{\max, RG}}$	$p^{2M}$	$\rho_{\max, RG}$	$\frac{\sigma_{\max, RG}}{\rho_{\max, RG}}$	$p^{2M}$
2	$1,3 \times 10^5$	3,994	4	$6,8 \times 10^8$	15,82	16	$5,4 \times 10^3$	15,54	16
3	$5,6 \times 10^4$	8,973	9	$1,4 \times 10^8$	79,17	81	$1,1 \times 10^3$	75,92	81
4	$3,1 \times 10^4$	15,93	16	$4,4 \times 10^7$	247,4	256	$3,7 \times 10^2$	229,7	256
5	$2,0 \times 10^4$	24,85	25	$1,8 \times 10^7$	597,2	625	$1,6 \times 10^2$	532,5	625
6	$1,4 \times 10^4$	35,73	36	$8,8 \times 10^6$	1224	1296	$8,1 \times 10^1$	1038	1296
7	$1,0 \times 10^4$	48,56	49	$4,8 \times 10^6$	2243	2401	$4,7 \times 10^1$	1798	2401
8	$7,9 \times 10^3$	63,33	64	$2,9 \times 10^6$	3783	4096	$3,0 \times 10^1$	2836	4096

Tabela 4.5: Valor singular máximo da matriz de estados para as estruturas direta e polifásica, algoritmo RG.

relação entre  $|D(z^{-1})|$  e  $|D_P(z^{-p})|$  aparece ao quadrado na relação  $\sigma_1/\rho_1$ , resulta que

$$\frac{\sigma_1}{\rho_1} = p^{2M}. \quad (4.99)$$

Para exemplificar como comparam-se esta relação e a relação efetivamente de interesse,  $\sigma_{\max, RG}/\rho_{\max, RG}$ , foram considerados os seguintes exemplos:  $H_a(z)$  com um polo em 0,999 e zero em 0,3;  $H_b(z)$  com polos em 0,999 e 0,99 e zeros em  $0,3\angle \pm 110^\circ$ ;  $H_c(z)$  com polos em  $0,99\angle \pm 10^\circ$  e zeros em  $0,9\angle \pm 110^\circ$ . Os resultados estão na Tabela 4.5, para  $p$  variando de 2 a 8. Podemos ver que à medida em que  $H(z)$  se afasta da situação considerada acima ( $M$  polos em 1), a relação  $\sigma_{\max, RG}/\rho_{\max, RG}$  tende a ser menor do que a dada por (4.99), sendo que a diferença aumenta com  $p$ .

Quando  $H(z)$  tem polos complexos, à medida em que  $p$  aumenta há uma aproximação entre as raízes de  $D_P(z^{-p})$ , podendo haver o aparecimento de raízes duplas, o que ocorre quando para dois polos  $\nu_i$  e  $\nu_j$  de  $H(z)$  valer  $\nu_i^p = \nu_j^p$  (isto decorre do fato de que, como visto no Item 4.1, os polos de  $D_P(z)$  são os polos de  $D(z)$  elevados à  $p$ -ésima potência), que é a mesma situação na qual  $\mathbf{R}_P(\mathbf{w}_P^*)$  torna-se singular, conforme discutido no item anterior. Nesta situação, os valores absolutos de  $D_P(z^{-p})$  nos pontos de interesse podem ser maiores do que os de  $D(z^{-1})$  e, conseqüentemente, a relação  $\sigma_{\max, RG}/\rho_{\max, RG}$  pode se tornar menor do que 1. Para a mesma função  $H_5(z)$  da Figura 4-4 isto ocorre para o caso  $p = 6$ , exemplificado na Figura 4-5. Em termos práticos, portanto, deve-se esperar que se para dois polos  $\nu_i$  e  $\nu_j$  de  $H(z)$  valer  $\nu_i^p \simeq \nu_j^p$  então  $\rho_{\max, RG}$  poderá ser até maior do que  $\sigma_{\max, RG}$ , caso em que não haveria ganho da estrutura polifásica em relação à direta quanto ao espalhamento dos valores singulares da matriz de estados.

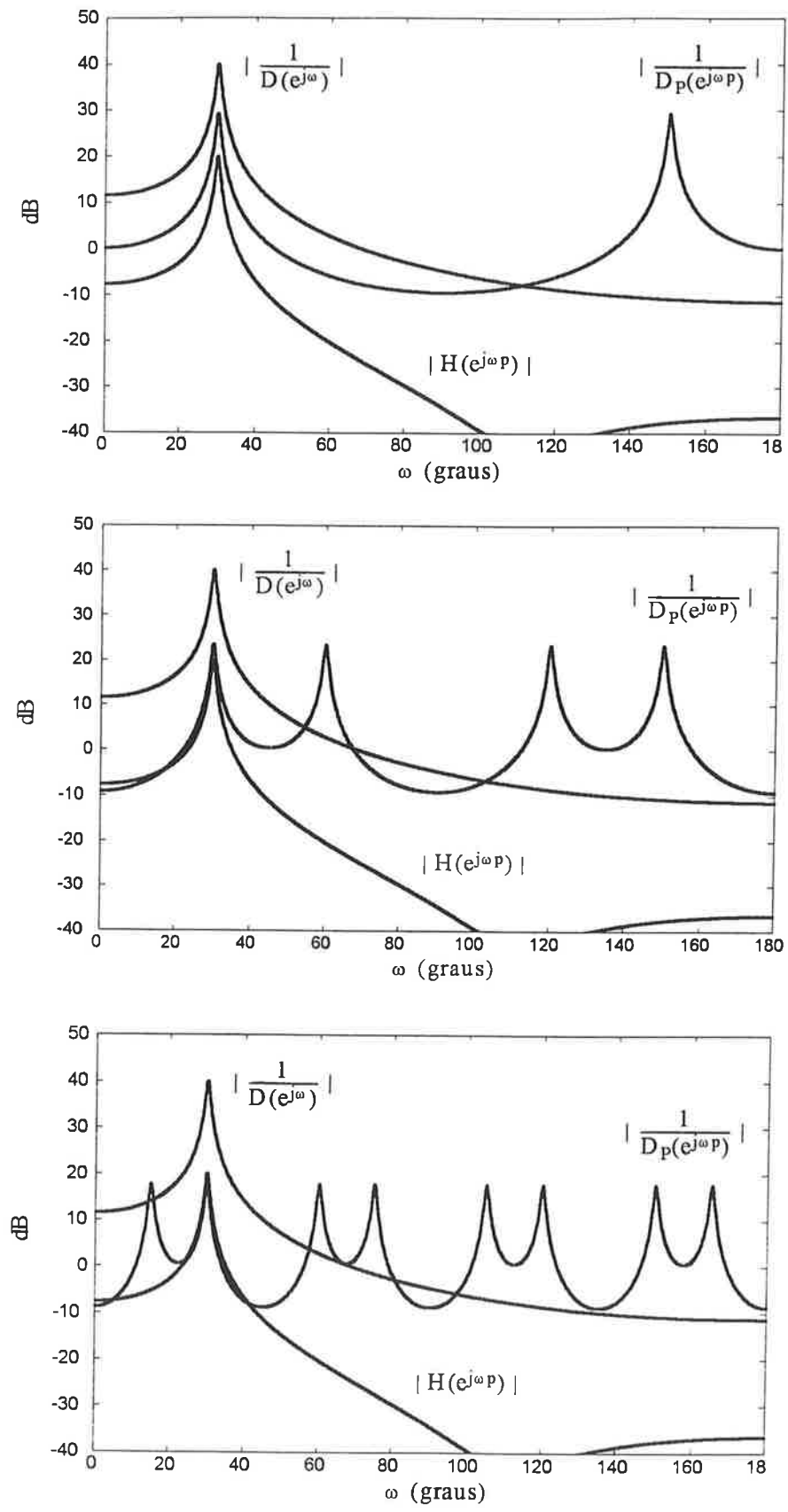


Figura 4-4: Funções que determinam  $\rho_1$  para  $H_5(z)$ , com  $p = 2$ ,  $p = 4$  e  $p = 8$ .

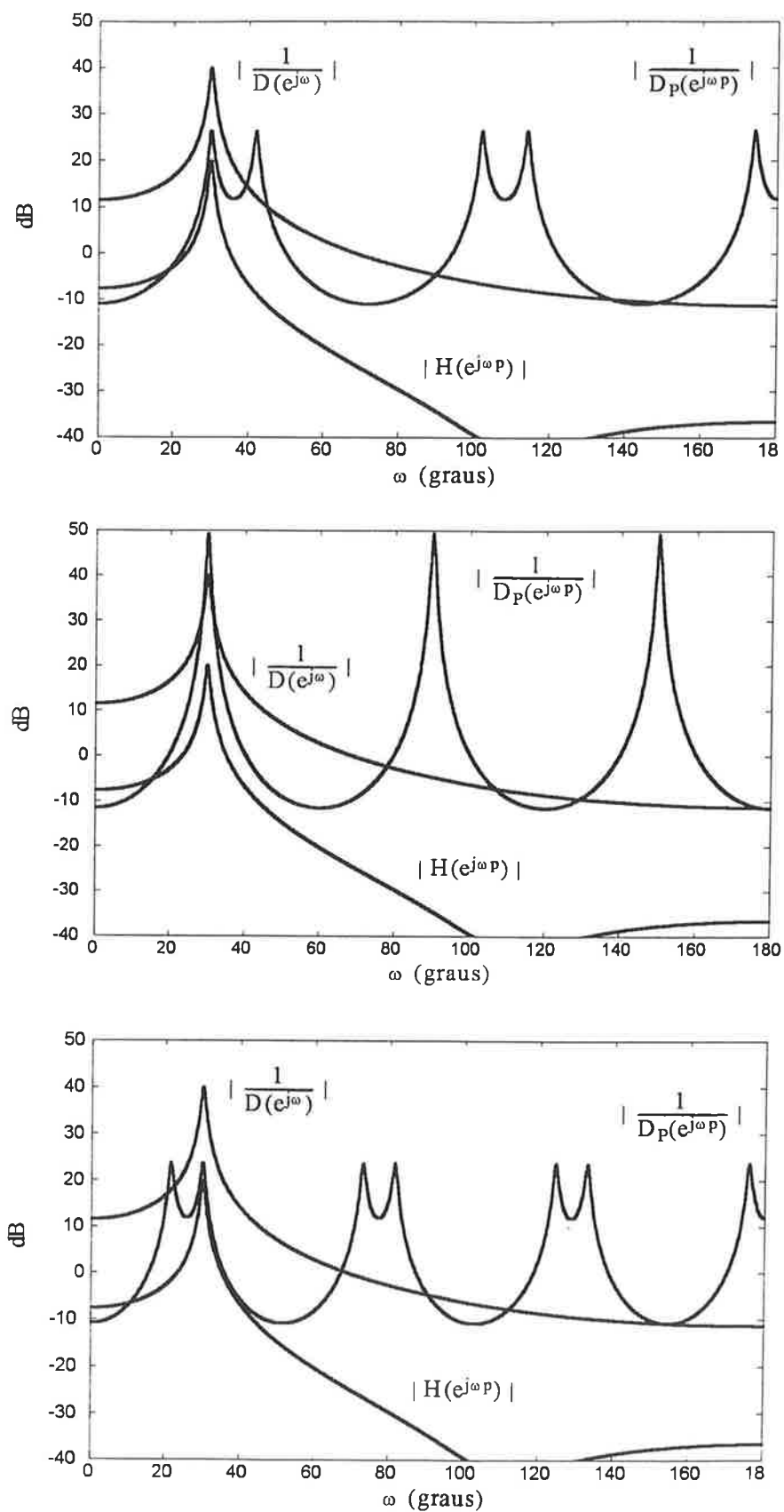


Figura 4-5: Funções que determinam  $\rho_1$  para  $H_5(z)$ , com  $p = 5$ ,  $p = 6$  e  $p = 7$ .

### Valor singular máximo de $\mathbf{R}_P(\mathbf{w}_P^*)$ para a Restrição 3

Para o caso em que  $H(z)$  é passa-tudo, isto é,  $H(z)H(z^{-1}) \equiv 1$ , de (4.92) temos

$$\rho_{\max, RG} \geq \left\langle \frac{1}{D_P(z^p)}, \frac{1}{D_P(z^p)} \right\rangle \triangleq \rho_6, \quad (4.100)$$

onde o símbolo  $\rho_6$  foi usado para manter uma notação análoga à empregada para a estrutura direta. Expressando o produto interno acima como em (4.95), vemos que, como neste caso  $H(z)$  não é seletivo em frequência, não basta considerar, como em (4.96), apenas os trechos ao redor de  $e^{j\theta_j}$  da integral em  $\omega$ , onde  $\theta_j$  são os ângulos dos polos de  $H(z)$  de próximos à circunferência unitária, mas sim os trechos em torno de todos os pontos  $e^{j(\theta_j + 2\pi \frac{l}{p})}$  relativos aos polos acrescentados pela expansão polifásica, isto é,

$$\left\langle \frac{1}{D_P(z^p)}, \frac{1}{D_P(z^p)} \right\rangle \simeq \sum_{l=0}^{p-1} \sum_{j=1}^m \frac{1}{2\pi} \int_{\theta_j + 2\pi \frac{l}{p} - \varepsilon}^{\theta_j + 2\pi \frac{l}{p} + \varepsilon} \left| \frac{1}{D_P(e^{j\omega p})} \right|^2 d\omega \quad (4.101)$$

(Para visualizar este fato pode-se tomar os gráficos da Figura 4-4 e assumir  $|H(e^{j\omega})| = 1$ ). Conseqüentemente, quando o raio de algum polo de  $H(z)$  tende a 1 a relação  $\sigma_6/\rho_6$  é menor do que a relação  $\sigma_1/\rho_1$ . Novamente assumindo que estes limites fornecem uma boa aproximação da ordem de grandeza dos valores singulares máximos das matrizes de estados  $\mathbf{R}(\mathbf{w}^*)$  e  $\mathbf{R}_P(\mathbf{w}_P^*)$ , resulta que no caso em que  $H(z)$  é passa-tudo a relação  $\sigma_{\max, RG}/\rho_{\max, RG}$  será menor do que no caso das outras duas restrições.

### Valor singular mínimo de $\mathbf{R}_P(\mathbf{w}_P^*)$

De (4.81) e da Propriedade 1.1 temos

$$\rho_{\min, RG} = \min_{\|\mathbf{e}\|=1} \sum_{l=0}^{p-1} \left\langle \frac{F_l(z)}{D_P(z)}, \frac{F_l(z)}{D_P(z)} \right\rangle. \quad (4.102)$$

Podemos aplicar a cada termo da somatória acima os mesmos argumentos utilizados no caso da estrutura direta. Resulta então que se  $H(z)$  tem  $m = M$  polos de raio  $r$ , quando  $r$  tende a 1  $\rho_{\min, RG}$  é obtido com  $R_P(z) \equiv 0$ ,  $Q_0(z) = D_P(z)/(\|D_P(z)\|)$  e  $Q_l(z) \equiv 0$ ,  $l = 1, 2, \dots, p-1$ . Temos, portanto,

$$\lim_{r \rightarrow 1} \rho_{\min, RG} = k_{P,1}, \quad m = M, \quad (4.103)$$

onde  $k_{P,1} = \lim_{r \rightarrow 1} 1/\|D_P(z)\|^2$ . No caso em que  $m < M$ , temos, para as restrições 1 e 3, como em (2.171),

$$\rho_{\min, RG} \leq k_{P,1}, \quad m < M. \quad (4.104)$$

Em comparação com o caso da estrutura direta, temos que quando os polos de  $H(z)$  estão próximos à circunferência unitária,  $k_{P,1}$  é da mesma ordem de grandeza que  $k_1$ . Neste caso, portanto, o possível ganho da estrutura polifásica quanto ao espalhamento dos valores singulares da matriz de estados resulta, essencialmente, da possível redução dos valores singulares máximos, vista acima. Para caracterizar de maneira mais precisa este ganho seria necessário analisar em maior detalhe a relação  $k_{P,1}/k_1$ . Nos casos práticos investigados, porém, isto não se mostrou necessário, e esta análise não será feita aqui.

Para a Restrição 2, também como no caso da estrutura direta,  $\rho_{\min, RG}$  tende a zero quando os polos de  $H(z)$  se aproximam da circunferência unitária. Quanto a este aspecto, então, não há ganho da estrutura polifásica, uma vez que o que determina que os valores singulares mínimos de  $\mathbf{R}(\mathbf{w})$  tendam a zero neste caso, como visto no Item 2.6.5, é o termo  $\langle C'(z)/D(z), C'(z)/D(z) \rangle$  em (2.174), cujo valor não muda ao se passar para a estrutura polifásica.

### 4.3.5 Valores singulares extremos de $\mathbf{R}_P(\mathbf{w}_P^*)$ para o algoritmo PLR

#### Valor singular máximo de $\mathbf{R}_P(\mathbf{w}_P^*)$ para as Restrições 1 e 2

Fazendo  $R_P(z) \equiv 1$  e  $Q_P(z) \equiv 0$  em (4.85), temos, analogamente a (2.154),

$$\rho_{\max, PLR} \geq \left| \left\langle \frac{C(z)}{D_P(z^P)D(z)}, \frac{C(z)}{D(z)} \right\rangle \right| \triangleq \rho_2. \quad (4.105)$$

Assim como no caso da estrutura direta, quando o raio  $r$  de algum polo de  $H(z)$  tende a 1, temos

$$\left| \left\langle \frac{C(z)}{D_P(z^P)D(z)}, \frac{C(z)}{D(z)} \right\rangle \right| \xrightarrow{r \rightarrow 1} |\gamma_P| \left\langle \frac{C(z)}{D(z)}, \frac{C(z)}{D(z)} \right\rangle, \quad |\gamma_P| \xrightarrow{r \rightarrow 1} \infty, \quad (4.106)$$

onde  $\gamma_P$  resulta da aplicação de (A.8) em  $1/D_P(z^P)$ . No mesmo caso limite considerado para o algoritmo RG, em que  $H(z)$  tem  $M$  polos em 1, os termos  $1/D_P(1/p_i^P)$  que aparecem em (A.8) são reais. Temos então  $\text{Re}[1/D_P(1/p_i^P)] = |1/D_P(1/p_i^P)|$  e, notando que  $1/p_i$  está próximo à circunferência unitária, podemos aplicar a mesma argumentação que no caso RG

$p$	$H_a(z), \sigma_{\max,PLR} = 5,0 \times 10^2$			$H_b(z), \sigma_{\max,PLR} = 1,0 \times 10^5$			$H_c(z), \sigma_{\max,PLR} = 1,1 \times 10^2$		
	$\rho_{\max,PLR}$	$\frac{\sigma_{\max,PLR}}{\rho_{\max,PLR}}$	$p^M$	$\rho_{\max,PLR}$	$\frac{\sigma_{\max,PLR}}{\rho_{\max,PLR}}$	$p^M$	$\rho_{\max,PLR}$	$\frac{\sigma_{\max,PLR}}{\rho_{\max,PLR}}$	$p^M$
2	$2,5 \times 10^2$	1,996	2	$2,5 \times 10^4$	3,971	4	$4,1 \times 10^1$	2,705	4
3	$1,7 \times 10^2$	2,988	3	$1,1 \times 10^4$	8,869	9	$2,4 \times 10^1$	4,755	9
4	$1,3 \times 10^2$	3,976	4	$6,4 \times 10^3$	15,65	16	$1,6 \times 10^1$	7,165	16
5	$1,0 \times 10^2$	4,961	5	$4,1 \times 10^3$	24,28	25	$1,1 \times 10^1$	10,03	25
6	$8,4 \times 10^1$	5,941	6	$2,9 \times 10^3$	34,71	36	8,3	13,51	36
7	$7,2 \times 10^1$	6,918	7	$2,1 \times 10^3$	46,91	49	6,3	17,82	49
8	$6,3 \times 10^1$	7,891	8	$1,6 \times 10^3$	60,82	64	4,8	23,19	64

Tabela 4.6: Valor singular máximo da matriz de estados para as estruturas direta e polifásica, algoritmo PLR.

para obter

$$\frac{\sigma_2}{\rho_2} = p^M, \quad (4.107)$$

onde agora temos  $p^M$  ao invés de  $p^{2M}$  pois a relação  $D(1/p_i)/D_P(1/p_i^p)$  não aparece ao quadrado em  $\sigma_2/\rho_2$ . Para exemplificar esta propriedade foram considerados os mesmos exemplos que para o caso RG, os resultados sendo apresentados na Tabela 4.6.

### Valor singular máximo de $R_P(w_p^*)$ para Restrição 3

Para o algoritmo PLR, analogamente ao que foi feito para o caso da estrutura direta, podemos considerar diretamente a expressão (4.85). Nesta expressão a segunda somatória tem sempre um valor finito. Para os termos da primeira somatória, usando o pressuposto  $H(z)H(z^{-1}) \equiv 1$  e, como dado por (4.77),  $F_P(z) = [Q_P(z) - z^p R_P(z^p)H(z)]$ , temos:

$$\begin{aligned} \left\langle \frac{F_P(z)}{D_P(z^p)}, z^{pi} H(z) \right\rangle &= \left\langle \frac{Q_P(z) - z^p R_P(z^p)H(z)}{D_P(z^p)}, z^{pi} H(z) \right\rangle \\ &= \left\langle \frac{Q_P(z)}{D_P(z^p)}, z^{pi} H(z) \right\rangle - \left\langle z^p \frac{R_P(z^p)}{D_P(z^p)}, z^{pi} \right\rangle. \end{aligned} \quad (4.108)$$

O segundo termo desta expressão é sempre finito, sendo que para o primeiro termo, usando o fato de que  $H(z) = 1/H(z^{-1})$ , temos

$$\begin{aligned} \left\langle \frac{Q_P(z)}{D_P(z^p)}, z^{pi} H(z) \right\rangle &= \left\langle \frac{Q_P(z)}{D_P(z^p)}, z^{pi} \frac{D_P(z^{-p})}{C_P(z^{-1})} \right\rangle \\ &= \left\langle \frac{Q_P(z)}{C_P(z)}, z^{pi} \right\rangle. \end{aligned} \quad (4.109)$$

O polinômio  $C_P(z)$  pode ser escrito como  $C_P(z) = C(z)C'_P(z)$ , onde  $C(z)$  é de fase não-mínima, pois  $H(z)$  é passa-tudo, e as  $(p-1)M$  raízes de  $C'_P(z)$  são as raízes de  $D_P(z^p)$  que devem ser canceladas na expansão polifásica. Portanto, o produto interno acima pode ser escrito como o produto interno de duas funções estáveis e causais:

$$\begin{aligned} \left\langle \frac{Q_P(z)}{C_P(z)}, z^{pi} \right\rangle &= \left\langle \frac{Q_P(z)}{C(z)C'_P(z)}, z^{pi} \right\rangle \\ &= \left\langle \frac{Q_P(z)}{C'_P(z)}, \frac{z^{pi}}{C(z^{-1})} \right\rangle \\ &= \left\langle \frac{Q_P(z)}{C'_P(z)}, \frac{z^{pi+M}}{\overline{C}(z)} \right\rangle, \end{aligned} \quad (4.110)$$

com  $\overline{C}(z) = z^M C(z^{-1})$ . Como  $H(z)$  é passa-tudo, se os polos de  $H(z)$  se aproximam da circunferência unitária o mesmo ocorre com as raízes de  $C'_P(z)$  e  $\overline{C}(z)$ . Em relação a estas raízes temos que, se for assumido, como anteriormente, que para nenhum par de polos  $\nu_i$  e  $\nu_j$  de  $H(z)$ ,  $\nu_i \neq \nu_j$ , vale  $\nu_i^p = \nu_j^p$ , então as raízes de  $C'_P(z)$  são sempre diferentes das raízes de  $\overline{C}(z)$ . Neste caso, então, assim como no caso da estrutura direta, para o algoritmo PLR e a Restrição 3, o produto interno acima, e, conseqüentemente, o valor singular máximo de  $\mathbf{R}(\mathbf{w}_P^*)$ , não tendem a infinito quando algum polo de  $H(z)$  se aproxima da circunferência unitária.

### Valor singular mínimo de $\mathbf{R}_P(\mathbf{w}_P^*)$

Para o caso das Restrições 1 e 3, com  $F_P(z) = [Q_P(z) - z^p R_P(z^p)H(z)]$  temos que fazendo  $R_P(z^p) \equiv 0$  e  $Q_P(z) = D_P(z^p)/\|D_P(z)\|$  em (4.85) obtém-se

$$\rho_{\min,PLR} \leq \sqrt{k_{P,1}} \left[ \sum_{i=1}^M \langle 1, z^{pi} H(z) \rangle^2 + \sum_{i=0}^{pM} \langle 1, z^i \rangle^2 \right] = \sqrt{k_{P,1}}, \quad (4.111)$$

onde como já definido no caso RG,  $k_{P,1} = \lim_{r \rightarrow 1} 1/\|D_P(z)\|^2$ . Para a Restrição 3, porém, não é possível aplicar o mesmo argumento que no caso da estrutura direta, o qual levava à conclusão que mesmo quando algum polo de  $H(z)$  se aproximava da circunferência unitária,  $\rho_{\min,PLR}$  não tendia a zero.

Para o caso da Restrição 2 e quando  $H(z)$  tem  $m < M$  polos que se aproximam da circunferência unitária, pode-se aplicar o mesmo argumento que para o caso da estrutura



direta, chegando-se à conclusão que  $\rho_{\min,PLR}$  tende a zero.

### 4.3.6 Valores singulares extremos de $\mathbf{R}_P(\mathbf{w}_P^*)$ para o algoritmo SHARF

Para a Restrição 1, o limite inferior para  $\sigma_{\max,SHARF}$  dado por (2.153) não muda quando se passa para a estrutura polifásica, e, portanto, novamente assumindo que estes limites inferiores também sejam boas aproximações para os valores singulares máximos em questão, pode-se esperar que  $\rho_{\max,SHARF}$  seja próximo de  $\sigma_{\max,SHARF}$ . Quanto ao valor singular mínimo de  $\mathbf{R}_P(\mathbf{w}_P^*)$ , por uma argumentação análoga à usada acima para o caso RG, chega-se às mesmas propriedades que no caso da estrutura direta. O mesmo vale para as propriedades dos valores singulares extremos de  $\mathbf{R}_P(\mathbf{w}_P^*)$  para as outras restrições.

Temos então, que quanto ao aspecto do espalhamento dos valores singulares da matriz de estados, não há para o algoritmo SHARF situações nas quais exista um ganho importante da estrutura polifásica em relação à direta. Deve-se notar, por outro lado, que está sendo considerado o algoritmo SHARF idealizado, isto é, com  $P(z) = D(z)$ . Em situações práticas em que  $P(z)$  estivesse consideravelmente afastado de  $D(z)$  as propriedades da convergência local seriam diferentes das do caso idealizado e a utilização do filtro polifásico poderia eventualmente ter então algum atrativo. Esta possibilidade, porém, não será investigada neste trabalho.

### 4.3.7 Exemplos numéricos do comportamento dos valores singulares

Para exemplificar o comportamento dos valores singulares de  $\mathbf{R}_P(\mathbf{w}_P^*)$  quando  $H(z)$  tem polos próximos à circunferência unitária, foram consideradas as mesmas funções  $H(z)$  que no Item 2.6.8 e um fator  $p = 4$ . Nestas condições a constante  $k_{P,1}$  assume os valores

$$k_{P,1} = \lim_{r \rightarrow 1} 1 / \|D(z)\|^2 = 0,3333, \quad H_1(z) \text{ a } H_8(z), \quad (4.112)$$

e

$$k_1 = \lim_{r \rightarrow 1} 1 / \|D(z)\|^2 = 0,4960, \quad H_9(z) \text{ a } H_{16}(z). \quad (4.113)$$

Os valores singulares extremos para cada caso estão apresentados nas Tabelas 4.7 e 4.8.

Em relação aos casos nos quais, na análise teórica, assim como para o caso da estrutura direta, foi obtido apenas um limite superior ou inferior para o valor singular mínimo, temos: 1) para casos nos quais  $m = M$  (todos os polos de  $H(z)$  próximos à circunferência unitária), que são os casos da Tabela 4.7, o valor de  $\rho_{\min}$  quando os polos de  $H(z)$  se aproximam da circunferência unitária, assim como no caso da estrutura direta, é relativamente grande, com exceção do caso da Restrição 3 com algoritmo PLR; 2) para os casos nos quais  $m < M$ , que são os casos da Tabela 4.8, há, assim como no caso da estrutura direta, duas situações distintas: para a Restrição 1 (funções  $H_9(z)$  a  $H_{11}(z)$ ) os valores de  $\rho_{\min}$  estão consideravelmente abaixo dos limites superiores obtidos, ao passo que para a Restrição 3 (funções  $H_{14}(z)$  a  $H_{16}(z)$ ), no caso do algoritmo RG,  $\sigma_{\min}$  é relativamente próximo do limite superior, e, no caso do algoritmo PLR,  $\sigma_{\min}$  é relativamente grande.

Uma vez que, como visto nos itens anteriores, o possível ganho sistemático da estrutura polifásica em relação à estrutura direta é sempre advindo de um menor valor singular máximo da matriz de estados, para facilitar a interpretação dos resultados, a relação  $\sigma_{\max}/\rho_{\max}$  já está computada na Tabela 4.9 para os casos relevantes. Pode-se ver nesta tabela que para as Restrições 1 e 2, com  $m = M$  (funções  $H_1(z)$  a  $H_5(z)$ ), quando os polos de  $H(z)$  se aproximam da circunferência unitária a relação  $\sigma_{\max}/\rho_{\max}$  tende a um valor de aproximadamente 58, enquanto que para a situação limite considerada acima, na qual os  $M$  polos de  $H(z)$  estão em 1, este valor seria de  $p^{2M} = 4^4 = 256$ . Para o caso do algoritmo PLR esta relação tende a um valor de aproximadamente 5,5 ao passo que na mesma situação limite este valor seria de  $p^M = 4^2 = 16$ . Como observado acima, para o algoritmo RG, esta relação é menor no caso da Restrição 3 (funções  $H_9(z)$  a  $H_{11}(z)$ ), tendendo a um valor de aproximadamente 11.

O espalhamento dos valores singulares de  $\mathbf{R}_P(\mathbf{w}_P^*)$ , isto é, a relação  $\rho_{\max}/\rho_{\min}$ , está computada para cada caso nas Tabelas 4.10 e 4.11, assim como a variação deste espalhamento relativa ao caso da estrutura direta, dada pela relação  $(\rho_{\max}/\rho_{\min})/(\sigma_{\max}/\sigma_{\min})$ .

## 4.4 Análise dos pontos estacionários da superfície de erro (entrada branca)

Neste item é analisada a existência de pontos estacionários da superfície de erro do filtro adaptativo polifásico. A necessidade de fazer-se esta análise pode ser entendida considerando

Função	Algoritmo						
	RG/SMM		PLR			SHARF	
	$\rho_{\max}$	$\rho_{\min}$	$\rho_{\max}$	$\rho_{\min}$	$\text{Re}[\lambda_{\min}]$	$\rho_{\max}$	$\rho_{\min}$
$H_1(z)$	4,4	0,23	2,5	0,18	0,19	1,8	0,13
$H_2(z)$	$1,5 \times 10^3$	0,34	$9,9 \times 10^1$	0,45	0,84	13	0,83
$H_3(z)$	$1,4 \times 10^6$	0,33	$9,0 \times 10^3$	0,45	0,96	130	0,98
$H_4(z)$	$1,7 \times 10^2$	0,34	$1,2 \times 10^1$	0,45	0,83	1,6	0,42
$H_5(z)$	$1,6 \times 10^4$	0,33	$1,0 \times 10^2$	0,45	0,96	1,5	0,49
$H_6(z)$	4,0	0,099	2,5	$7,7 \times 10^{-2}$	0,082	1,9	$5,7 \times 10^{-2}$
$H_7(z)$	$2,4 \times 10^1$	0,29	3,1	$7,9 \times 10^{-2}$	0,11	2,0	$9,5 \times 10^{-3}$
$H_8(z)$	$2,3 \times 10^2$	0,33	3,1	$7,9 \times 10^{-2}$	0,13	2,0	$1,0 \times 10^{-3}$

Tabela 4.7: Valores singulares extremos de  $\mathbf{R}_P(\mathbf{w}_P^*)$ ,  $p = 4$ .

Função	Algoritmo						
	RG/SMM		PLR			SHARF	
	$\rho_{\max}$	$\rho_{\min}$	$\rho_{\max}$	$\rho_{\min}$	$\text{Re}[\lambda_{\min}]$	$\rho_{\max}$	$\rho_{\min}$
$H_9(z)$	$1,1 \times 10^1$	$9,0 \times 10^{-3}$	4,3	$9,4 \times 10^{-3}$	$9,4 \times 10^{-3}$	2,0	$9,8 \times 10^{-3}$
$H_{10}(z)$	$7,1 \times 10^3$	$1,4 \times 10^{-2}$	$2,8 \times 10^2$	$1,4 \times 10^{-2}$	$1,4 \times 10^{-2}$	21	$1,5 \times 10^{-2}$
$H_{11}(z)$	$6,9 \times 10^6$	$1,4 \times 10^{-2}$	$2,7 \times 10^4$	$1,5 \times 10^{-2}$	$1,5 \times 10^{-2}$	210	$1,5 \times 10^{-2}$
$H_{12}(z)$	$6,8 \times 10^2$	$1,7 \times 10^{-3}$	$2,7 \times 10^1$	$1,7 \times 10^{-3}$	$1,7 \times 10^{-3}$	2,0	$1,7 \times 10^{-3}$
$H_{13}(z)$	$6,4 \times 10^4$	$1,8 \times 10^{-4}$	$2,5 \times 10^2$	$1,8 \times 10^{-4}$	$1,8 \times 10^{-4}$	2,0	$1,8 \times 10^{-4}$
$H_{14}(z)$	5,6	0,20	2,5	0,11	0,12	1,9	$6,0 \times 10^{-2}$
$H_{15}(z)$	$5,4 \times 10^1$	0,30	3,1	0,13	0,15	2,0	$9,4 \times 10^{-3}$
$H_{16}(z)$	$5,4 \times 10^2$	0,29	3,2	0,14	0,15	2,0	$9,8 \times 10^{-4}$

Tabela 4.8: Valores singulares extremos de  $\mathbf{R}_P(\mathbf{w}_P^*)$ ,  $p = 4$ .

Função	$\sigma_{\max}/\rho_{\max}$		Função	$\sigma_{\max}/\rho_{\max}$	
	Algoritmo			Algoritmo	
	RG/SMM	PLR		RG/SMM	PLR
$H_1(z)$	27,3	4,40	$H_9(z)$	21,8	4,19
$H_2(z)$	56,0	5,35	$H_{10}(z)$	31,0	5,36
$H_3(z)$	58,6	5,44	$H_{11}(z)$	31,9	5,56
$H_4(z)$	55,9	5,08	$H_{12}(z)$	29,4	5,19
$H_5(z)$	58,3	5,50	$H_{13}(z)$	31,3	5,60
$H_6(z)$	7,00	—	$H_{14}(z)$	5,18	—
$H_7(z)$	10,4	—	$H_{15}(z)$	5,56	—
$H_8(z)$	10,9	—	$H_{16}(z)$	5,74	—

Tabela 4.9: Relação entre o valor singular máximo da matrizes de estado para a estrutura direta e o para a estrutura polifásica,  $p = 4$ .

Função	$\rho_{\max}/\rho_{\min}$			$(\sigma_{\max}/\sigma_{\min})/(\rho_{\max}/\rho_{\min})$		
	Algoritmo			Algoritmo		
	RG/SMM	PLR	SHARF	RG/SMM	PLR	SHARF
$H_1$	$1,9 \times 10^1$	$1,4 \times 10^1$	$1,4 \times 10^1$	$3,1 \times 10^1$	2,7	1,6
$H_2$	$4,4 \times 10^3$	$2,2 \times 10^2$	$1,6 \times 10^1$	$9,5 \times 10^1$	7,1	1,2
$H_3$	$4,2 \times 10^6$	$2,0 \times 10^4$	$1,3 \times 10^2$	$9,7 \times 10^1$	7,2	1,3
$H_4$	$5,0 \times 10^2$	$2,7 \times 10^1$	3,8	$9,5 \times 10^1$	6,7	3,8
$H_5$	$4,8 \times 10^4$	$2,2 \times 10^2$	3,1	$9,6 \times 10^1$	7,3	4,8
$H_6$	$4,0 \times 10^1$	$3,2 \times 10^1$	$3,3 \times 10^1$	3,2	0,37	2,1
$H_7$	$8,3 \times 10^1$	$3,9 \times 10^1$	$2,1 \times 10^2$	$1,5 \times 10^1$	0,42	3,5
$H_8$	$7,0 \times 10^2$	$3,9 \times 10^1$	$2,0 \times 10^3$	$1,8 \times 10^1$	0,42	3,7

Tabela 4.10: Espalhamento dos valores singulares extremos de  $\mathbf{R}_P(\mathbf{w}_P^*)$  e relação com espalhamento para a estrutura direta,  $p = 4$ .

Função	$\rho_{\max}/\rho_{\min}$			$(\sigma_{\max}/\sigma_{\min})/(\rho_{\max}/\rho_{\min})$		
	Algoritmo			Algoritmo		
	RG/SMM	PLR	SHARF	RG/SMM	PLR	SHARF
$H_9(z)$	$1,2 \times 10^3$	$4,6 \times 10^2$	$2,0 \times 10^2$	$1,2 \times 10^1$	3,3	1,2
$H_{10}(z)$	$5,1 \times 10^5$	$2,0 \times 10^4$	$1,4 \times 10^3$	$2,1 \times 10^1$	5,4	1,6
$H_{11}(z)$	$4,9 \times 10^8$	$1,8 \times 10^6$	$1,4 \times 10^4$	$2,1 \times 10^1$	6,0	1,6
$H_{12}(z)$	$4,0 \times 10^5$	$1,6 \times 10^4$	$1,2 \times 10^3$	$2,4 \times 10^1$	6,3	1,9
$H_{13}(z)$	$3,6 \times 10^8$	$1,4 \times 10^6$	$1,1 \times 10^4$	$2,6 \times 10^1$	7,2	2,0
$H_{14}(z)$	$2,8 \times 10^1$	2,3	$3,2 \times 10^1$	3,7	2,8	1,3
$H_{15}(z)$	$1,8 \times 10^2$	$2,4 \times 10^1$	$2,1 \times 10^2$	6,4	0,31	2,1
$H_{16}(z)$	$1,9 \times 10^3$	$2,3 \times 10^1$	$2,0 \times 10^3$	6,4	0,32	2,2

Tabela 4.11: Espalhamento dos valores singulares extremos de  $\mathbf{R}_P(\mathbf{w}_P^*)$  e relação com espalhamento para a estrutura direta,  $p = 4$ .

um filtro adaptativo hipotético da forma

$$\widehat{H}(z) = \frac{\beta_0 + \beta_1 z + \dots + \beta_{pM} z^{pM}}{1 + \alpha_1 z + \dots + \alpha_{pM} z^{pM}}, \quad (4.114)$$

ou seja, um filtro que diferisse do filtro polifásico definido em (4.40) (Item 4.2.2) por ter todos os coeficientes do denominador, de  $\alpha_1$  a  $\alpha_{pM}$ . Se este filtro adaptativo hipotético fosse utilizado para identificar um sistema  $H(z)$  de ordem  $M$ , então a análise feita no Capítulo 3 garantiria que para todos os pontos estacionários da superfície de erro teríamos  $\widehat{H}(z) = H(z)$ . Poderia-se pensar que o filtro polifásico é apenas um caso particular deste filtro hipotético e que esta propriedade dos pontos estacionários é válida, de imediato, para o filtro polifásico. Isto, porém não é correto: para o filtro polifásico, um ponto estacionário só precisa ter derivadas nulas em relação aos coeficientes  $a_1$  a  $a_M$  dos termos  $z^p, z^{2p} \dots z^{pM}$  do denominador (além, é claro, de derivadas nulas em relação aos coeficientes do numerador), enquanto que para o filtro acima é necessário que as derivadas em relação aos coeficientes  $\alpha_1$  a  $\alpha_{pM}$  de todas as potências  $z, z^2, \dots, z^{pM}$  do denominador sejam nulas. A princípio, portanto, nada garante que para o filtro polifásico todos os pontos estacionários sejam da forma  $\widehat{H}(z) = H(z)$ . Demonstra-se, no entanto, o seguinte teorema:

**Teorema 4.1** *Se  $\widehat{H}(z)$  tem a forma dada (4.40) e se  $M \geq \deg[H(z)]$ , então  $\widehat{H}(z)$  é um ponto estacionário de  $\|H(z) - \widehat{H}(z)\|^2$  se, e somente, se  $\widehat{H}(z) = H(z)$ .*

*Prova:* A demonstração deste teorema é dada a seguir. Como trata-se de uma demonstração longa foi continuada a divisão do texto em itens e sub-itens.

#### 4.4.1 Forma polifásica do erro de aproximação

Usando (4.40) e as formas (4.9) e (4.10) das funções  $H_l(z)$  da expansão polifásica de  $H(z)$ , pode-se escrever também para  $\widehat{H}(z)$ ,

$$\begin{aligned} \widehat{H}(z) &= \frac{b_{p,0} + b_{p,1}z + \dots + b_{p,pM}z^{pM}}{1 + a_{p,1}z^p + \dots + a_{p,M}z^{pM}} \\ &= \sum_{l=0}^{p-1} \frac{B_l(z)}{A_p(z^p)}, \end{aligned} \quad (4.115)$$

onde

$$\frac{B_0(z)}{A_P(z)} = \frac{b_{0,0} + b_{0,1}z + \dots + b_{0,M} z^M}{1 + a_{P,1}z + \dots + d_{P,M} z^M} = \widehat{H}_0(z) \quad (4.116)$$

e

$$\frac{B_l(z)}{A_P(z)} = \frac{b_{l,0} + b_{l,1}z + \dots + b_{l,M-1} z^{M-1}}{1 + d_{P,1}z + \dots + d_{P,M} z^M} = \widehat{H}_l(z), \quad l = 1, 2, \dots, p-1. \quad (4.117)$$

Utilizando agora a Propriedade 4.4, o erro de aproximação  $\|H(z) - \widehat{H}(z)\|^2$  pode ser escrito em função do erro de aproximação  $\|H_l(z) - \widehat{H}_l(z)\|^2$  de cada fase como:

$$\begin{aligned} \|H(z) - \widehat{H}(z)\|^2 &= \langle H(z) - \widehat{H}(z), H(z) - \widehat{H}(z) \rangle \\ &= \sum_{l=0}^{p-1} \langle H_l(z) - \widehat{H}_l(z), H_l(z) - \widehat{H}_l(z) \rangle \\ &= \sum_{l=0}^{p-1} \|H_l(z) - \widehat{H}_l(z)\|^2. \end{aligned} \quad (4.118)$$

#### 4.4.2 Otimização dos zeros com os polos fixados

Uma condição necessária para a otimização dos zeros de  $\widehat{H}(z)$  é que

$$\frac{\partial}{\partial b_{l,k}} \|H(z) - \widehat{H}(z)\|^2 = 0, \quad \begin{cases} l = 0, & k = 0, 1, \dots, M, \\ l = 1, 2, \dots, p-1, & k = 0, 1, \dots, M-1. \end{cases} \quad (4.119)$$

O procedimento seguido em [1] pode agora ser generalizado para o caso de um filtro polifásico.

Temos, inicialmente, usando (4.118) e (1.21) que:

$$\begin{aligned} \frac{\partial}{\partial b_{l,k}} \|H(z) - \widehat{H}(z)\|^2 &= \frac{\partial}{\partial b_{l,k}} \sum_{j=0}^{p-1} \| [H_j(z) - \widehat{H}_j(z)] \|^2 \\ &= \frac{\partial}{\partial b_{l,k}} \| [H_l(z) - \widehat{H}_l(z)] \|^2 \\ &= \frac{\partial}{\partial b_{l,k}} \langle H_l(z) - \widehat{H}_l(z), H_l(z) - \widehat{H}_l(z) \rangle \\ &= 2 \left\langle \frac{\partial}{\partial b_{l,k}} [H_l(z) - \widehat{H}_l(z)], H_l(z) - \widehat{H}_l(z) \right\rangle. \end{aligned} \quad (4.120)$$

### Coefficientes da fase 0

Consideremos primeiro o caso  $l = 0$ . Reunindo (4.119) e (4.120) obtemos o seguinte sistema de equações:

$$\left\langle \left[ \frac{1}{A_P(z)} \frac{z}{A_P(z)} \cdots \frac{z^M}{A_P(z)} \right]^\top, H_0(z) - \widehat{H}_0(z) \right\rangle = \mathbf{0}_{M+1}. \quad (4.121)$$

Notar que nesta notação sintética temos na realidade  $M + 1$  produtos internos no lado esquerdo da igualdade. Em consequência da Propriedade 3.3 a expressão acima implica em que  $H_0(z) - \widehat{H}_0(z)$  é dado por

$$H_0(z) - \widehat{H}_0(z) = g_0(z)V_P(z), \quad z^{-1}g_0(z) \in \mathcal{H}_2, \quad (4.122)$$

onde  $V_P(z)$  é dado por

$$V_P(z) = \frac{z^M A_P(z^{-1})}{A_P(z)}. \quad (4.123)$$

Notar que  $z^{-1}g_0(z) \in \mathcal{H}_2$  significa que  $g_0(z)$  é estritamente causal.

A expressão (4.122) pode ser escrita como

$$V_P(z^{-1})H_0(z) - V_P(z^{-1})\widehat{H}_0(z) = g_0(z), \quad (4.124)$$

pois  $V_P(z)$  é passa-tudo e portanto  $V_P(z)V_P(z^{-1}) = 1$ . Além disso temos que

$$\begin{aligned} V_P(z)\widehat{H}_0(z^{-1}) &= \frac{z^M B_0(z^{-1})}{A_P(z)} \\ &= \frac{b_{0,M} + b_{0,M-1}z^1 + \dots + b_{0,0}z^M}{A_P(z)} \in \mathcal{H}_2, \end{aligned} \quad (4.125)$$

e portanto  $V_P(z^{-1})\widehat{H}_0(z)$  é anti-causal. Usando este fato em (4.124) resulta que  $g_0(z)$  é dada apenas pela projeção estritamente causal (ver definição no Item 1.5.1) de  $V_P(z^{-1})H_0(z)$ :

$$g_0(z) = [V_P(z^{-1})H_0(z)]_+. \quad (4.126)$$

### Coefficientes das demais fases

Consideremos agora as demais fases. Reunindo (4.119) e (4.120) obtemos os  $p - 1$  sistemas de equações a seguir:

$$\left\langle \left[ \frac{1}{A_P(z)} \frac{z}{A_P(z)} \cdots \frac{z^{M-1}}{A_P(z)} \right]^\top, H_l(z) - \widehat{H}_l(z) \right\rangle = \mathbf{0}_M, \quad l = 1, 2, \dots, p - 1, \quad (4.127)$$

o que, em consequência da propriedade 3.3 implica em que  $H_l(z) - \widehat{H}_l(z)$  é dado por

$$H_l(z) - \widehat{H}_l(z) = g_l(z)V_P(z), \quad g_l(z) \in \mathcal{H}_2, \quad l = 1, 2, \dots, p - 1. \quad (4.128)$$

Analogamente ao que foi feito para  $l = 0$ , a expressão acima pode ser escrita como

$$V_P(z^{-1})H_l(z) - V_P(z^{-1})\widehat{H}_l(z) = g_l(z). \quad (4.129)$$

Neste caso, porém,  $V_P(z)\widehat{H}_l(z^{-1}) = z^M B_l(z^{-1})/A_P(z)$  é *estritamente* causal pois  $b_{l,M} = 0$  para  $l = 1, 2, \dots, p - 1$ :

$$\begin{aligned} z^{-1}V_P(z)\widehat{H}_l(z^{-1}) &= z^{-1} \frac{z^M B_l(z^{-1})}{A_P(z)} \\ &= z^{-1} \frac{b_{l,M-1}z^1 + \dots + b_{l,0}z^M}{A_P(z)} \in \mathcal{H}_2, \quad l = 1, 2, \dots, p - 1. \end{aligned} \quad (4.130)$$

Portanto,  $V_P(z^{-1})\widehat{H}_l(z)$  é *estritamente* anti-causal e de (4.128) e (4.129) resulta que  $g_l(z)$  é dada apenas pela projeção *causal* de  $V_P(z^{-1})H_l(z)$ :

$$g_l(z) = [V_P(z^{-1})H_l(z)]_{(+)}, \quad l = 1, 2, \dots, p - 1. \quad (4.131)$$

De (4.118), (4.122) e (4.128) resulta ainda que, quando os zeros são otimizados com os polos fixados, o erro de aproximação é dado por

$$\begin{aligned} \|H(z) - \widehat{H}(z)\|^2 &= \sum_{l=0}^{p-1} \langle H_l(z) - \widehat{H}_l(z), H_l(z) - \widehat{H}_l(z) \rangle \\ &= \sum_{l=0}^{p-1} \langle g_l(z)V_P(z), g_l(z)V_P(z) \rangle \end{aligned}$$



$$\begin{aligned}
&= \sum_{l=0}^{p-1} \langle g_l(z) V_P(z) V_P(z^{-1}), g_l(z) \rangle \\
&= \sum_{l=0}^{p-1} \langle g_l(z), g_l(z) \rangle.
\end{aligned} \tag{4.132}$$

### 4.4.3 Otimização dos zeros e polos

Uma condição necessária para a otimização dos polos de  $\widehat{H}(z)$  é que

$$\frac{\partial}{\partial a_{P,k}} \left\| H(z) - \widehat{H}(z) \right\|^2 = 0, \quad k = 1, 2, \dots, M. \tag{4.133}$$

Assim como no caso da otimização dos zeros, o procedimento seguido em [1] pode ser generalizado para o caso de um filtro polifásico. Temos então, usando (4.118) e (1.21):

$$\begin{aligned}
\frac{\partial}{\partial a_{P,k}} \left\| H(z) - \widehat{H}(z) \right\|^2 &= \frac{\partial}{\partial a_{P,k}} \sum_{l=0}^{p-1} \langle H_l(z) - \widehat{H}_l(z), H_l(z) - \widehat{H}_l(z) \rangle \\
&= \sum_{l=0}^{p-1} \frac{\partial}{\partial a_{P,k}} \langle H_l(z) - \widehat{H}_l(z), H_l(z) - \widehat{H}_l(z) \rangle \\
&= 2 \sum_{l=0}^{p-1} \left\langle \frac{\partial}{\partial a_{P,k}} [H_l(z) - \widehat{H}_l(z)], H_l(z) - \widehat{H}_l(z) \right\rangle.
\end{aligned} \tag{4.134}$$

Agora, as derivadas acima são dadas por

$$\begin{aligned}
\frac{\partial}{\partial a_{P,k}} \widehat{H}_l(z) &= \frac{\partial}{\partial a_{P,k}} \frac{B_l(z)}{A_P(z)} \\
&= -\frac{B_l(z)}{A_P^2(z)} \frac{\partial A_P(z)}{\partial a_{P,k}} \\
&= -\widehat{H}_l(z) \frac{1}{A_P(z)} \frac{\partial A_P(z)}{\partial a_{P,k}} = -\widehat{H}_l(z) \frac{z^k}{A_P(z)}.
\end{aligned} \tag{4.135}$$

Reunindo então (4.134) e (4.135) obtemos o seguinte sistema de equações:

$$\sum_{l=0}^{p-1} \left\langle z^k \widehat{H}_l(z), H_l(z) - \widehat{H}_l(z) \right\rangle = \left\langle \mathcal{C}(z), \sum_{l=0}^{p-1} z^{-l} \widehat{H}_l(z^{-1}) [H_l(z) - \widehat{H}_l(z)] \right\rangle = \mathbf{0}_M, \tag{4.136}$$

onde, como já visto anteriormente,  $\mathcal{C}(z) = [1/A_P(z) \ z/A_P(z) \ \dots \ z^{M-1}/A_P(z)]^\top$ .

Quando os zeros são otimizados, como visto no item anterior, temos  $H_l(z) - \hat{H}_l(z) = g_l(z)V_P(z)$  e, portanto, a otimização conjunta de polos e zeros resulta em:

$$\left\langle \mathcal{C}(z), \sum_{l=0}^{p-1} z^{-1} \hat{H}_l(z^{-1}) V_P(z) g_l(z) \right\rangle = \mathbf{0}_M \quad (4.137)$$

Definindo o polinômio recíproco, relativo a  $z^M$ , de um polinômio  $P(z)$  qualquer como  $\bar{P}(z) = z^M P(z^{-1})$ , temos:

$$\begin{aligned} \hat{H}_l(z^{-1}) V_P(z) &= \frac{B_l(z^{-1}) z^M A_P(z^{-1})}{A_P(z^{-1}) A_P(z)} \\ &= \frac{z^M B_l(z^{-1})}{A_P(z)} \\ &= \frac{\bar{B}_l(z)}{A_P(z)} \end{aligned} \quad (4.138)$$

e usando isto em (4.137) temos

$$\left\langle \mathcal{C}(z), \sum_{l=0}^{p-1} z^{-1} \frac{\bar{B}_l(z)}{A_P(z)} g_l(z) \right\rangle = \mathbf{0}_M \quad (4.139)$$

Para aplicar agora (3.3) devemos mostrar que o segundo termo no produto interno acima é uma função que pertence a  $\mathcal{H}_2$ . De fato, temos:

- $l = 0 : \bar{B}_0(z)/A_P(z) \in \mathcal{H}_2$ ; e, conforme (4.122),  $z^{-1}g_0(z) \in \mathcal{H}_2$
- $l = 1, 2, \dots, p-1 : z^{-1}\bar{B}_l(z)/A_P(z)$  é causal pois, com  $b_{l,M} = 0$ ,  $\bar{B}_l(z) = z^M B_l(z^{-1}) = z^M [b_{l,0} + b_{l,1}z^{-1} + \dots + b_{l,M-1}z^{-M+1}]$  é estritamente causal; e, conforme (4.128),  $g_l(z) \in \mathcal{H}_2$ .

Aplicando então (3.3) em (4.139), e substituindo  $V_P(z)$  por  $\bar{A}_P(z)/A_P(z)$  obtemos uma condição necessária para a otimização conjunta dos polos e zeros do filtro adaptativo polifásico:

$$\sum_{l=0}^{p-1} \bar{B}_l(z) z^{-1} g_l(z) = \bar{A}_P(z) Q(z), \quad Q(z) \in \mathcal{H}_2. \quad (4.140)$$

A seguir será analisada a forma das funções  $g_l(z)$ , o que será necessário para a análise dos pontos estacionários da superfície de erro.

#### 4.4.4 Forma das funções $g_l(z)$

##### Função $g_l(z)$ para a fase 0

Como expresso em (4.126),  $g_0(z) = [V_P(z^{-1})H_0(z)]_+$ . Escrevendo  $H_0(z)$  como

$$H_0(z) = \frac{C_0(z)}{D_P(z)} = \frac{c_{0,0} + c_{0,1}z + \dots + c_{0,N_{c,0}} z^{N_{c,0}}}{1 + d_{P,1}z + \dots + d_{P,N_d} z^{N_d}}, \quad (4.141)$$

com  $c_{0,N_{c,0}} \neq 0$  e  $d_{P,N_d} \neq 0$ , e notando que  $V_P(z^{-1}) = A_P(z)/\bar{A}_P(z)$ , temos que

$$\deg[A_P(z)C_0(z)] \leq M + N_{c,0}, \quad (4.142)$$

e

$$\deg[\bar{A}_P(z)D_P(z)] = M + N_d. \quad (4.143)$$

Fazendo uma expansão em frações parciais de  $V_P(z^{-1})H_0(z)$  e reagrupando os termos convenientemente podemos escrever

$$V_P(z^{-1})H_0(z) = \frac{R_{d,0}(z)}{D_P(z)} + \frac{R_{a,0}(z)}{\bar{A}_P(z)} + r_{0,0} + r_{0,1}z + \dots + r_{0,N_{r,0}} z^{N_{r,0}}, \quad (4.144)$$

onde

$$N_{r,0} \leq \max(0, M + N_{c,0} - M - N_d) = \max(0, N_{c,0} - N_d) \quad (4.145)$$

e

$$\deg[R_{d,0}(z)] \leq \max(0, N_d - 1). \quad (4.146)$$

Agora, para uma função racional  $F(z)$  como abaixo (onde com  $q_N \neq 0$ ), sempre podemos escrever

$$\begin{aligned} F(z) &= \frac{P(z)}{Q(z)} = \frac{p_0 + p_1z + \dots + p_{N-1}z^{N-1}}{1 + q_1z + \dots + q_Nz^N} \\ &= p_0 + \frac{P(z) - p_0Q(z)}{Q(z)} \end{aligned}$$

$$= p_0 + \frac{P'(z)}{Q(z)}, \quad (4.147)$$

onde  $P'(z)$  resulta estritamente causal e  $\deg[P'(z)] = N$ . Podemos escrever portanto

$$\frac{R_{d,0}(z)}{D_P(z)} = r'_{0,0} + \frac{R'_0(z)}{D_P(z)}, \quad (4.148)$$

onde  $\deg[R'_0(z)] = N_d$  e  $R'_0(z)$  é estritamente causal.

Além disso, uma vez que  $\deg[R_{a,0}(z)] \leq M - 1$ , temos que

$$\frac{R_{a,0}(z)}{A_P(z)} = \frac{z^{-M} R_{a,0}(z)}{A_P(z^{-1})} \quad (4.149)$$

é estritamente anti-causal e, portanto,

$$\begin{aligned} [V_P(z^{-1})H_0(z)]_+ &= \left[ r'_{0,0} + \frac{R'_0(z)}{D_P(z)} \right]_+ + \left[ \frac{R_{a,0}(z)}{A_P(z)} \right]_+ + [r_{0,0} + r_{0,1}z + \dots + r_{0,N_r,0}z^{N_r,0}]_+ \\ &= \frac{R'_0(z)}{D_P(z)} + r_{0,1}z + \dots + r_{0,N_r,0}z^{N_r,0} \\ &= \frac{R_0(z)}{D_P(z)}, \end{aligned} \quad (4.150)$$

onde

$$\begin{aligned} \deg[R_0(z)] &= N_d + N_{r,0} \\ &= N_d + \max(0, N_{c,0} - N_d) \\ &= \max(N_d, N_{c,0}) \\ &\leq \deg[H_0(z)]. \end{aligned} \quad (4.151)$$

### Função $g_l(z)$ para as demais fases

Procedendo analogamente ao que foi feito acima, para  $g_l(z)$  dado por 4.131 chegamos a

$$\begin{aligned} [V_P(z^{-1})H_l(z)]_{(+)} &= \left[ \frac{R_{d,l}(z)}{D_P(z)} \right]_{(+)} + \left[ \frac{R_{a,l}(z)}{A(z)} \right]_{(+)} + [r_{l,0} + r_{l,1}z + \dots + r_{l,N_r,l}z^{N_r,l}]_{(+)} \\ &= \frac{R_{d,l}(z)}{D_P(z)} + r_{l,0} + r_{l,1}z + \dots + r_{l,N_r,l}z^{N_r,l} \end{aligned}$$

$$= \frac{R_l(z)}{D_P(z)}, \quad l = 1, 2, \dots, p-1. \quad (4.152)$$

onde

$$\begin{aligned} \deg[R_l(z)] &= N_d + \max(0, N_{c,l} - N_d) \\ &= \max(N_d, N_{c,l}) \\ &\leq \deg[H_l(z)]. \end{aligned}$$

Genericamente temos então

$$g_l(z) = \frac{R_l(z)}{D_P(z)}, \quad \text{com } \deg[R_l(z)] \leq \deg[H_l(z)], \quad l = 0, 1, \dots, p-1. \quad (4.153)$$

#### 4.4.5 Pontos estacionários da superfície de erro

##### Condição suficiente para ponto estacionário

Inicialmente obtemos uma condição suficiente para que  $\hat{H}(z)$  seja um ponto estacionário da superfície de erro. Consideremos o caso hipotético de um filtro adaptativo  $\hat{H}_h(z)$  com  $pM$  coeficientes no numerador e no denominador, isto é,

$$\hat{H}_h(z) = \frac{\beta_0 + \beta_1 z + \dots + \beta_{pM} z^{pM}}{1 + \alpha_1 z + \dots + \alpha_{pM} z^{pM}}. \quad (4.154)$$

Qualquer  $\hat{H}_h(z)$  tal que  $\hat{H}_h(z) = H(z)$  atende à condição (3.14) com  $Q(z) = g(z) \equiv 0$  e é, portanto, um ponto estacionário da superfície de erro. Segue que, para  $\hat{H}(z)$  polifásico, isto é,

$$\hat{H}(z) = \frac{b_{P,0} + b_{P,1}z + \dots + b_{P,pM} z^{pM}}{1 + a_{P,1}z^p + \dots + a_{P,M} z^{pM}}, \quad (4.155)$$

qualquer  $\hat{H}(z)$  tal que  $\hat{H}(z) = H(z)$  é um ponto estacionário da superfície de erro, uma vez que  $\hat{H}(z)$  faz parte do conjunto de funções da forma definida em (4.154). Portanto  $\hat{H}(z) = H(z)$  é uma condição suficiente para que  $\hat{H}(z)$  seja um ponto estacionário da superfície de erro. A seguir obtemos uma condição necessária para que  $\hat{H}(z)$  seja um ponto estacionário da superfície de erro.

### Condição necessária para ponto estacionário

Reunindo (4.122),(4.128) e (4.153) podemos escrever

$$\begin{aligned} H_l(z) - \widehat{H}_l(z) &= g_l(z)V(z) \\ &= \frac{R_l(z) \overline{A}_P(z)}{D_P(z) A_P(z)}, \quad l = 0, 1 \dots p-1, \end{aligned} \quad (4.156)$$

e, portanto, usando (4.116) e (4.117) temos

$$\frac{C_l(z)}{D_P(z)} - \frac{B_l(z)}{A_P(z)} = \frac{R_l(z) \overline{A}_P(z)}{D_P(z) A_P(z)}, \quad l = 0, 1 \dots p-1, \quad (4.157)$$

o que resulta em

$$B_l(z) = \frac{C_l(z)A_P(z) - R_l(z)\overline{A}_P(z)}{D_P(z)}, \quad (4.158)$$

ou, equivalentemente,

$$z^M B_l(z^{-1}) = \frac{z^M C_l(z^{-1})z^M A_P(z^{-1}) - z^M R_l(z^{-1})z^M \overline{A}_P(z^{-1})}{z^M D_P(z^{-1})}, \quad (4.159)$$

ou ainda

$$\overline{B}_l(z) = \frac{\overline{C}_l(z)\overline{A}_P(z) - \overline{R}_l(z)A_P(z)}{\overline{D}_P(z)}, \quad l = 0, 1 \dots p-1. \quad (4.160)$$

Utilizando as expressões acima para  $\overline{B}_l(z)$  e  $g_l(z)$  em (4.140), que descreve a condição a ser obedecida nos pontos estacionários, temos

$$\sum_{l=0}^{p-1} z^{-1} \frac{\overline{C}_l(z)\overline{A}_P(z) - \overline{R}_l(z)A_P(z)}{\overline{D}_P(z)} \frac{R_l(z)}{D_P(z)} = \overline{A}_P(z)Q(z). \quad (4.161)$$

Nos zeros  $\{z_j\}_{j=1}^M$  de  $\overline{A}_P(z)$  a expressão acima se reduz a

$$\sum_{l=0}^{p-1} \overline{R}_l(z_j)R_l(z_j) = 0, \quad j = 1, 2 \dots M, \quad (4.162)$$

pois  $\overline{A}_P(z)$  e  $z^{-1}A_P(z)$  não tem nenhum zero em comum. Conseqüentemente, nos zeros

$\{z_j^{-1}\}_{j=1}^M$  de  $A_P(z)$  temos:

$$\sum_{l=0}^{p-1} \overline{R_l}(z_j^{-1}) R_l(z_j^{-1}) = z_j^{-2M} \sum_{l=0}^{p-1} R_l(z_j) \overline{R_l}(z_j) = 0, \quad j = 1, 2, \dots, M. \quad (4.163)$$

Por outro lado, sempre podemos escrever

$$\begin{aligned} \sum_{l=0}^{p-1} \overline{R_l}(z) R_l(z) &= z^M \sum_{l=0}^{p-1} R_l(z^{-1}) R_l(z) \\ &= z^M s(z^{-1}) s(z) \\ &= \overline{s}(z) s(z), \end{aligned} \quad (4.164)$$

para algum  $s(z)$ . E de (4.162) e (4.163) decorre que  $\overline{s}(z)s(z)$  é divisível por  $\overline{A}_P(z)$  e  $A_P(z)$ , isto é, tem a forma

$$\overline{s}(z)s(z) = \overline{A}_P(z) A_P(z) q(z^{-1}) q(z), \quad (4.165)$$

para algum  $q(z) = \sum_{k=0}^{N_q} q_k z^k \in \mathcal{H}_2$ .

Sempre é possível escolher  $\overline{s}(z)$  e  $s(z)$  tais que todos seus zeros estejam, respectivamente, dentro e fora da circunferência unitária, o mesmo valendo para  $q(z)$  e  $q(z^{-1})$ . Desta maneira, de (4.165) resulta que

$$\overline{s}(z) = \overline{A}_P(z) q(z), \quad q(z) \in \mathcal{H}_2, \quad (4.166)$$

e também  $s(z) = A_P(z) q(z^{-1})$ . Notar ainda que, devido a (4.164) e (4.153) temos que

$$\begin{aligned} \deg[\overline{s}(z)] &\leq \max_l \{\deg[H_l(z)]\} \\ &\leq \deg[H(z)]. \end{aligned} \quad (4.167)$$

Usando agora (4.164) em (4.161) obtemos

$$\left[ \sum_{l=0}^{p-1} \frac{\overline{C}_l(z) R_l(z) \overline{A}_P(z)}{\overline{D}_P(z) D_P(z)} \right] - \frac{\overline{s}(z) s(z) A_P(z)}{\overline{D}_P(z) D_P(z)} = z \overline{A}_P(z) Q(z), \quad (4.168)$$

o que, com (4.166) resulta em

$$\left[ \sum_{l=0}^{p-1} \frac{\overline{C}_l(z) R_l(z) \overline{A}_P(z)}{\overline{D}_P(z) D_P(z)} \right] - \frac{\overline{A}_P(z) q(z) q(z^{-1}) [A_P(z)]^2}{\overline{D}_P(z) D_P(z)} = z \overline{A}_P(z) Q(z), \quad (4.169)$$

ou

$$\left[ \sum_{l=0}^{p-1} \overline{C}_l(z) R_l(z) \right] - q(z)q(z^{-1})[A_P(z)]^2 = zQ(z)\overline{D}_P(z)D_P(z), \quad (4.170)$$

isto é, o lado esquerdo da igualdade acima deve ser uma função estritamente causal.

Consideremos agora que  $M = N = \deg[H(z)]$ . De (4.166) e (4.167) resulta que  $\deg[q(z)] = 0$ , ou seja,  $q(z) = q_0$ . Portanto, (4.170) pode ser escrito como

$$\left[ \sum_{l=0}^{p-1} \overline{C}_l(z) R_l(z) \right] - q_0^2[A_P(z)]^2 = zQ(z)\overline{D}_P(z)D_P(z). \quad (4.171)$$

Verifiquemos como pode ser atendida a condição de causalidade estrita. De sua parte, a função  $\sum_{l=0}^{p-1} \overline{C}_l(z) R_l(z)$  é estritamente causal, como pode ser visto considerando as suas parcelas:

- $l = 0$ :  $\overline{C}_0(z)$  é causal; e segundo (4.153) e (4.126)  $R_0(z)$  é estritamente causal;
- $l = 1, 2, \dots, p-1$ :  $R_l(z)$  é causal; e, uma vez que, conforme (4.117),  $c_{l,N} = 0$  neste caso,

$$\overline{C}_l(z) = z^M(c_{l,0} + c_{l,1}z^{-1} + \dots + c_{l,N-1}z^{-N+1})$$

é estritamente causal.

Por outro lado,  $[A_P(z)]^2$  não é estritamente causal, e portanto, para atender (4.171) é necessário que  $q_0 = 0$ . Portanto,  $\overline{s}(z) \equiv 0$  e  $\sum_{l=0}^{p-1} \overline{R}_l(z) R_l(z) = \overline{s}(z)s(z) \equiv 0$ . Usando este resultado com (4.132) e (4.153) temos

$$\begin{aligned} \left\| H(z) - \widehat{H}(z) \right\|_2^2 &= \sum_{l=0}^{p-1} \langle g_l(z), g_l(z) \rangle \\ &= \left\langle \sum_{l=0}^{p-1} g_l(z)g_l(z^{-1}), 1 \right\rangle \\ &= \left\langle \frac{\sum_{l=0}^{p-1} R_l(z)R_l(z^{-1})}{D_P(z)D_P(z^{-1})}, 1 \right\rangle \\ &= \left\langle \frac{z^{-M}\overline{s}(z)s(z)}{D_P(z)D_P(z^{-1})}, 1 \right\rangle \\ &= 0 \end{aligned}$$



e conseqüentemente  $H(z) = \hat{H}(z)$ . Ou seja, quando  $N = M$ , uma condição necessária para  $\hat{H}(z)$  ser um ponto estacionário da superfície de erro é  $\hat{H}(z) = H(z)$ .

Consideremos agora que  $M > N = \deg[H(z)]$ . De (4.166) e (4.167) resulta que  $q(z) \equiv 0$ , o que, como acima, implica em  $\hat{H}(z) = H(z)$ , isto é, há um cancelamento de polos e zeros em  $\hat{H}(z)$ .

Com isso está completa a demonstração do Teorema 4.1.

## 4.5 Efeito da posição dos polos na superfície de erro reduzida

Neste item mostra-se que quando  $H(z)$  tem polos próximos à circunferência unitária a superfície de erro reduzida para a estrutura polifásica não conterà regiões tão planas quanto a superfície de erro reduzida para a estrutura direta.

### 4.5.1 Expressão para superfície de erro reduzida

Inicialmente, conforme (4.122 e (4.128), quando os zeros das funções  $\hat{H}_l(z)$  são otimizados em função de seus polos temos:

$$H_l(z) - \hat{H}_l(z) = g_l(z)V_P(z), \quad z^{-1}g_0(z) \in \mathcal{H}_2, \quad g_l(z) \in \mathcal{H}_2, \quad l = 1, 2, \dots, p-1. \quad (4.172)$$

Neste caso podemos escrever então:

$$\begin{aligned} H(z) - \hat{H}(z) &= \sum_{l=0}^{p-1} z^l [H_l(z^p) - \hat{H}_l(z^p)] \\ &= \sum_{l=0}^{p-1} z^l g_l(z^p) V_P(z^p) \\ &= g_P(z) V_P(z^p), \end{aligned} \quad (4.173)$$

onde  $g_P(z) = \sum_{l=0}^{p-1} z^l g_l(z^p)$  é estritamente causal, pois  $g_0(z)$  é estritamente causal. Analogamente ao caso da estrutura direta (Item 3.2.3), uma expressão conveniente para  $g_P(z)$  pode

ser obtida re-escrevendo inicialmente (4.173) como

$$V_P(z^{-p})H(z) - V_P(z^{-p})\widehat{H}(z) = V_P(z^{-p})V_P(z^p)g_P(z) = g_P(z), \quad z^{-1}g_P(z) \in \mathcal{H}_2. \quad (4.174)$$

Usando (4.40), consideremos agora a parcela

$$\begin{aligned} V_P(z^{-p})\widehat{H}(z) &= \frac{z^{-pM}A_P(z^p)}{A_P(z^{-p})} \frac{B_P(z)}{A_P(z^p)} = \frac{z^{-pM}B_P(z)}{A_P(z^{-p})} \\ &= \frac{b_{P,pM} + b_{P,p(M-1)}z^{-1} + \dots + b_{P,0}z^{-pM}}{A_P(z^{-p})}, \end{aligned} \quad (4.175)$$

que analogamente ao caso da estrutura direta é anti-causal pois  $z^{-pM}B_P(z)$  não tem potências positivas de  $z$ . Como  $g_P(z)$  é estritamente causal, segue que  $g_P(z)$  é dado apenas pela projeção estritamente causal de  $V_P(z^{-p})H(z)$ . Segue então a propriedade:

**Propriedade 4.5** *A função  $g_P(z)$  em 4.173 pode ser escrita como*

$$g_P(z) = [V_P(z^{-p})H(z)]_+ \quad (4.176)$$

De (4.173) temos que quando os zeros de  $\widehat{H}(z)$  na forma polifásica são otimizados em função dos seus polos temos

$$\|H(z) - \widehat{H}(z)\|^2 = \|V_P(z^p)g_P(z)\|^2 = \|g_P(z)\|^2, \quad (4.177)$$

pois  $V_P(z^p)$  é passa-tudo. De onde segue a definição:

**Definição 4.2** *A superfície de erro reduzida de  $H(z) - \widehat{H}(z)$  quando  $\widehat{H}(z)$  é implementada na forma polifásica (4.40) é dada por  $\|g_P(z)\|^2$  como função dos parâmetros de  $A_P(z)$ , com  $g_P(z)$  dado por 4.176.*

#### 4.5.2 Efeito da posição dos polos de $H(z)$

Podemos agora, analogamente ao feito no Item 3.6 para a estrutura direta, verificar o efeito da existência de polos de  $H(z)$  próximos à circunferência unitária na superfície de erro reduzida para a estrutura polifásica. Seguindo um procedimento análogo ao que leva a (3.92), resulta

que, quando existem  $m$  polos  $\nu_i$  de  $H(z)$  próximos à circunferência unitária, temos

$$\|g_P(z)\|^2 \simeq \sum_{i=1}^m \frac{|R_{P,i}|^2}{|\nu_i|^2 - 1}, \quad (4.178)$$

onde o termo  $|R_{P,i}|$  é dado por

$$|R_{P,i}| = |V_P(\nu_i^{-p})| \left| \frac{r_i}{\nu_i} \right|. \quad (4.179)$$

Agora, uma vez que  $\nu_i^p$  está mais longe da circunferência unitária do que  $\nu_i$ , resulta que, na situação em que os polos de  $\hat{H}(z)$  estão distantes dos polos de  $H(z)$ ,  $|V_P(\nu_i^{-p})|$  não será tão constante quanto  $|V(\nu_i^{-1})|$ . Portanto, em regiões afastadas do mínimo global, a superfície de erro reduzida para a estrutura polifásica não será tão plana quanto a superfície de erro reduzida para a estrutura direta, o que como exemplificado no Item 4.6 resulta, para os algoritmos de passo constante testados, em uma convergência global mais rápida.

Para exemplificar a análise acima, foram consideradas, como no Item 3.6, as funções  $H_1(z)$ ,  $H_4(z)$ ,  $H_9(z)$  e  $H_{12}(z)$ , definidas na Tabela 2.2 do Item 2.6.8. As superfícies de erro reduzidas dadas por (4.177) foram computadas para serem comparadas com as do caso da estrutura direta. Os resultados estão nas Figuras 4-6 a 4-9, onde, para maior facilidade de comparação, foram repetidas os gráficos das superfícies de erro para o caso da estrutura direta. Nas figuras pode ser visto o efeito da redução da planicidade da superfície de erro em regiões afastadas do mínimo global.

## 4.6 Exemplos da simulação dos algoritmos de adaptação para a estrutura polifásica

Neste item são apresentados resultados práticos referentes ao filtro polifásico adaptativo proposto. Partindo, em cada caso, de um valor inicial de  $\hat{H}(z)$  dado pela expansão polifásica do ponto inicial de  $\hat{H}(z)$  adotado no caso correspondente para a estrutura direta (ver Tabela 2.9), foram testados os algoritmos RG e PLR. Partindo da origem do espaço de parâmetros foram testados os algoritmos RG, PLR e SMM. O algoritmo SHARF não foi testado uma vez que, pela análise do Item 4.3.4, não seriam esperados, em termos de velocidade de convergência local, grandes ganhos de desempenho, e, como discutido no Item 3.7, também

Função	Algoritmo			
	RG		PLR	
	$\mu_P$	$n_{60,P}$	$\mu_P$	$n_{60,P}$
$H_1(z)$	$4,0 \times 10^{-2}$	$1,0 \times 10^3$	$6 \times 10^{-2}$	$7,0 \times 10^2$
$H_4(z)$	$6,5 \times 10^{-4}$	$6,4 \times 10^4$	$1,5 \times 10^{-2}$	$2,6 \times 10^3$
$H_6(z)$	$6 \times 10^{-2}$	$1,6 \times 10^3$	$6 \times 10^{-2}$	$1,6 \times 10^3$
$H_7(z)$	$1 \times 10^{-2}$	$1,3 \times 10^5$	$2 \times 10^{-2}$	$6,3 \times 10^4$
$H_9(z)$	$1,5 \times 10^{-2}$	$2,7 \times 10^4$	$7 \times 10^{-2}$	$8,6 \times 10^3$
$H_{14}(z)$	$4 \times 10^{-2}$	$1,1 \times 10^3$	$8 \times 10^{-2}$	$1,1 \times 10^3$
$H_{15}(z)$	$4 \times 10^{-3}$	$3,1 \times 10^4$	$1,5 \times 10^{-2}$	$3,0 \times 10^4$

Tabela 4.12: Resultados da execução dos algoritmos RG e PLR para a estrutura polifásica,  $p = 4$ , com início fora da origem;  $\mu$ : passo de adaptação máximo;  $n_{60}$ : número de iterações para erro quadrático de saída ficar 60 dB abaixo da potência de saída.

não haveria sentido em testar a sua convergência global, em face à idealização  $P(z) = D(z)$  adotada.

Os algoritmos foram testados na identificação das funções  $H_1(z)$ ,  $H_4(z)$ ,  $H_6(z)$ ,  $H_7(z)$ ,  $H_9(z)$ ,  $H_{14}(z)$  e  $H_{15}(z)$  definidas na Tabela 2.2 do Item 2.6. Todos os procedimentos relativos à execução dos algoritmos mencionados no Item 2.7 permanecem válidos. Os valores de  $\mu_P$  utilizados e o tempo de convergência  $n_{60,P}$  obtidos (ver Item 2.6 para a metodologia envolvendo estas grandezas) para o caso do ponto inicial da adaptação fora da origem, estão apresentados na Tabela 4.12. Foram computados também, assim como feito no caso da estrutura direta e descrito no Item 2.7.1, os valores do produto  $\mu_P \rho_{\max}$  e da velocidade de convergência local do erro de saída,  $\delta_P$  (em dB/iteração). A discussão sobre o papel destas grandezas está no Item 2.2.4. Conforme visto, a velocidade de convergência local além de poder ser determinante para o tempo de convergência do algoritmo de adaptação reflete sua capacidade em rastrear variações temporais do sistema sendo modelado. Os resultados estão na Tabela 4.13.

Para o caso do ponto inicial da adaptação na origem os resultados de  $\mu_P$  e  $n_{60,P}$  estão apresentados na Tabela 4.14. Para este caso são apresentados também gráficos com a adaptação dos coeficientes e a evolução do erro quadrático  $(y(n) - \hat{y}(n))^2$ . O número da figura para cada caso está na Tabela 4.15.

Função	Algoritmo			
	RG		PLR	
	$\mu_P \rho_{\max}$	$\delta_P$	$\mu_P \rho_{\max}$	$\delta_P$
$H_1(z)$	0,18	$7,6 \times 10^{-2}$	0,15	$7,7 \times 10^{-2}$
$H_4(z)$	0,11	$1,7 \times 10^{-3}$	0,18	$3,3 \times 10^{-2}$
$H_6(z)$	0,24	$5,9 \times 10^{-2}$	0,15	$3,8 \times 10^{-2}$
$H_7(z)$	0,24	$1,7 \times 10^{-2}$	0,062	$1,0 \times 10^{-2}$
$H_9(z)$	0,17	$1,5 \times 10^{-3}$	0,3	$5,7 \times 10^{-3}$
$H_{14}(z)$	0,22	$6,9 \times 10^{-2}$	0,2	$7,3 \times 10^{-2}$
$H_{15}(z)$	0,22	$1,0 \times 10^{-2}$	0,047	$1,5 \times 10^{-2}$

Tabela 4.13: Valores de  $\mu_P \rho_{\max}$  e  $\delta$  (decaimento medido do erro de saída, em db/iteração) obtidos das execuções dos algoritmos RG e PLR

Função	Algoritmo					
	RG		PLR		SMM	
	$\mu_P$	$n_{60,P}$	$\mu_P$	$n_{60,P}$	$\mu_P$	$n_{60,P}$
$H_1(z)$	$4 \times 10^{-2}$	$8,8 \times 10^2$	$6 \times 10^{-2}$	$8,3 \times 10^2$	$4 \times 10^{-2}$	$8,8 \times 10^2$
$H_4(z)$	$5,5 \times 10^{-4}$	$4,9 \times 10^4$	$1,5 \times 10^{-2}$	$2,4 \times 10^3$	$8 \times 10^{-4}$	$1,9 \times 10^4$
$H_6(z)$	$5 \times 10^{-2}$	$1,1 \times 10^3$	$6 \times 10^{-2}$	$1,6 \times 10^3$	$6 \times 10^{-2}$	$1,2 \times 10^3$
$H_7(z)$	$8 \times 10^{-3}$	$4,5 \times 10^4$	$2 \times 10^{-2}$	$4,2 \times 10^4$	$8 \times 10^{-3}$	$1,5 \times 10^4$
$H_9(z)$	$2 \times 10^{-2}$	$2,9 \times 10^4$	$7 \times 10^{-2}$	$9,4 \times 10^3$	$2 \times 10^{-2}$	$2,6 \times 10^4$
$H_{14}(z)$	$4 \times 10^{-2}$	$1,7 \times 10^3$	$8 \times 10^{-2}$	$1,7 \times 10^3$	$4 \times 10^{-2}$	$1,4 \times 10^3$
$H_{15}(z)$	$4 \times 10^{-3}$	$1,6 \times 10^5$	$1,5 \times 10^{-2}$	$1,2 \times 10^5$	$4 \times 10^{-3}$	$2,3 \times 10^4$

Tabela 4.14: Resultados da execução dos algoritmos RG, PLR e SMM para a estrutura polifásica,  $p = 4$ , com início na origem;  $\mu$ : passo de adaptação máximo;  $n_{60}$ : número de iterações para erro quadrático de saída ficar 60 dB abaixo da potência de saída.

Função	Número da figura		
	RG	PLR	SMM
$H_1(z)$	4-10	4-11	4-12
$H_4(z)$	4-13	4-14	4-15
$H_6(z)$	4-16	4-17	4-18
$H_7(z)$	4-19	4-20	4-21
$H_9(z)$	4-22	4-23	4-24
$H_{14}(z)$	4-25	4-26	4-27
$H_{15}(z)$	4-28	4-29	4-30

Tabela 4.15: Número da figura com gráficos da execução dos algoritmos para cada caso de  $H(z)$

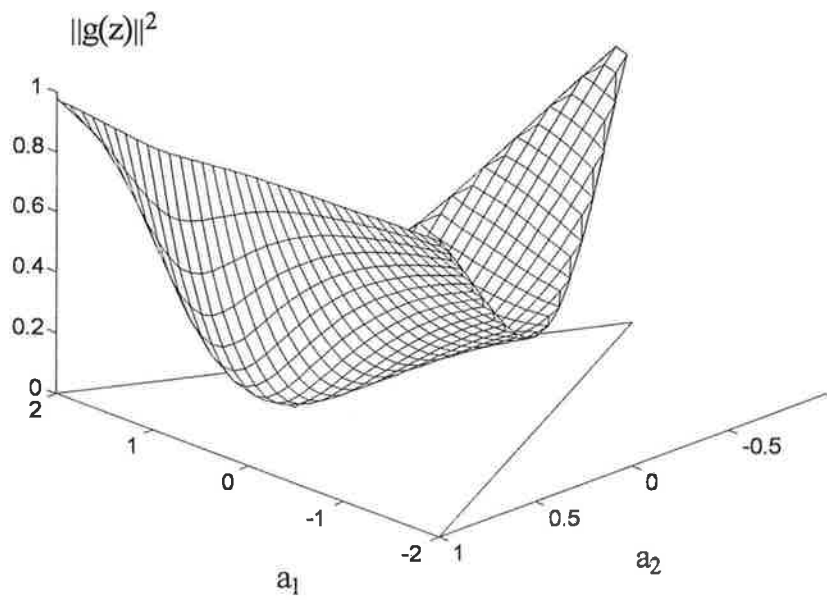
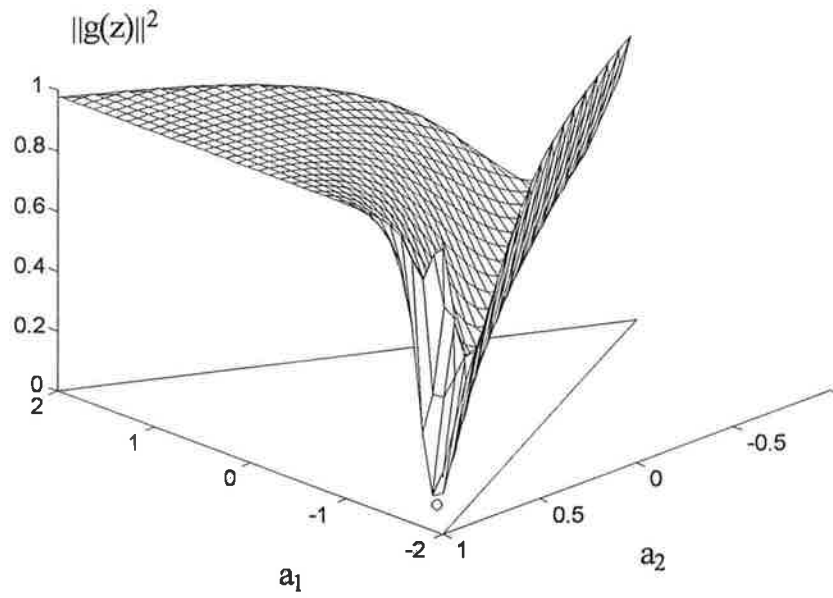


Figura 4-6: Superfície de erro reduzida para  $H_1(z)$ : estruturas direta (em cima) e polifásica,  $p = 4$  (em baixo).

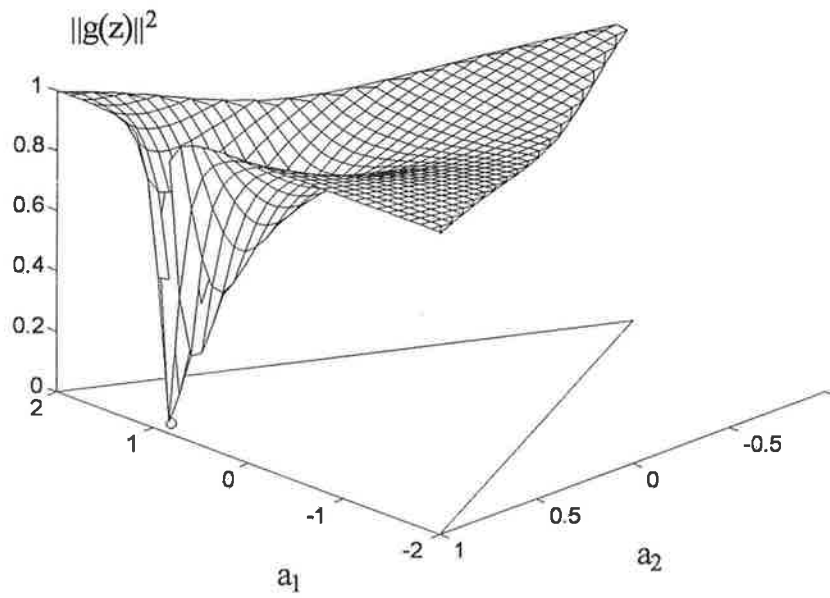
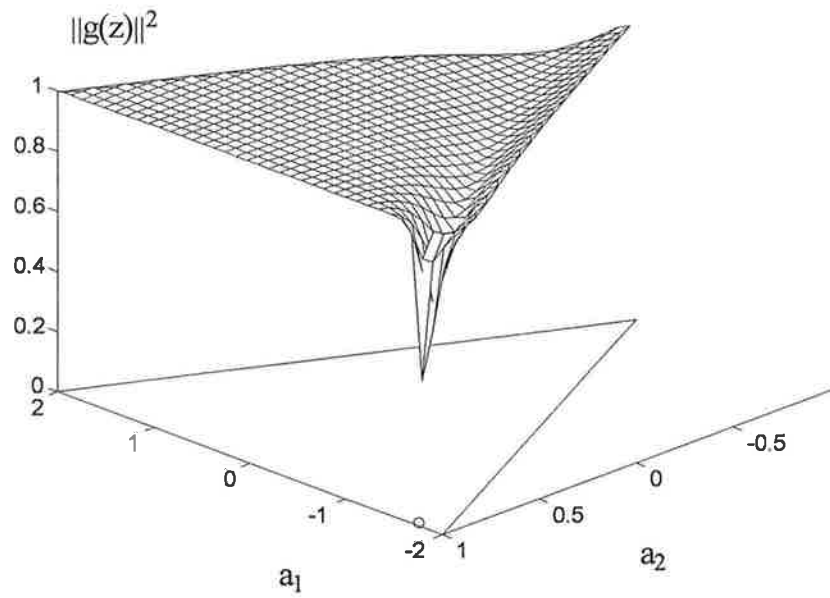


Figura 4-7: Superfície de erro reduzida para  $H_4(z)$ : estruturas direta (em cima) e polifásica,  $p = 4$  (em baixo).

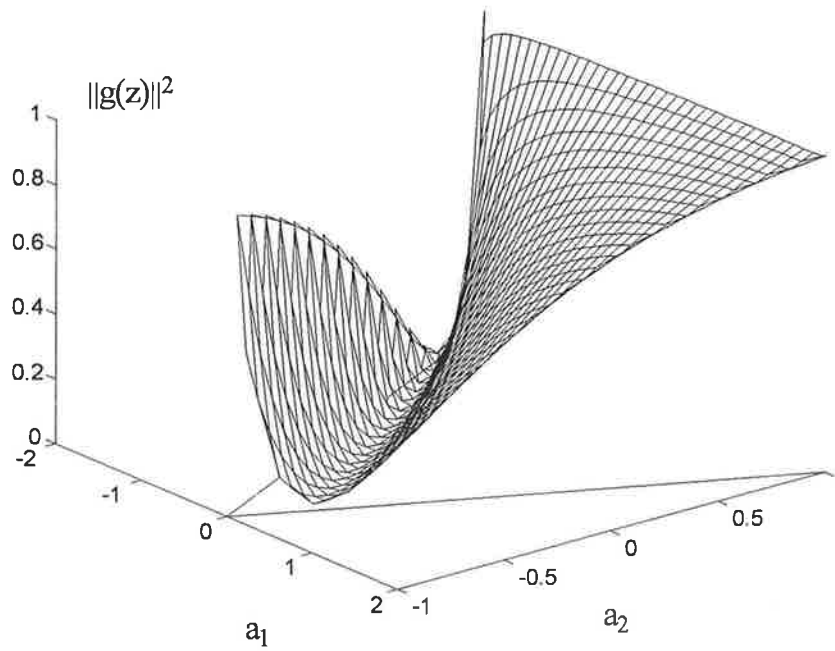
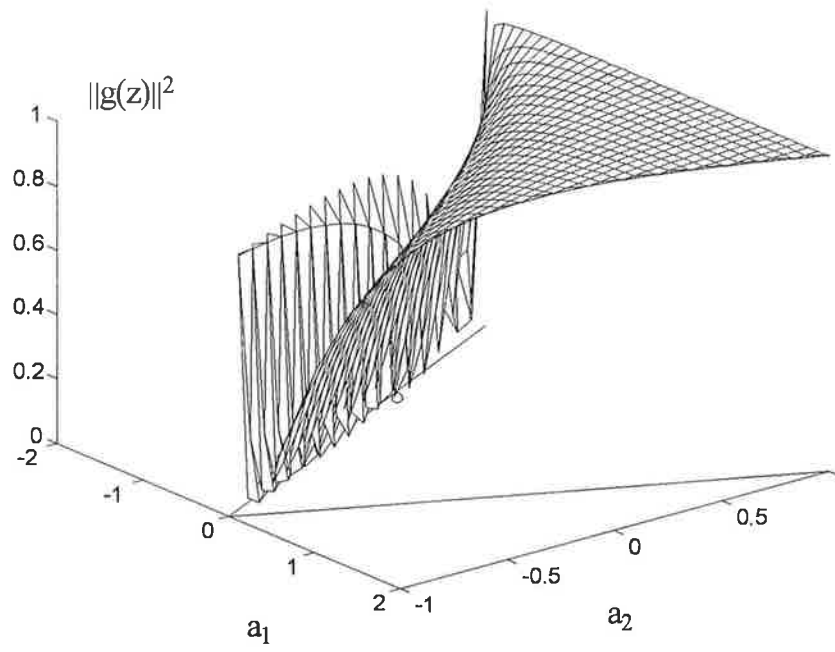


Figura 4-8: Superfície de erro reduzida para  $H_9(z)$ : estruturas direta (em cima) e polifásica,  $p = 4$  (em baixo).



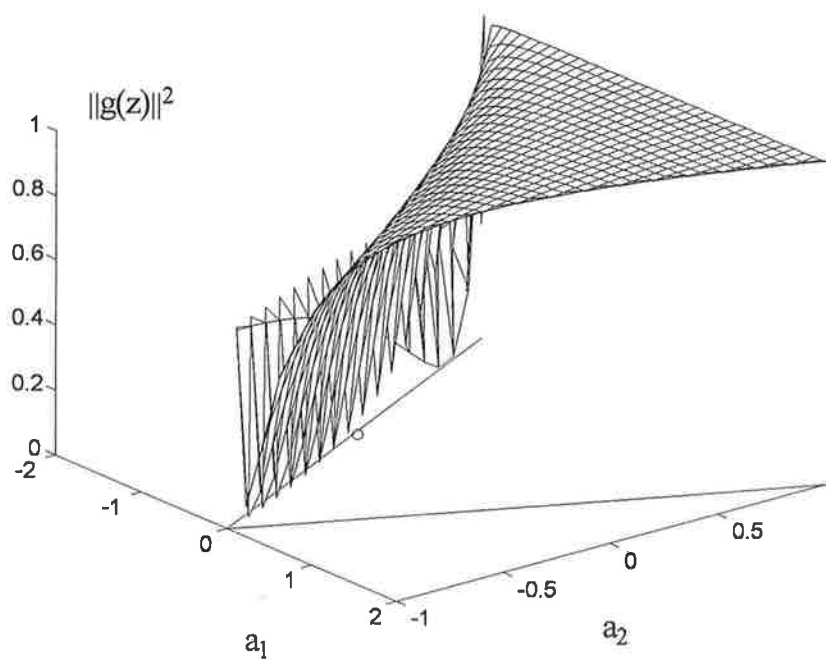
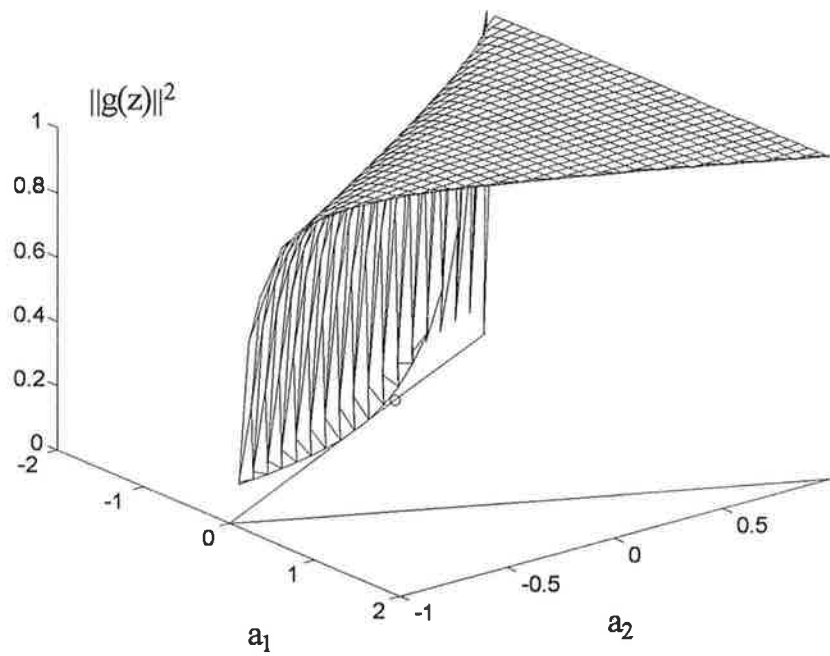


Figura 4-9: Superfície de erro reduzida para  $H_{12}(z)$ : estruturas direta (em cima) e polifásica,  $p = 4$  (em baixo).

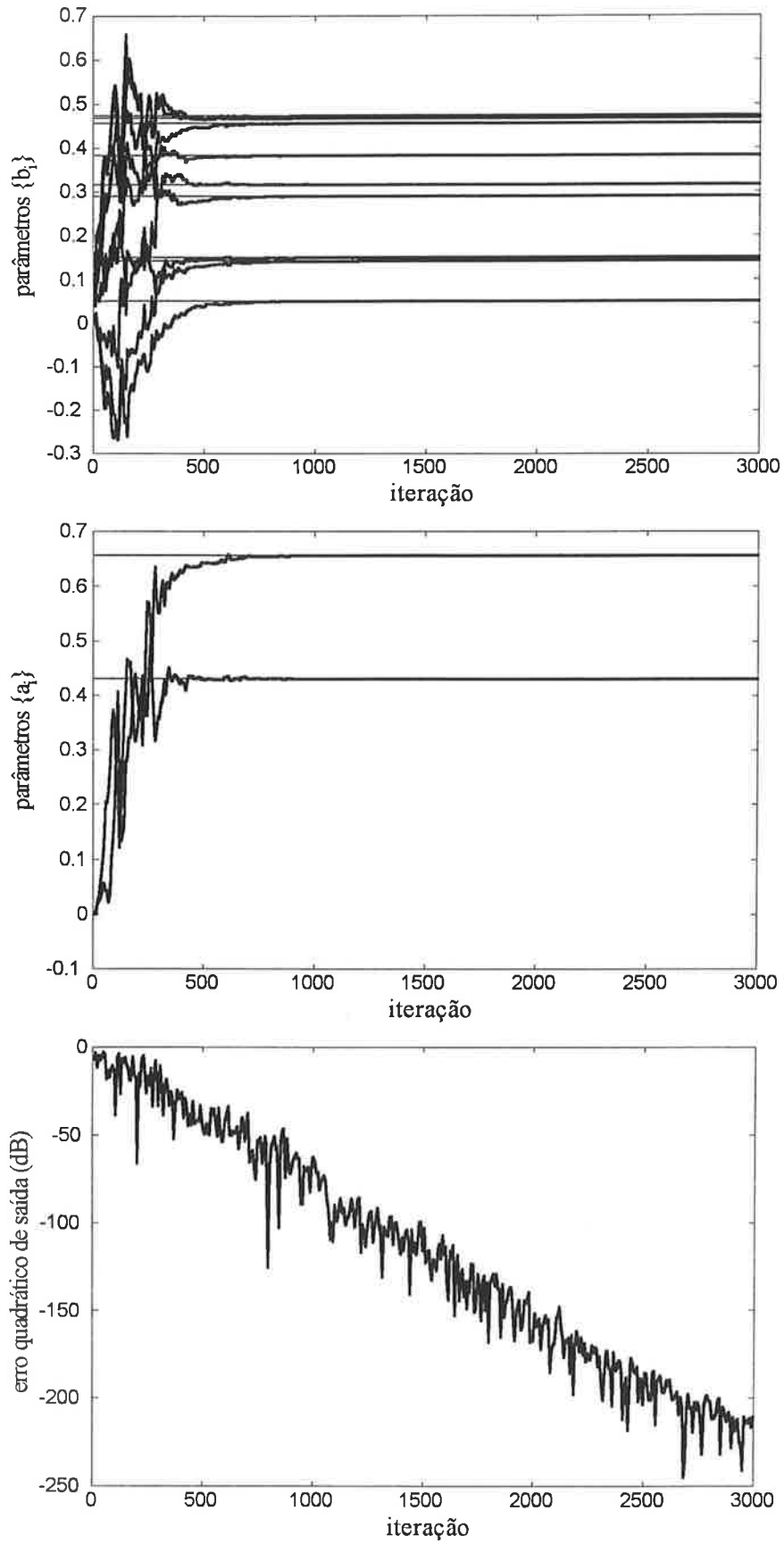


Figura 4-10: Filtro polifásico, algoritmo RG, sistema  $H_1(z)$  (parâmetros em traços claros)

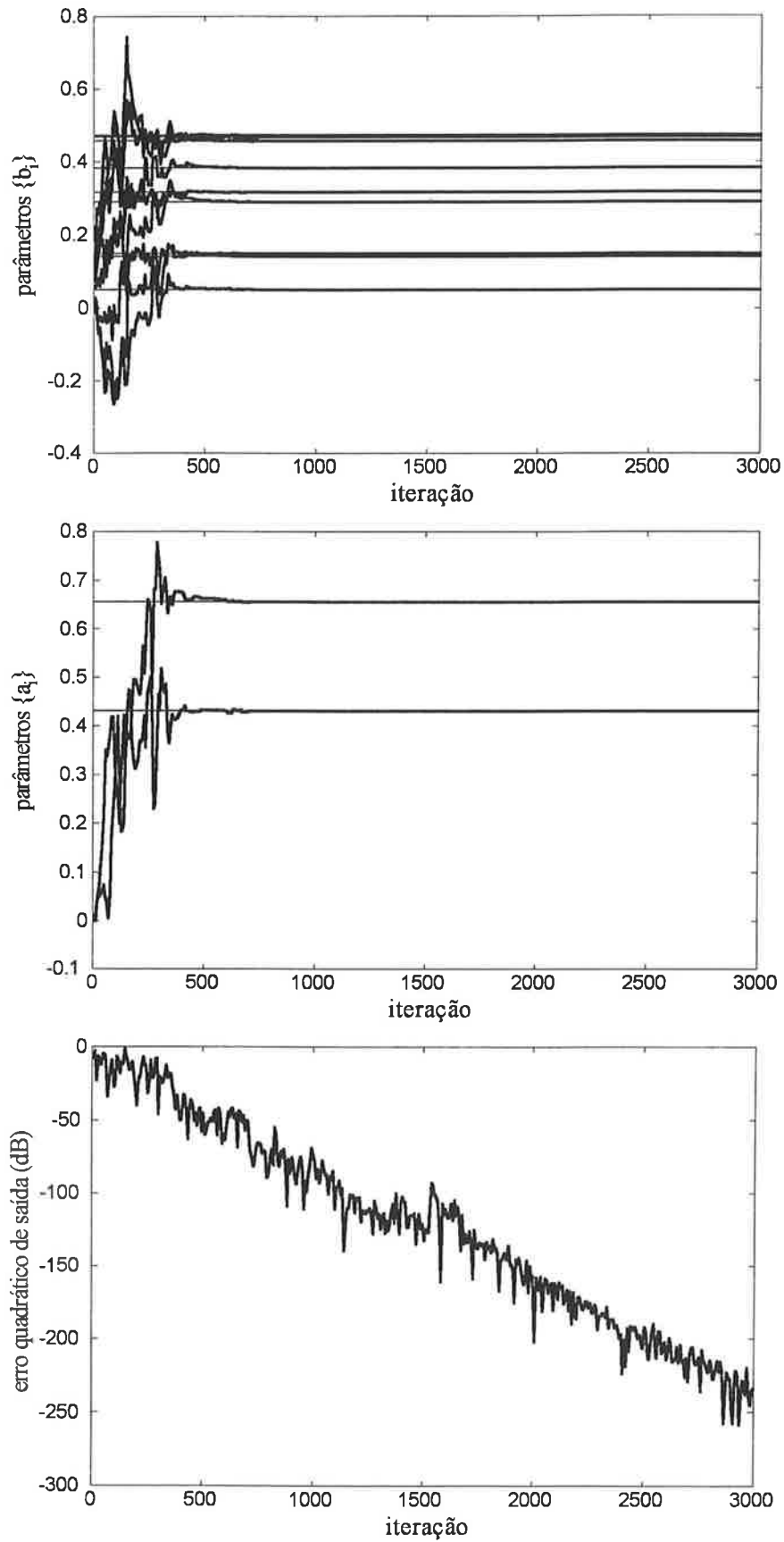


Figura 4-11: Filtro polifásico, algoritmo PLR, sistema  $H_1(z)$  (parâmetros em traços claros)

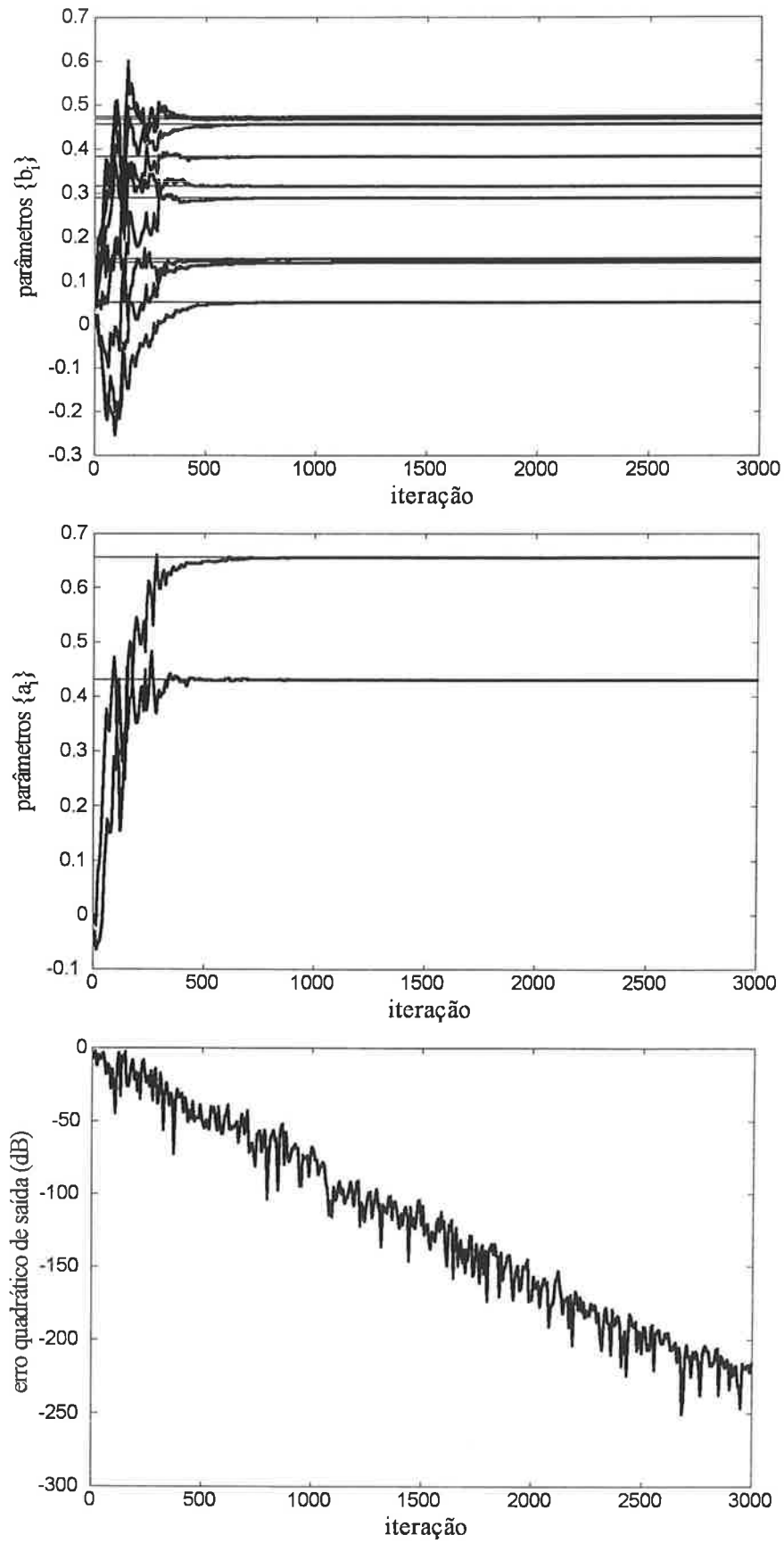


Figura 4-12: Filtro polifásico, algoritmo SMM, sistema  $H_1(z)$  (parâmetros em traços claros)

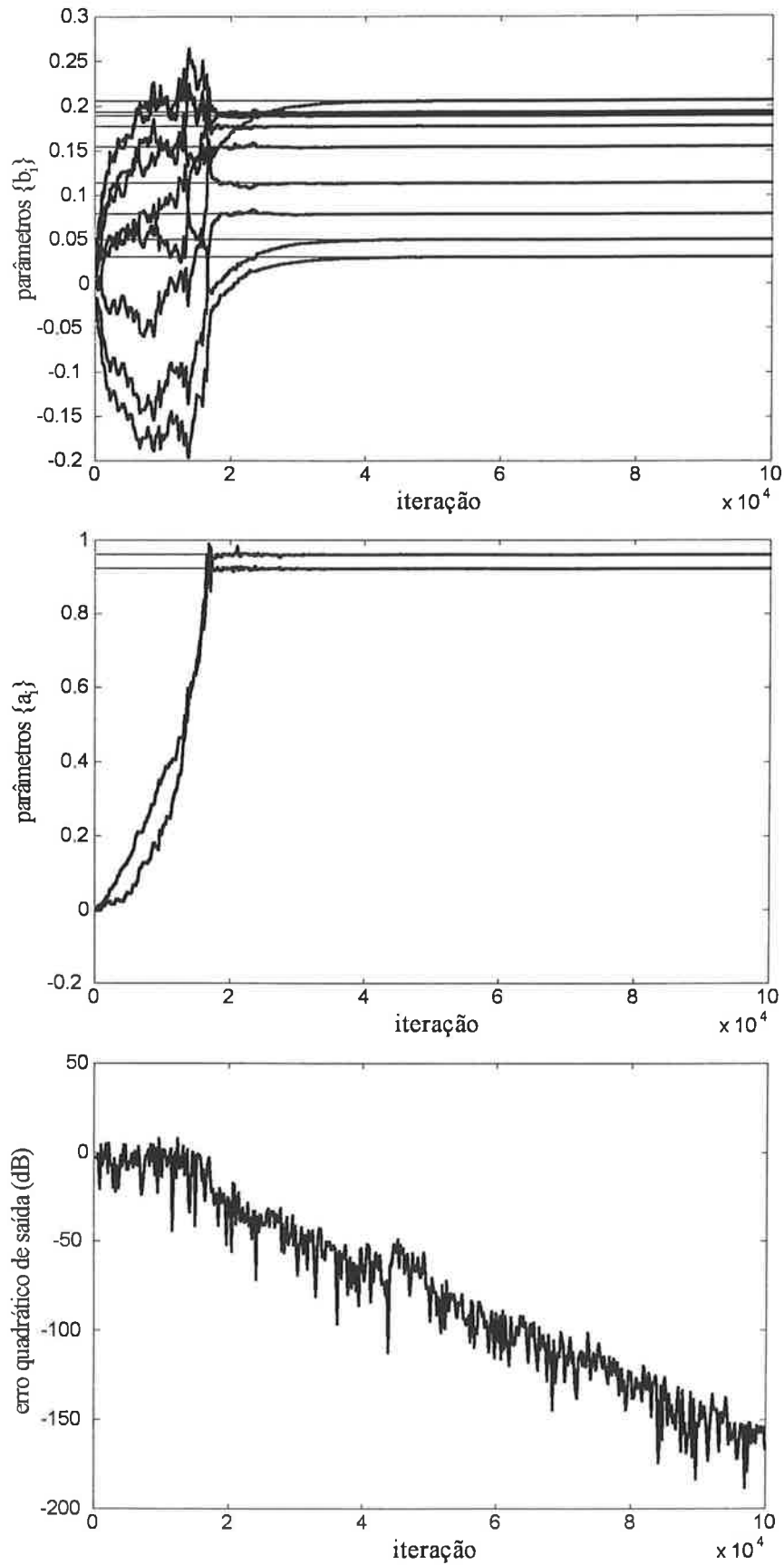


Figura 4-13: Filtro polifásico, algoritmo RG, sistema  $H_4(z)$  (parâmetros em traços claros)

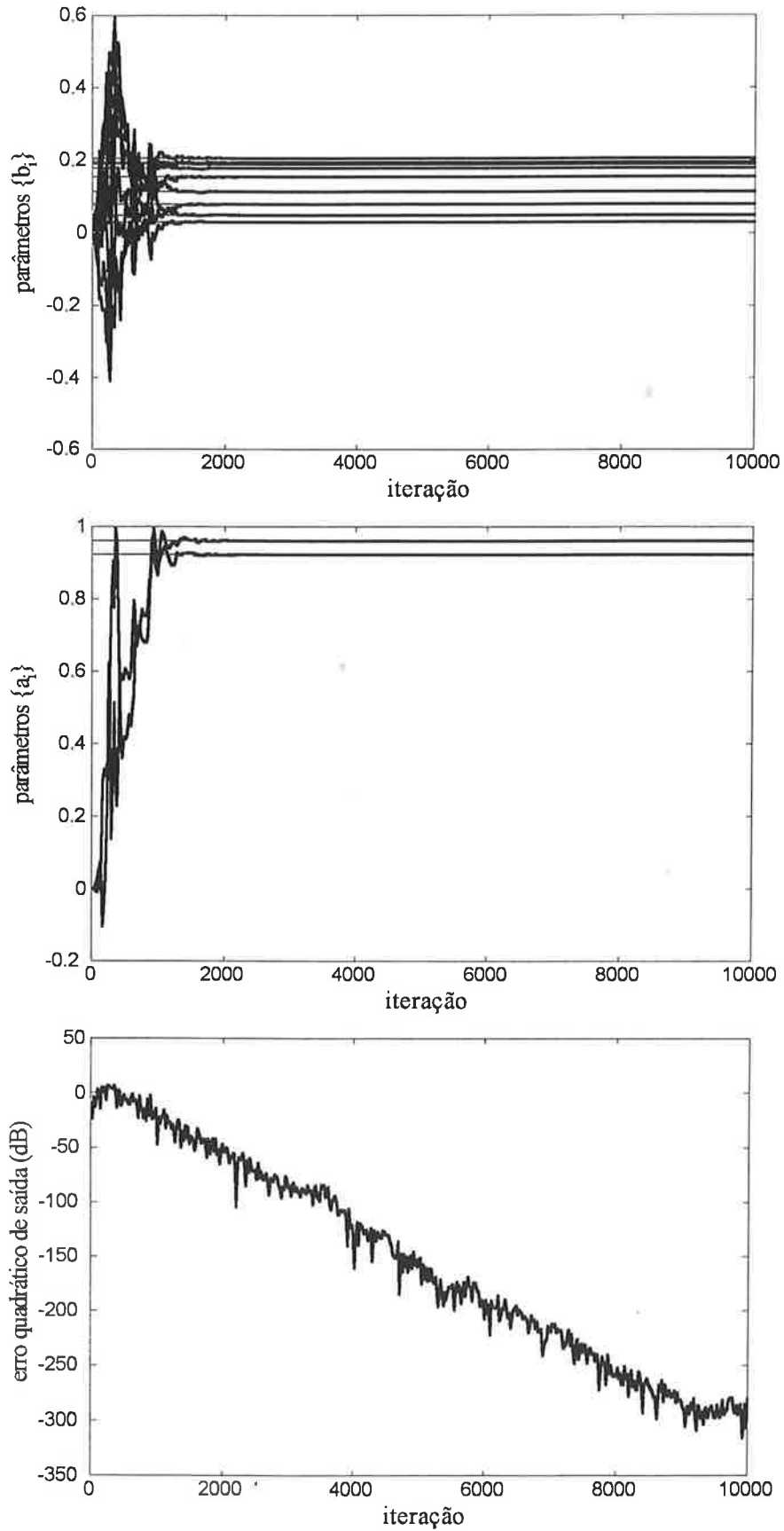


Figura 4-14: Filtro polifásico, algoritmo PLR, sistema  $H_4(z)$  (parâmetros em traços claros)

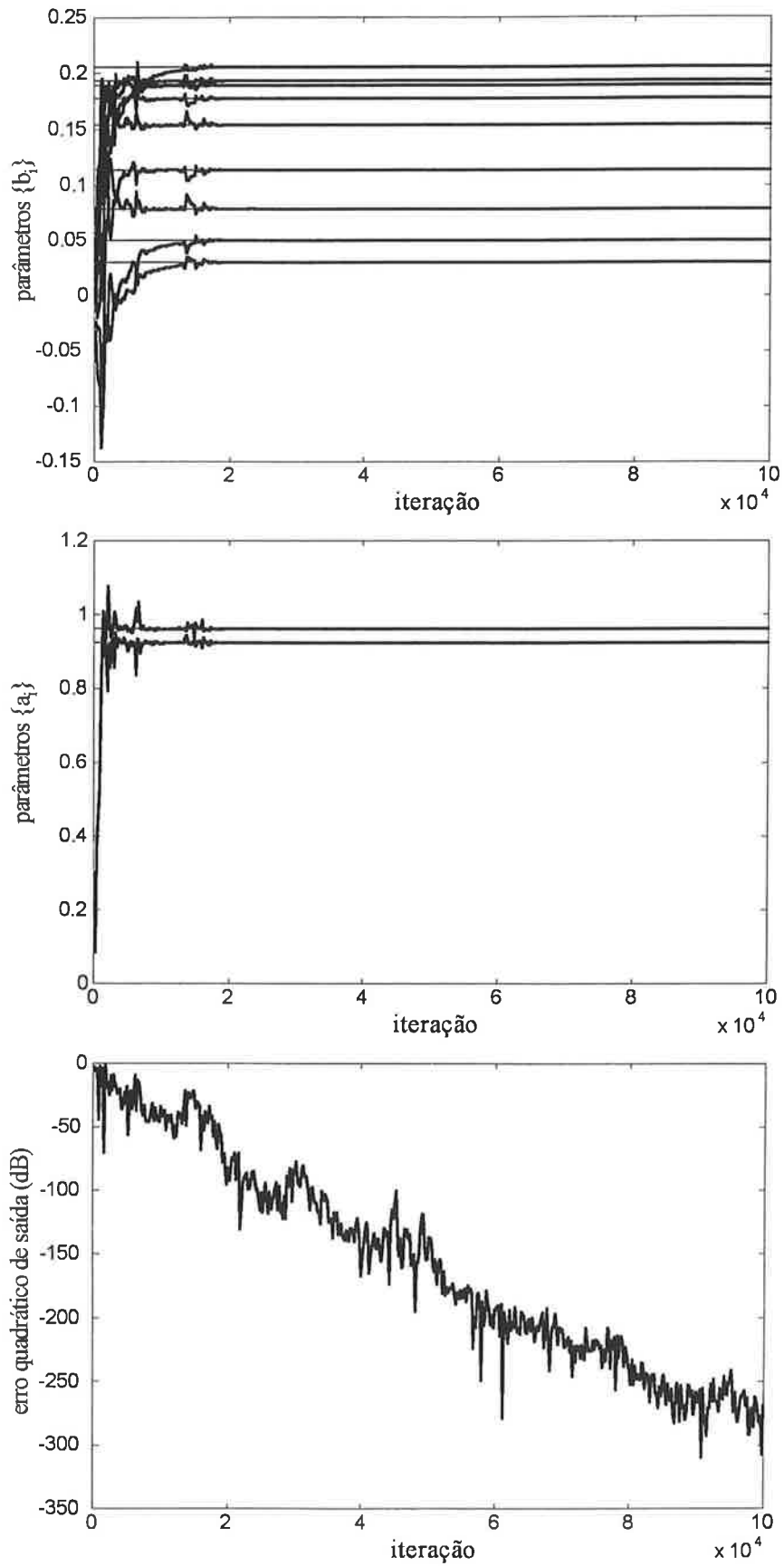


Figura 4-15: Filtro polifásico, algoritmo SMM, sistema  $H_4(z)$  (parâmetros em traços claros)

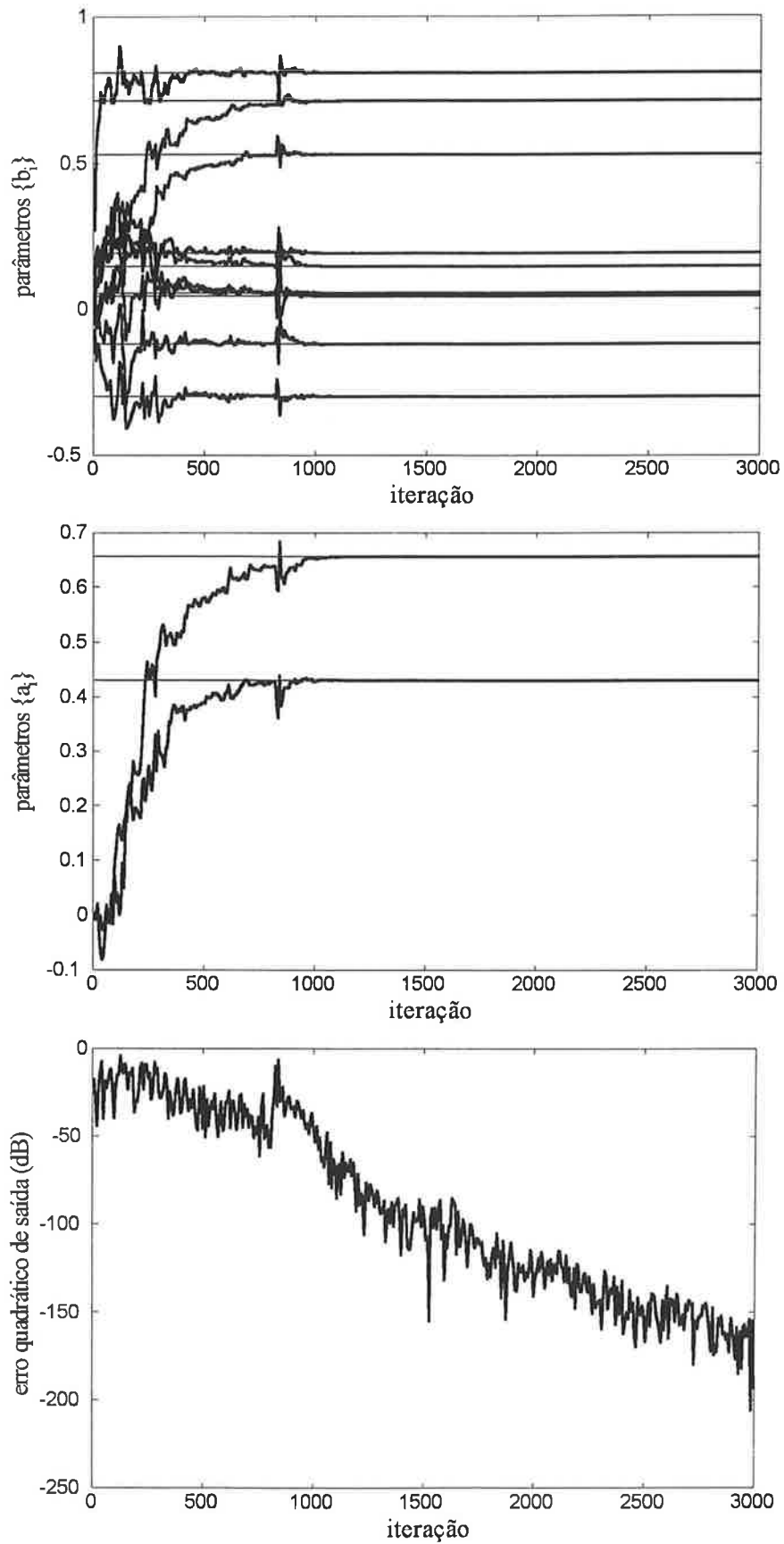


Figura 4-16: Filtro polifásico, algoritmo RG, sistema  $H_6(z)$  (parâmetros em traços claros)



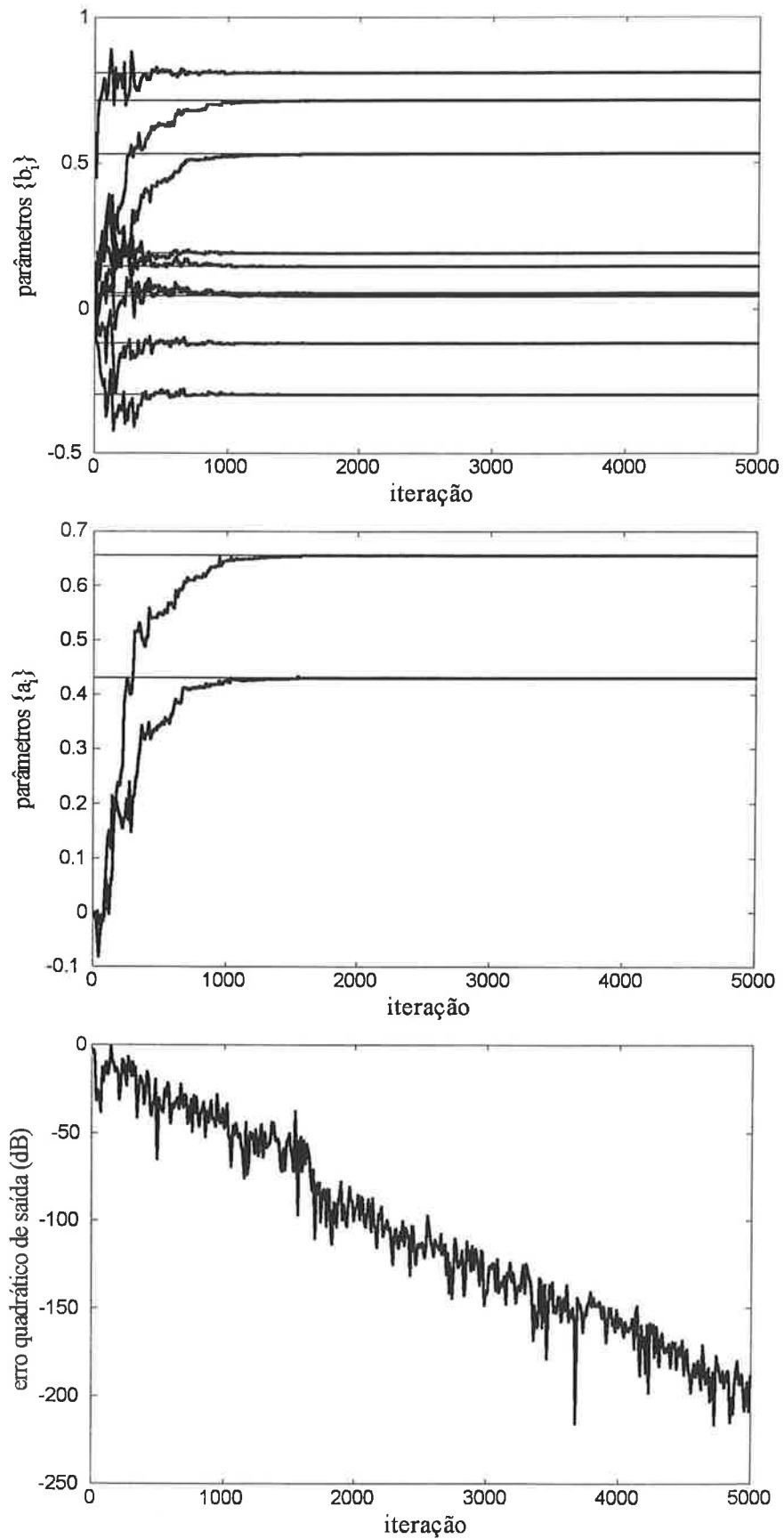


Figura 4-17: Filtro polifásico, algoritmo PLR, sistema  $H_6(z)$  (parâmetros em traços claros)

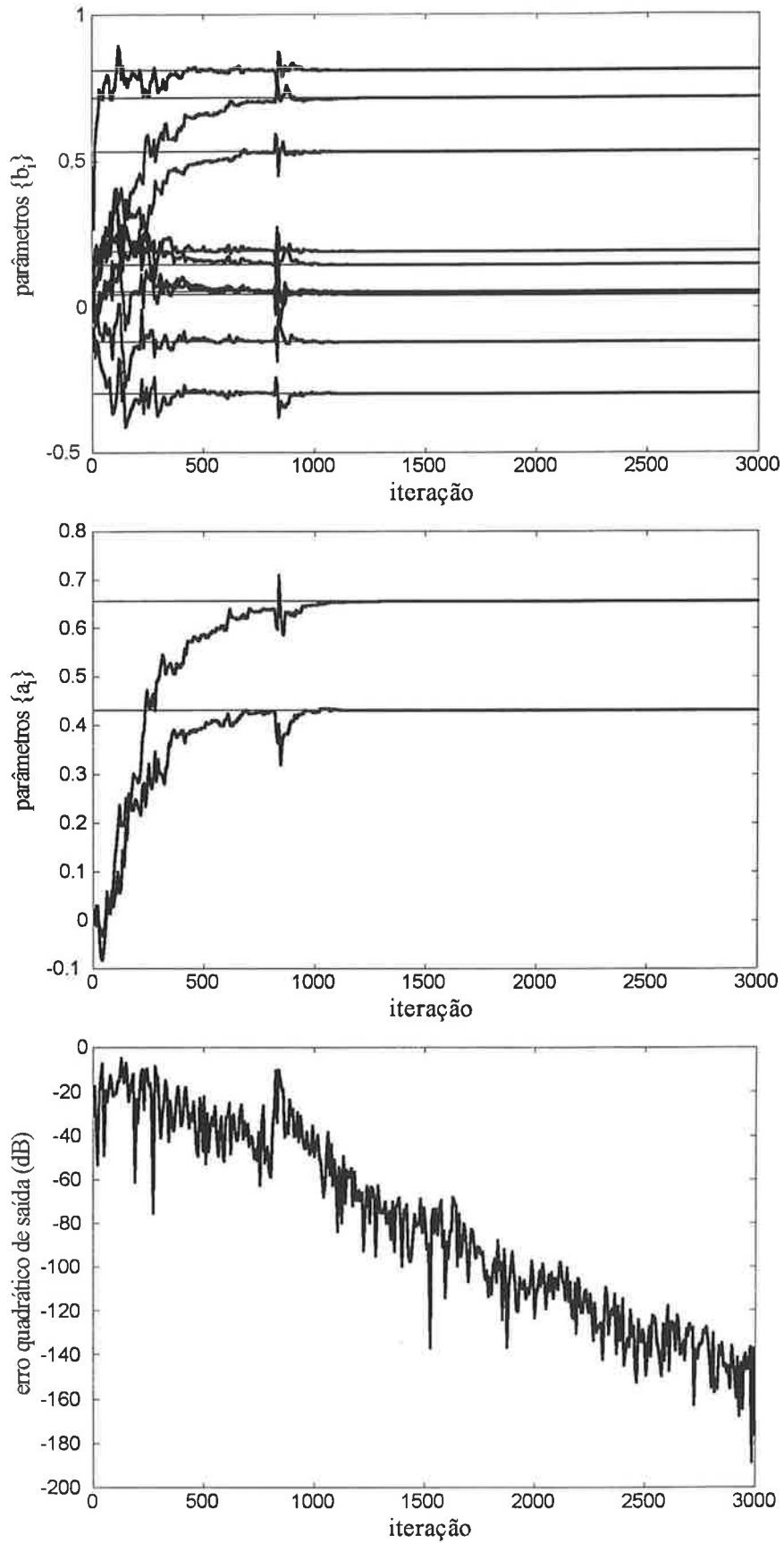


Figura 4-18: Filtro polifásico, algoritmo SMM, sistema  $H_6(z)$  (parâmetros em traços claros)

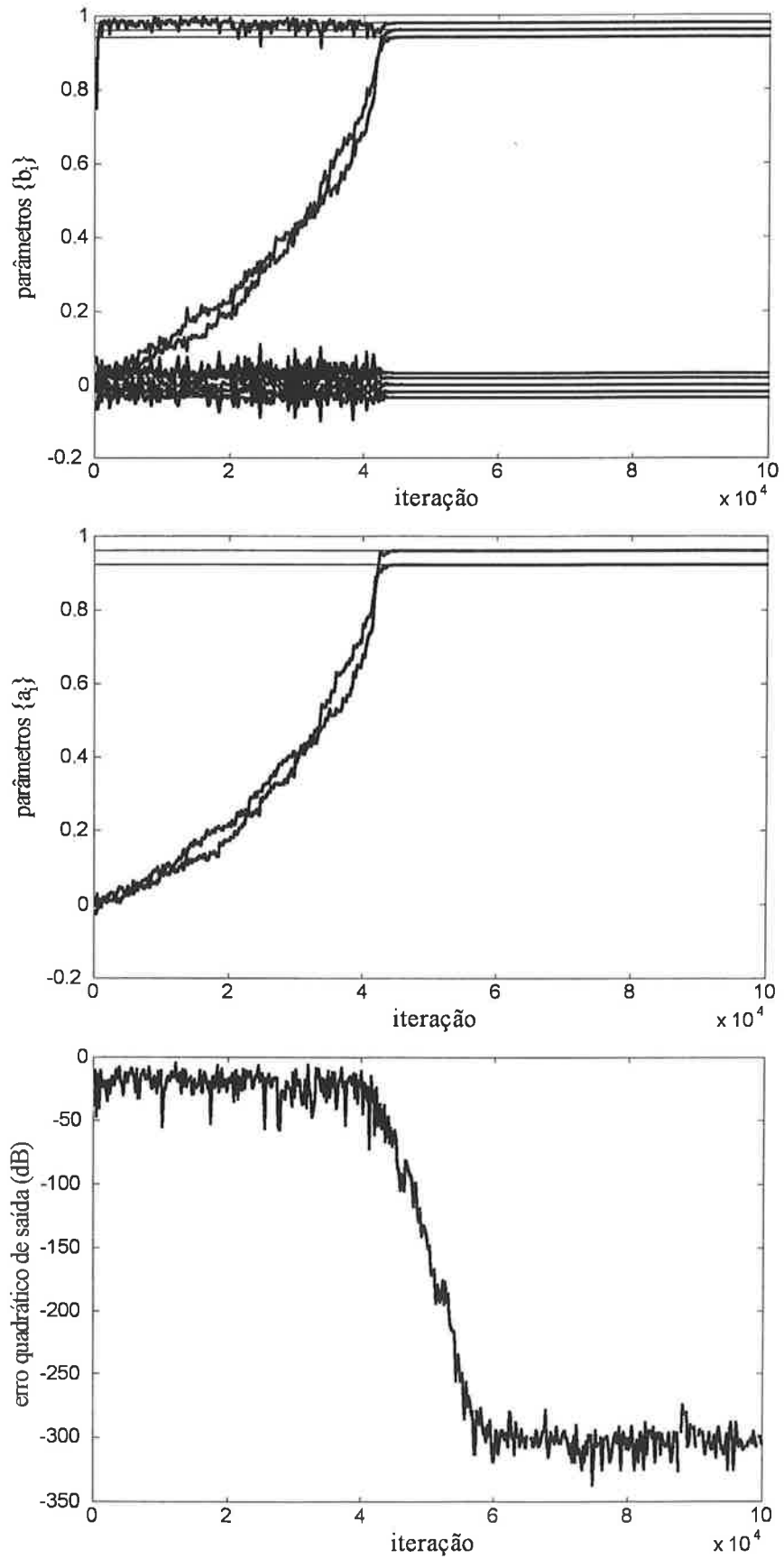


Figura 4-19: Filtro polifásico, algoritmo RG, sistema  $H_7(z)$  (parâmetros em traços claros)

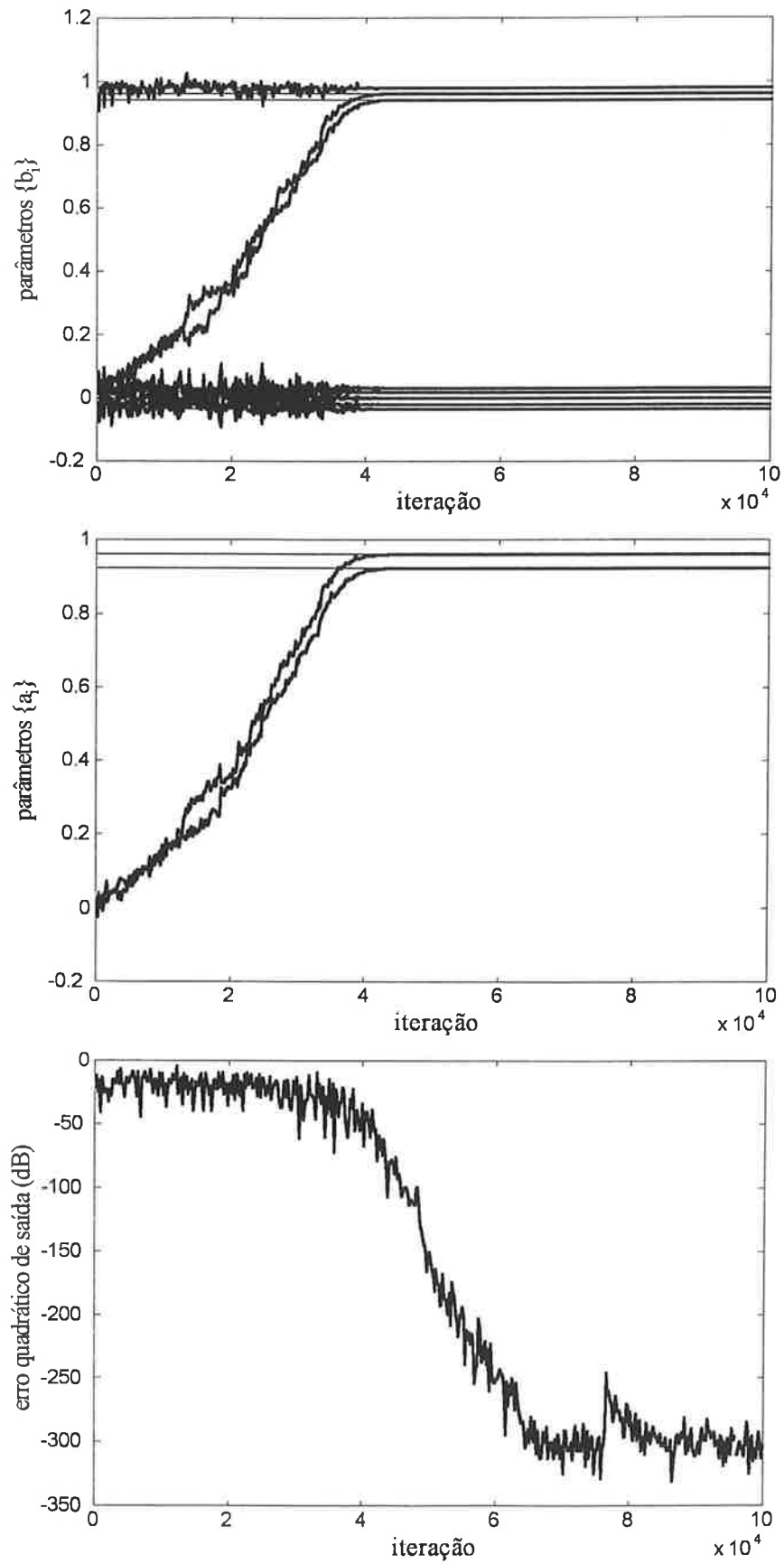


Figura 4-20: Filtro polifásico, algoritmo PLR, sistema  $H_7(z)$  (parâmetros em traços claros)

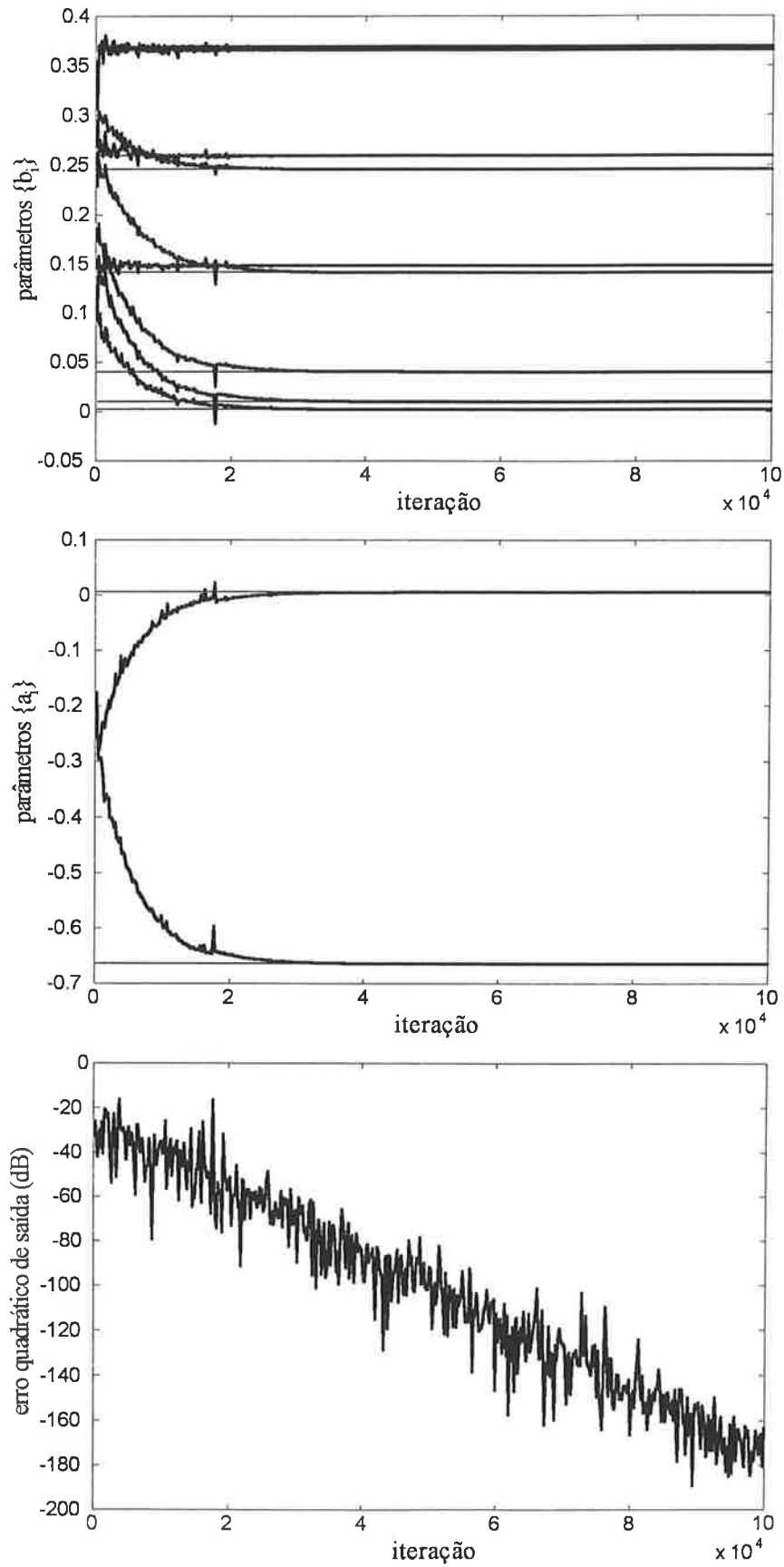


Figura 4-22: Filtro polifásico, algoritmo RG, sistema  $H_9(z)$  (parâmetros em traços claros)

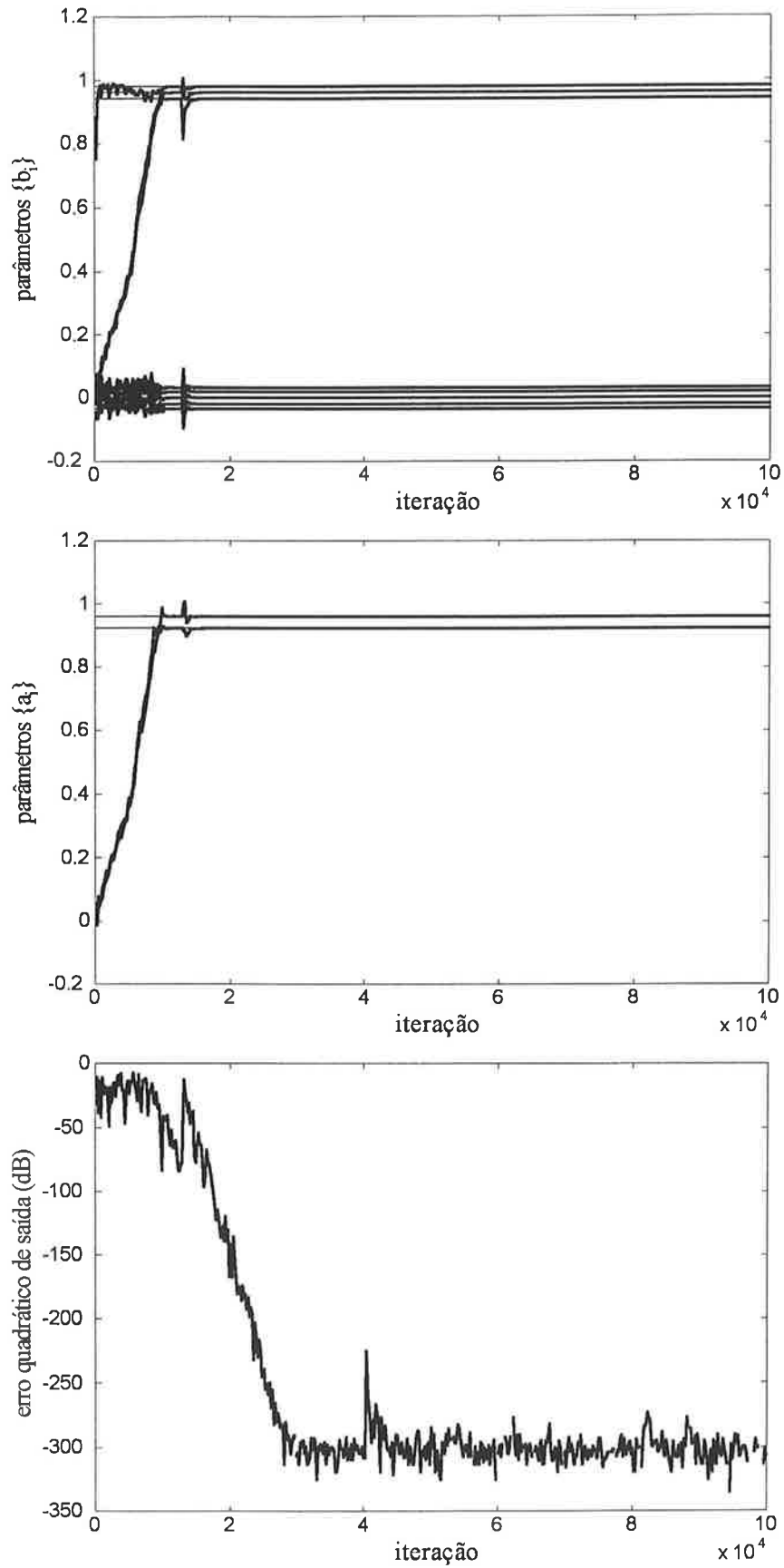


Figura 4-21: Filtro polifásico, algoritmo SMM, sistema  $H_7(z)$  (parâmetros em traços claros)

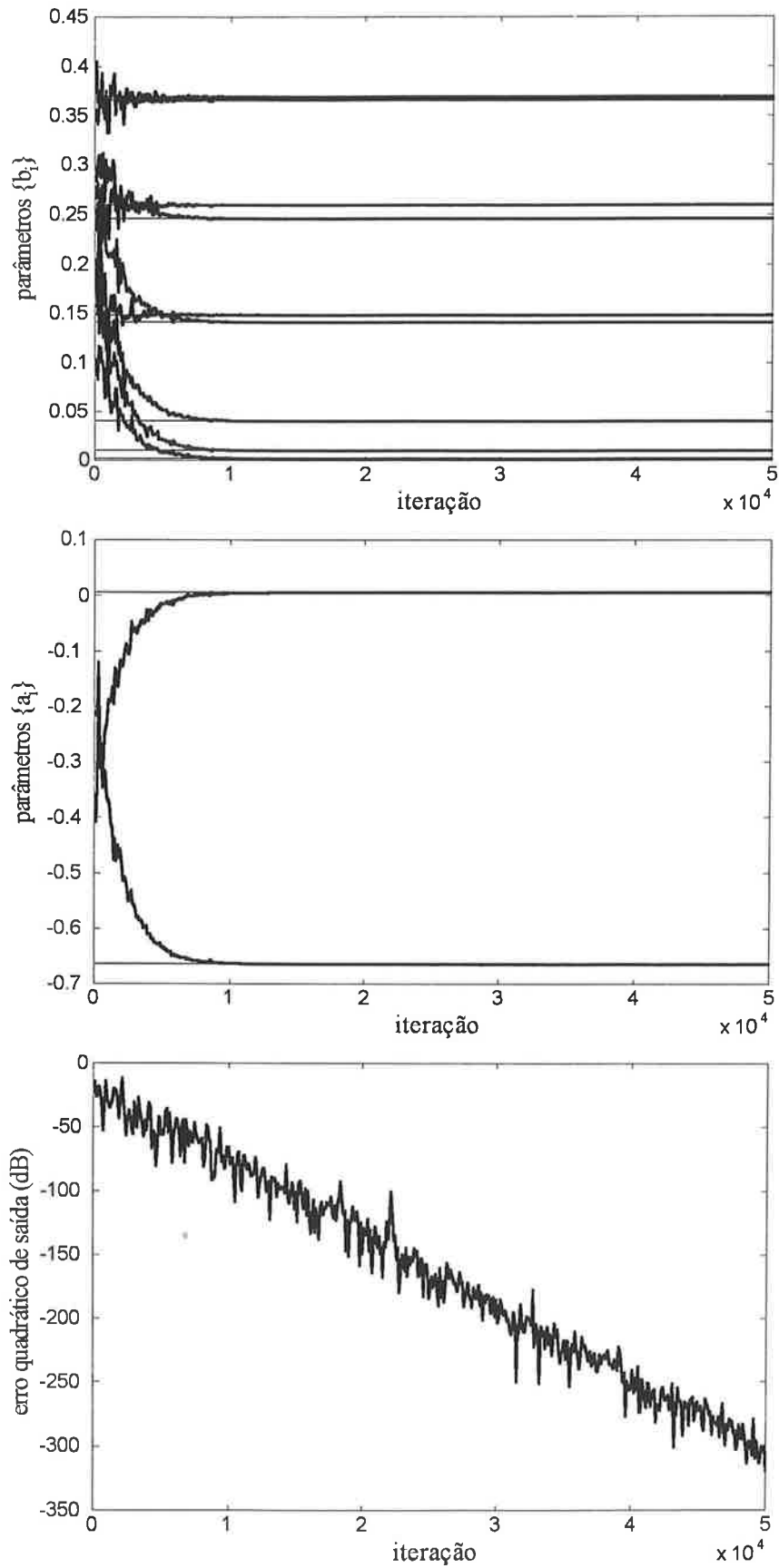


Figura 4-23: Filtro polifásico, algoritmo PLR, sistema  $H_9(z)$  (parâmetros em traços claros)

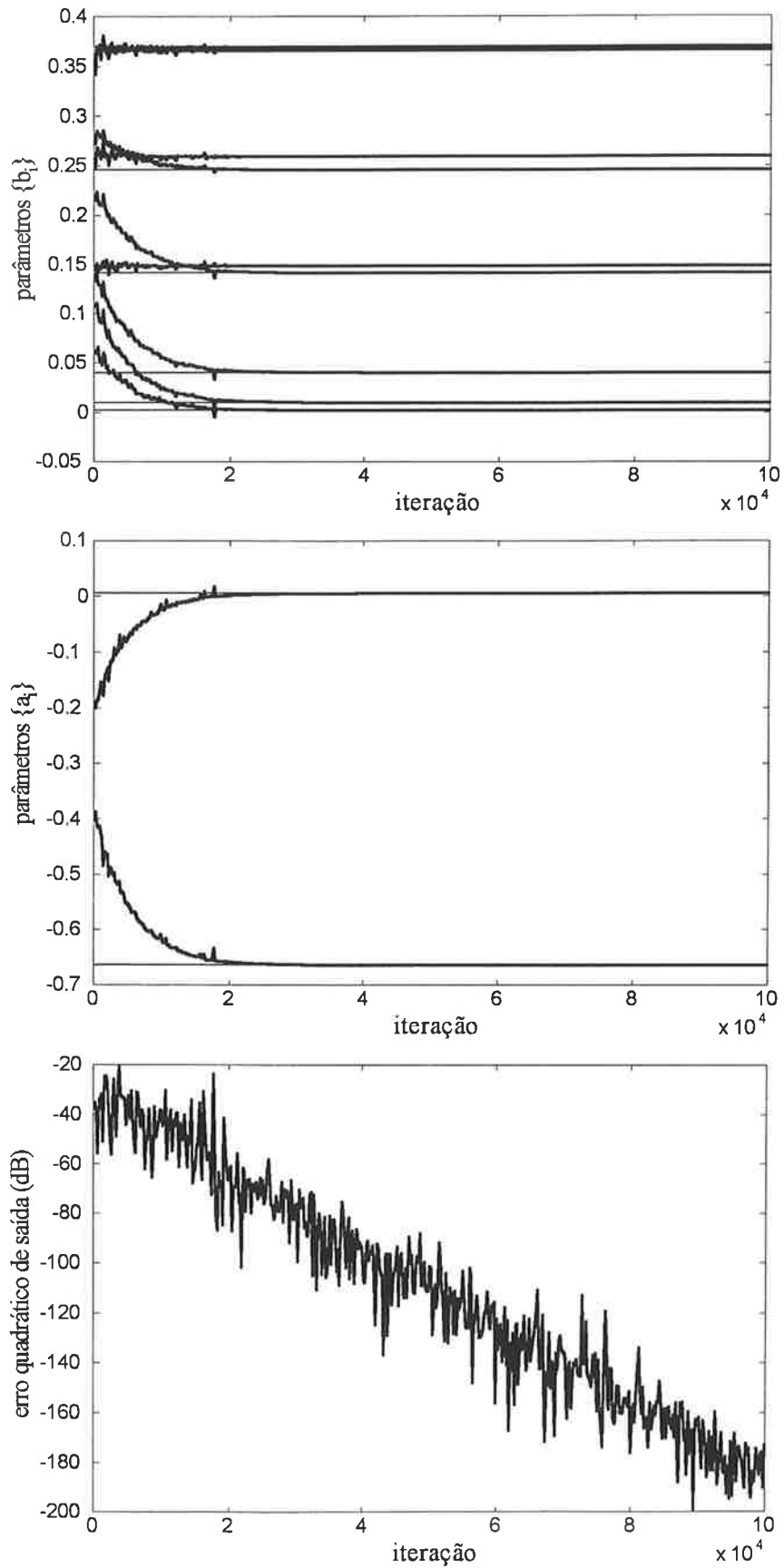


Figura 4-24: Filtro polifásico, algoritmo SMM, sistema  $H_9(z)$  (parâmetros em traços claros)



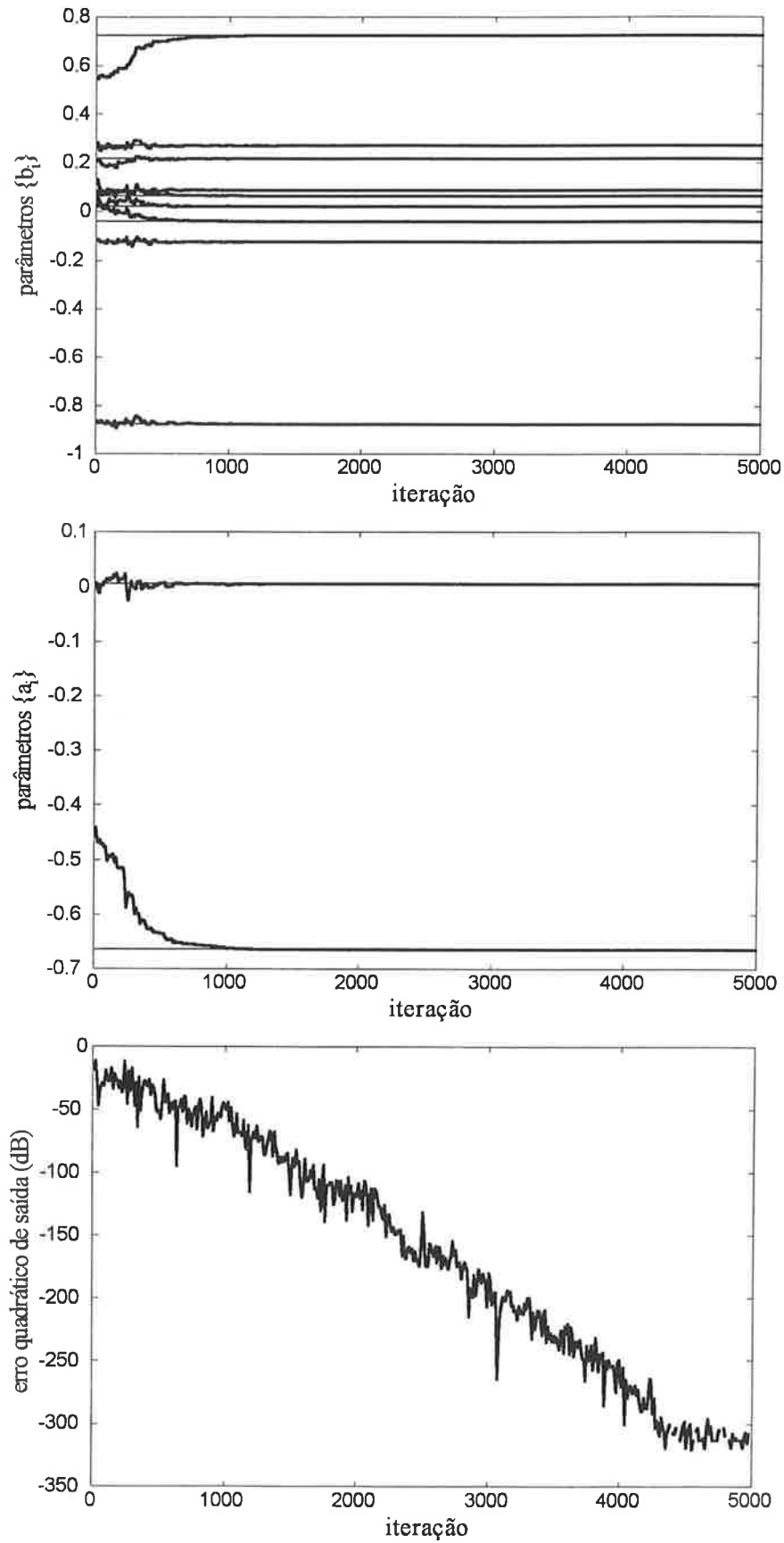


Figura 4-25: Filtro polifásico, algoritmo RG, sistema  $H_{14}(z)$  (parâmetros em traços claros)

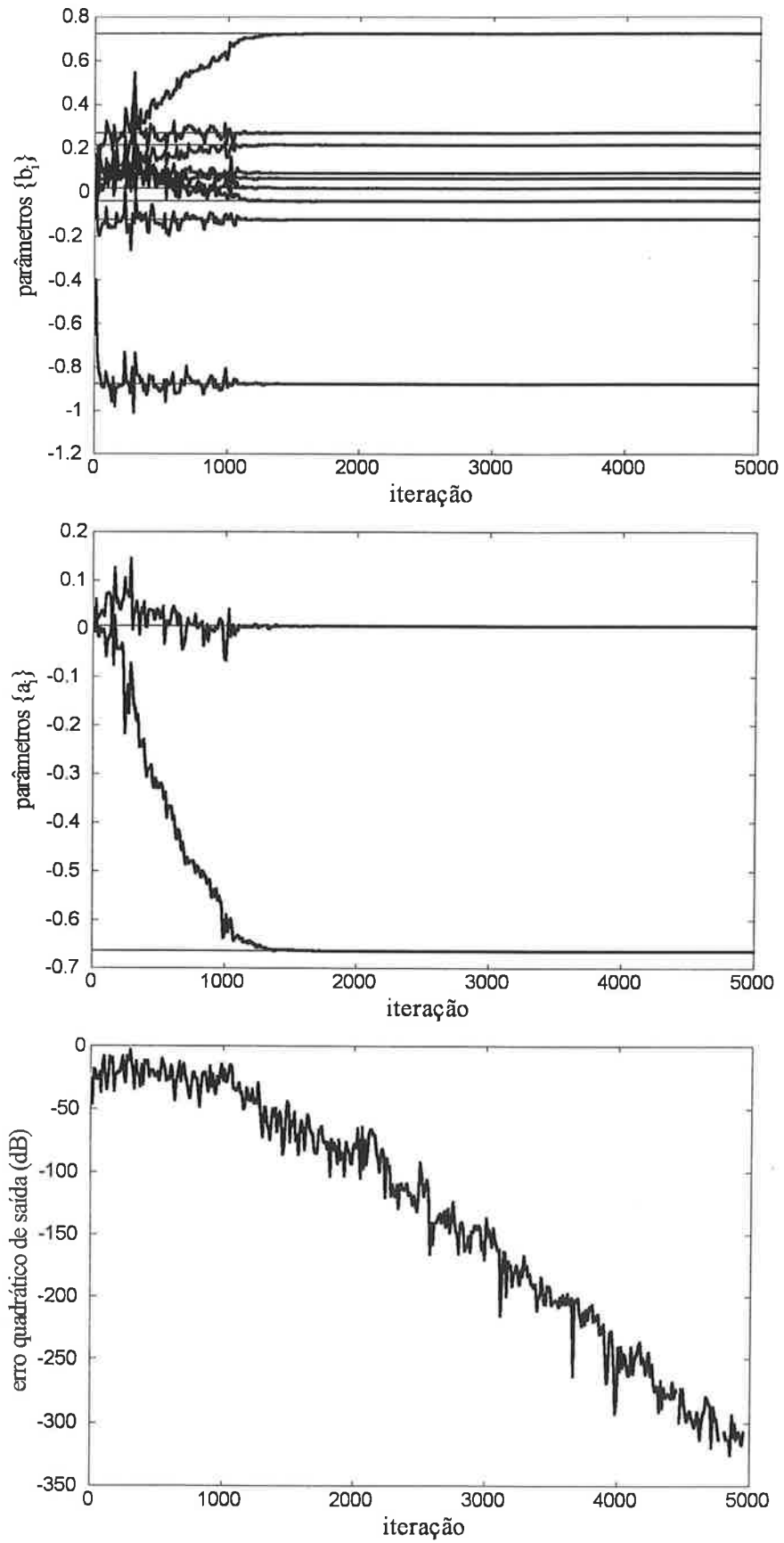


Figura 4-26: Filtro polifásico, algoritmo PLR, sistema  $H_{14}(z)$  (parâmetros em traços claros)

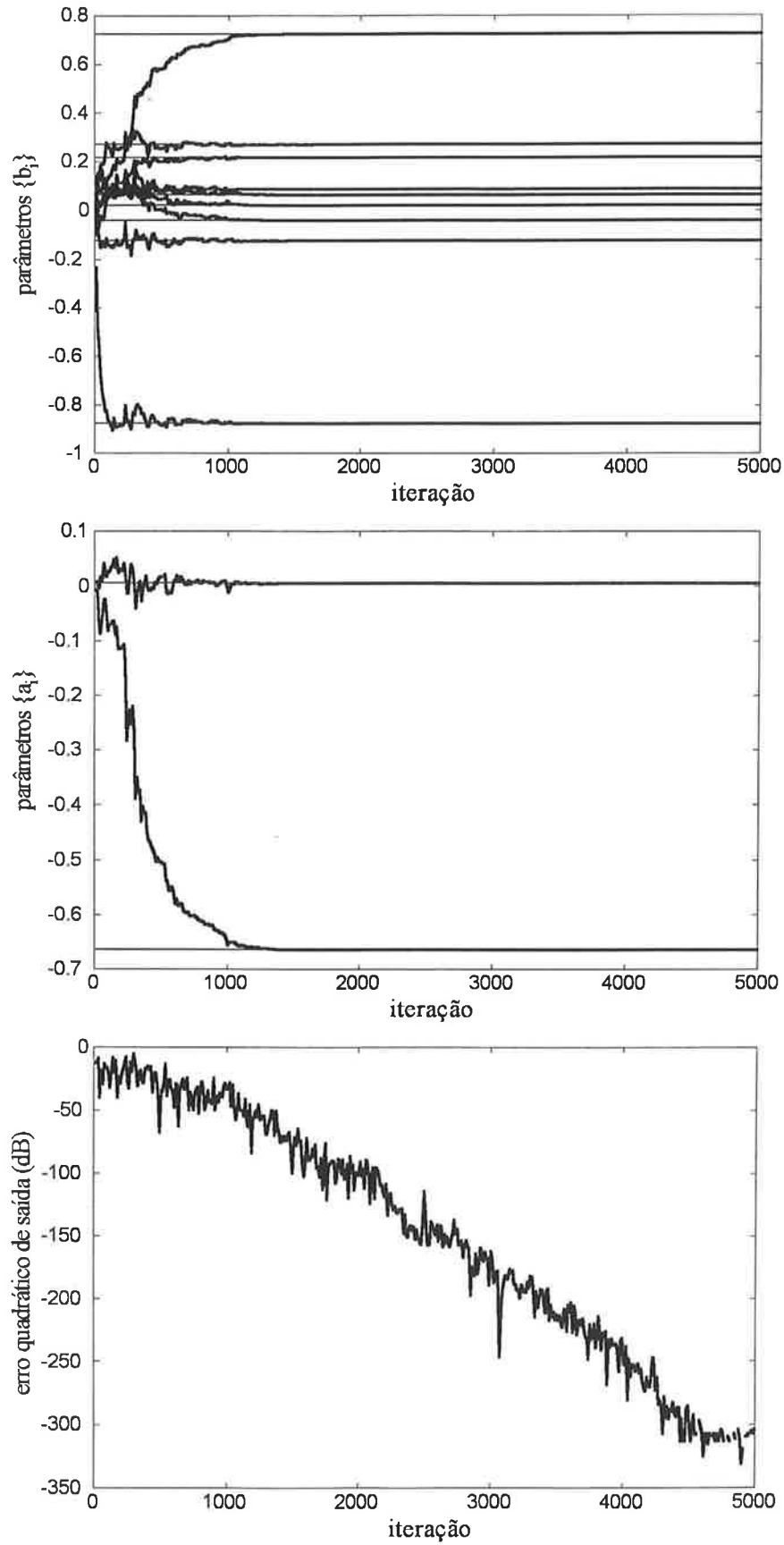


Figura 4-27: Filtro polifásico, algoritmo SMM, sistema  $H_{14}(z)$  (parâmetros em traços claros)

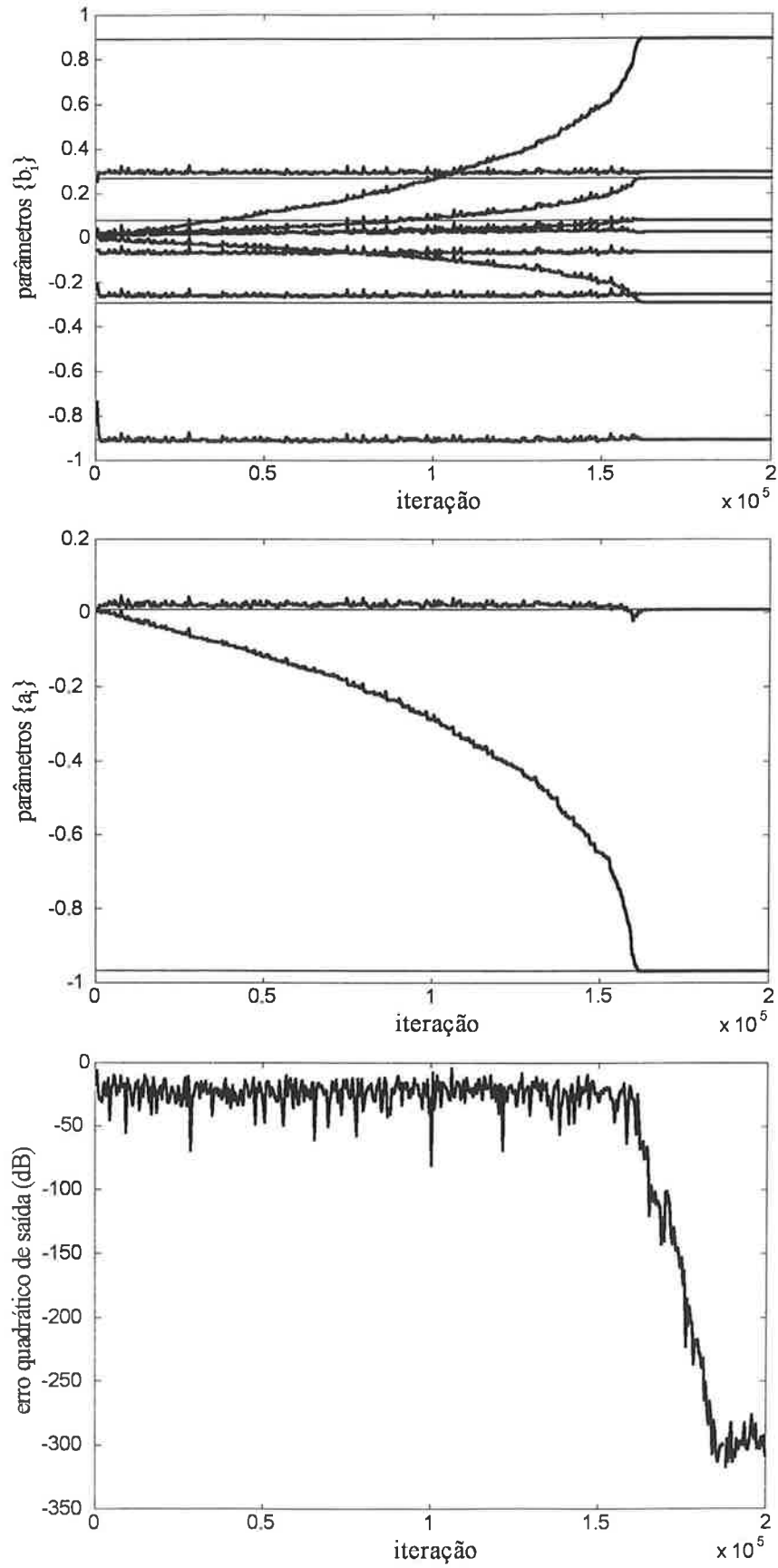


Figura 4-28: Filtro polifásico, algoritmo RG, sistema  $H_{15}(z)$  (parâmetros em traços claros)

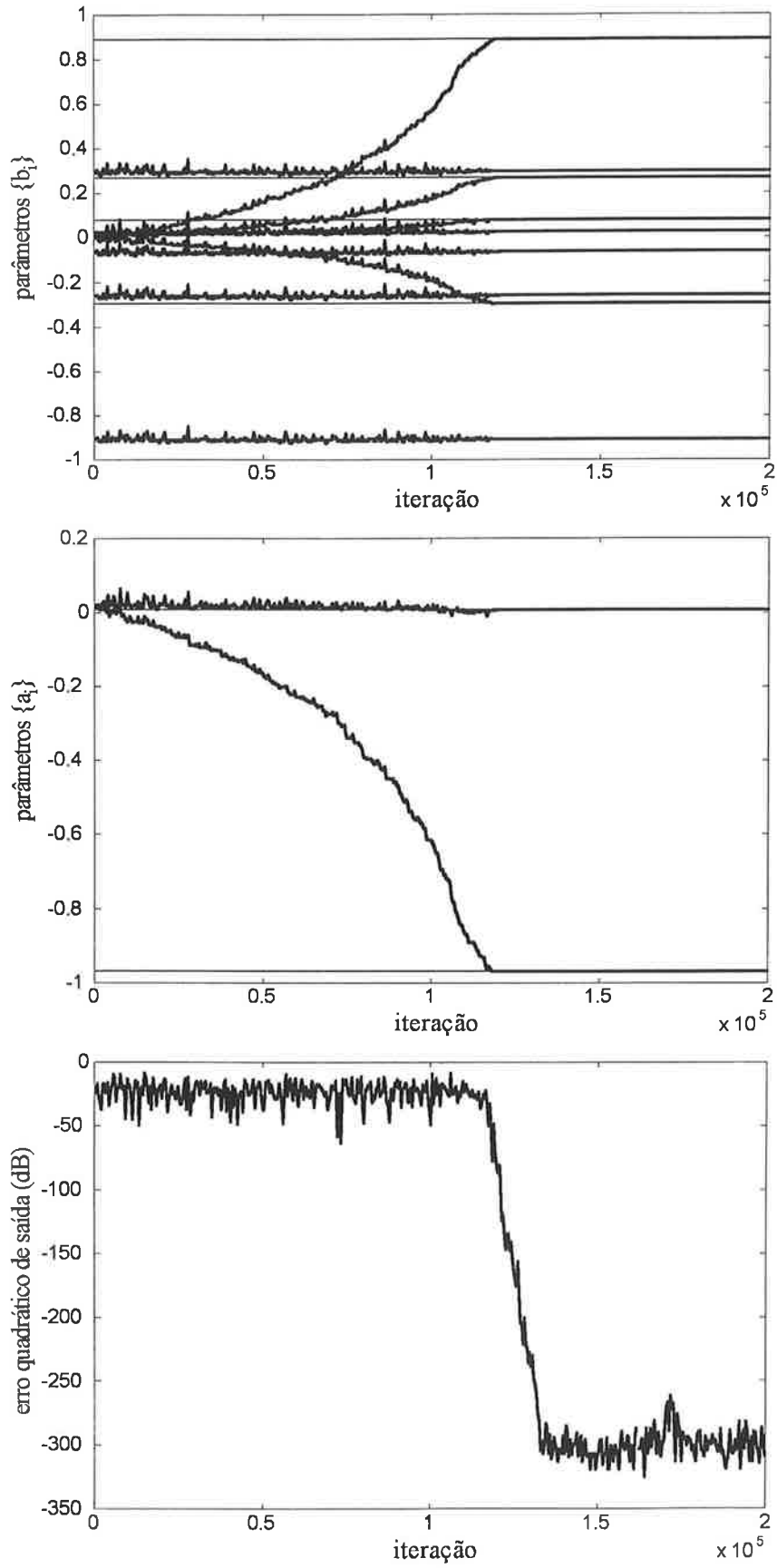


Figura 4-29: Filtro polifásico, algoritmo PLR, sistema  $H_{15}(z)$  (parâmetros em traços claros)

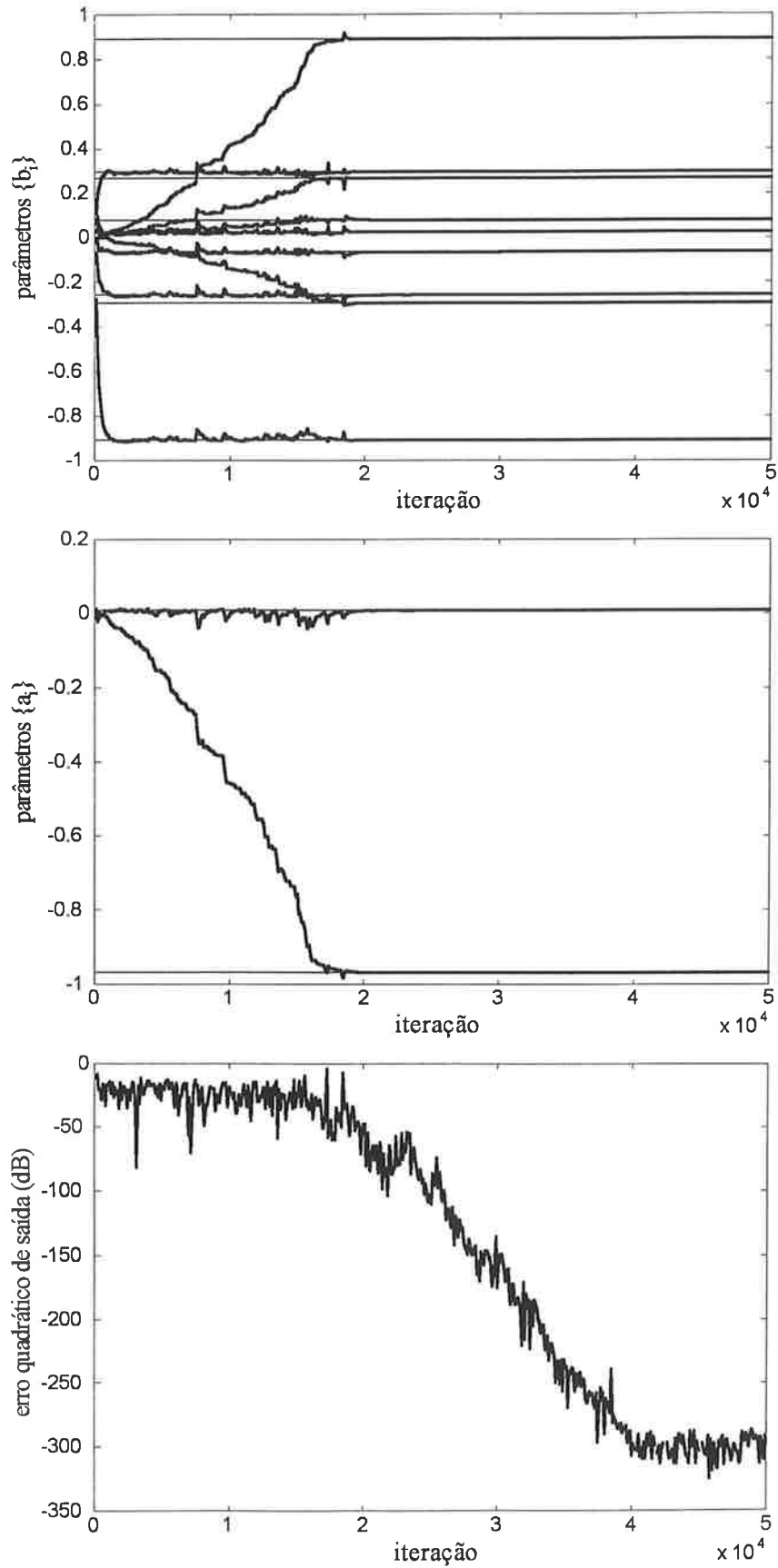


Figura 4-30: Filtro polifásico, algoritmo SMM, sistema  $H_{15}(z)$  (parâmetros em traços claros)

Comparemos, inicialmente, os resultados obtidos para o filtro polifásico adaptativo com o ponto inicial da adaptação  $\hat{H}(z, 0)$  na origem do espaço de parâmetros ( $\mathbf{w} = \mathbf{0}$ ) e com o ponto inicial fora da origem. Pode-se ver nas Tabelas acima que em 6 dos 14 casos considerados o início da adaptação na origem resultou em um tempo de adaptação *menor* do que com o ponto inicial fora da origem. Computando as funções  $H(z)$  e  $\hat{H}(z, 0)$  em cada caso pode-se constatar que a razão disto é que nestes casos mencionados, os pontos iniciais utilizados (que, como comentado acima, são resultantes da expansão polifásica do ponto inicial  $\hat{H}(z, 0)$  adotado para o caso da estrutura direta) estão mais distantes do valor correto dos parâmetros da expansão polifásica de  $H(z)$  do que a origem. Isto resulta da menor sensibilidade da posição dos polos de  $\hat{H}(z)$  em relação aos parâmetros da estrutura polifásica do que em relação aos parâmetros da estrutura direta. Em termos práticos, como consequência, temos que na escolha do ponto inicial da adaptação para a estrutura polifásica é importante levar em conta o grau de incerteza sobre  $H(z)$  e seu possível efeito na distância aos parâmetros corretos da expansão polifásica. Em função desta análise poderia-se chegar à conclusão, que, para o caso específico em questão, seria mais seguro começar a adaptação, por exemplo, na origem.

#### 4.6.1 Comparação com estrutura direta: ponto inicial fora da origem

Pode-se comparar, agora, os resultados obtidos para a estrutura polifásica com os obtidos para a estrutura direta (ver Tabelas 2.10 e 2.12), no caso do ponto inicial da adaptação fora da origem. Para facilitar a análise, as relações

$$n_r = \frac{n_{60}}{n_{60,P}} \quad (4.180)$$

entre os valores do tempo de convergência para cada estrutura,

$$\delta_r = \frac{\delta_P}{\delta} \quad (4.181)$$

entre as velocidades de convergência local (em dB/iteração), e

$$(\mu\sigma)_r = \frac{\mu_P \rho_{\max}}{\mu\sigma_{\max}} \quad (4.182)$$

Função	Algoritmo							
	RG				PLR			
	$n_r$	$\delta_r$	$\chi_r$	$(\mu\sigma)_r$	$n_r$	$\delta_r$	$\chi_r$	$(\mu\sigma)_r$
$H_1(z)$	15	28	31	1,0	2,6	2,1	2,7	0,88
$H_4(z)$	23	—	97	0,79	32	38	6,7	10
$H_6(z)$	1,9	3,7	3,2	1,1	1,0	0,71	0,37	1,4
$H_7(z)$	1,5	63	15	6,3	11	2,2	0,42	7,4
$H_9(z)$	6,7	7,5	12	0,47	0,99	1,4	3,3	0,33
$H_{14}(z)$	2,2	3,6	3,7	1,1	0,33	0,76	2,8	2,7
$H_{15}(z)$	6,1	43	6,4	7,3	0,6	1,7	0,31	2,1

Tabela 4.16: relações entre grandezas obtidas para a estrutura direta e para a estrutura polifásica ( $p = 4$ ), com início da adaptação fora da origem.

(ver discussão sobre papel de  $\mu\sigma_{\max}$  no Item 2.7.1) estão apresentados na Tabela 4.16. Para maior conveniência a relação entre o espalhamento dos valores singulares da matriz de estados da ODE,

$$\frac{(\sigma_{\max}/\sigma_{\min})}{(\rho_{\max}/\rho_{\min})} = \chi_r \quad (4.183)$$

já apresentada nas Tabelas 4.10 e 4.11 está repetida na Tabela 4.16.

A seguir faz-se se uma análise caso a caso dos resultados.

### Algoritmo RG

Para  $H_1(z)$ ,  $H_6(z)$  e  $H_{14}(z)$  obteve-se  $(\mu\sigma)_r$  próximo de 1 o que resultou em  $\delta_r$  dado, aproximadamente, pela melhoria do espalhamento  $\chi_r$ . Para  $H_4(z)$  há uma certa redução em  $(\mu\sigma)_r$ . O fato dos valores de  $n_r$  serem menores do que  $\delta_r$ , reflete a contribuição da parcela do tempo em que a convergência dos algoritmos não ocorre conforme suas características locais (comparar as Figuras 2-5, 2-6, 2-12, 4-10, 4-13 e 4-16).

Para  $H_7(z)$  e  $H_{15}(z)$  o valor de  $(\mu\sigma)_r$  é consideravelmente maior do que 1, o que resulta em um valor de  $\delta_r$  também consideravelmente maior do que a melhoria do espalhamento  $\chi_r$ . Este é um comportamento interessante, mostrando, em certos casos (notar que  $H_7(z)$  e  $H_{15}(z)$  são passa-tudo), outra origem para o ganho na velocidade de convergência local do polifásico. Isto reforça o comentário feito no Item 2.2.4 sobre o interesse em uma análise teórica referente ao comportamento de  $\mu\sigma_{\max}$ . Quanto aos valores de  $n_r$  para  $H_7(z)$  e  $H_{15}(z)$ , pode-se ver que são menores do que  $\delta_r$ , o que resulta, como pode ser visto nas Figuras 2-13, 2-23, 4-19 4-28, das características não locais da convergência em cada caso.

Para  $H_9(z)$ , em contraposição ao que ocorre para  $H_7(z)$  e  $H_{15}(z)$ ,  $(\mu\sigma)_r$  é menor do que



Função	$\min \omega \operatorname{Re}[1/D_P(e^{j\omega})]$		$\min\{\lambda_i[\mathbf{R}_P(\mathbf{w}_P^*)+\mathbf{R}_P^\top(\mathbf{w}_P^*)]\}$	
	$p = 1$	$p = 4$	$p = 1$	$p = 4$
$H_1(z)$	-1, 2	0, 48	-0, 030	0, 34
$H_4(z)$	-24	-0, 38	-0, 81	0, 63
$H_6(z)$	-1, 2	0, 48	-0, 017	0, 13
$H_7(z)$	-24	-0, 38	-0, 60	0, 087

Tabela 4.17: Grandezas relacionadas com a positividade da matriz de estados da ODE

1 o que resulta em  $\delta_r$  menor do que  $\chi_r$ . O valor de  $n_r$  é semelhante a  $\delta_r$ , o que reflete o fato da natureza da convergência em ambos os casos ser essencialmente local (comparar Figuras 2-19 e 4-22).

### Algoritmo PLR

Para o algoritmo PLR torna-se importante a questão da positividade da matriz de estados da ODE,  $\mathbf{R}_P(\mathbf{w}_P^*)$ . No Item 4.3.2 mostrou-se que uma condição suficiente para a positividade de  $\mathbf{R}_P(\mathbf{w}_P^*)$  é dada por (4.71). Além disso, conforme a discussão relativa ao caso da estrutura direta no Item 2.7.2, a não-positividade desta matriz para o algoritmo PLR parece estar relacionada com o fato de o desempenho deste algoritmo, em termos de velocidade de convergência, em especial nos casos das funções  $H_4(z)$  e  $H_7(z)$ , estar aquém do que seria sugerido pelo espalhamento dos valores singulares da mesma matriz. Conforme também discutido em 2.7.2, a importância da não-positividade de  $\mathbf{R}_P(\mathbf{w}_P^*)$  pode ser medida pelo menor autovalor de  $\mathbf{R}_P(\mathbf{w}_P^*)+\mathbf{R}_P^\top(\mathbf{w}_P^*)$ .

Para comparar a estrutura direta e a polifásica em termos deste aspecto, foram relacionados na Tabela 4.17 o menor valor assumido pela função  $\operatorname{Re}[1/D_P(e^{j\omega})]$  que aparece na condição (4.71) e o menor autovalor de  $\mathbf{R}_P(\mathbf{w}_P^*)+\mathbf{R}_P^\top(\mathbf{w}_P^*)$  no caso de  $u(n)$  branco, para os casos  $p = 1$  (que corresponde à estrutura direta) e  $p = 4$ . Podemos ver que a utilização da estrutura polifásica, ao contrário da estrutura direta, garante a positividade de  $\mathbf{R}_P(\mathbf{w}_P^*)$  para qualquer  $u(n)$  no caso da função  $H_1(z)$  (e  $H_6(z)$ , pois tem os mesmos polos que  $H_1(z)$ ) e, para  $u(n)$  branco, resulta em  $\mathbf{R}_P(\mathbf{w}_P^*)$  positiva definida nos quatro casos. Isto tem consequências importantes para a velocidade de convergência local do algoritmo, conforme visto a seguir.

Levando em conta esta discussão podemos retornar agora aos resultados da Tabela 4.16. Vemos que para  $H_4(z)$ , pode-se usar para a estrutura polifásica um valor de  $\mu$  que resulta em  $(\mu\sigma)_r = 10$ , o que, em conjunto com uma melhoria de 6,7 vezes no espalhamento  $\chi_r$ ,

resultou em uma melhoria de 32 vezes e 38 vezes, respectivamente, no tempo de convergência e na velocidade de convergência local. Analogamente ao discutido para o caso da estrutura direta, pode-se então fazer a conjectura de que este ganho de  $(\mu\sigma)_r$  resulta justamente do fato de que a matriz de estados da ODE passa a ser positiva definida ao se empregar a estrutura polifásica.

Como visto nos Itens 2.6 e 4.3, para  $H(z)$  passa-tudo (funções  $H_6(z)$ ,  $H_7(z)$ ,  $H_{14}(z)$  e  $H_{15}(z)$ ) o condicionamento da matriz de estados já é favorável no caso da estrutura direta, não havendo vantagens, quanto a este aspecto, em utilizar a estrutura polifásica. No entanto, a questão do efeito da positividade da matriz de estados também precisa ser levada em conta: podemos ver que para  $H_7(z)$  o valor de  $(\mu\sigma)_r$  também é apreciavelmente maior que 1, o que, apesar de  $\chi_r$  ser menor do que um ainda resultou em  $\delta_r$  de 2,2. Além disso, a melhoria do tempo de convergência,  $n_r$ , foi de 11 vezes, o que resulta do fato de, no caso da estrutura polifásica, ao contrário do que ocorre para a estrutura direta (ver Figura 2-15), os parâmetros não ficam um longo tempo em uma região onde a convergência do algoritmo não tem características locais. O fato de que para a estrutura polifásica isto não ocorre, reforça a conjectura feita no Item 2.7.2 sobre a possibilidade de também este efeito ter relação com a positividade da matriz de estados da ODE.

#### 4.6.2 Comparação com estrutura direta: ponto inicial na origem

Na Tabela 4.18 temos a relação

$$n_r = \frac{n_{60}}{n_{60,P}} \quad (4.184)$$

para o caso do ponto inicial da adaptação sendo a origem do espaço de parâmetros. Os valores de  $n_{60}$  foram retirados da Tabela 3.1 e os de  $n_{60,P}$  foram retirados da Tabela 4.14. Notar que em vários casos a convergência global dos algoritmos de passo constante aplicados à estrutura direta é tão lenta que não foi possível calcular o valor de  $n_r$ . Nestes caso, para se ter uma idéia dos valores de  $n_r$  pode-se comparar diretamente as figuras relacionadas na Tabela 3.2 com as relacionadas na Tabela 4.15.

Os resultados para o caso de o ponto inicial da adaptação ser a origem do espaço de parâmetros exemplificam os efeitos da melhoria das características da superfície de erro reduzida que pode resultar da utilização da estrutura polifásica, o que havia sido analisado teoricamente no Item 4.5.

Função	Algoritmo		
	RG	PLR	SMM
	$n_r$	$n_r$	$n_r$
$H_1(z)$	28	8,7	16
$H_4(z)$	(*)	(*)	(*)
$H_6(z)$	16,8	7,4	3,4
$H_7(z)$	(*)	(*)	(*)
$H_9(z)$	(*)	4,4	9,2
$H_{14}(z)$	10,5	4,8	3,1
$H_{15}(z)$	(*)	(*)	(*)

Tabela 4.18:  $n_r$ : melhoria do tempo de convergência da estrutura polifásica,  $p = 4$ , em relação à estrutura direta; (\*): convergência global excessivamente lenta da estrutura direta impede cálculo de  $n_r$ .

# Apêndice A

## Expressões usadas na análise local de algoritmos de adaptação

### A.1 Dedução de (2.156)

Escrevendo o produto interno em (2.154) como uma integral no plano  $z$  temos

$$\begin{aligned} \left\langle \frac{C(z)}{D^2(z)}, \frac{C(z)}{D(z)} \right\rangle &= \left\langle H(z) \frac{1}{D(z)}, H(z) \right\rangle \\ &= \frac{1}{2\pi j} \oint_{|z|=1} H(z^{-1}) H(z) \frac{1}{D(z)} \frac{dz}{z}. \end{aligned} \quad (\text{A.1})$$

Consideremos para maior simplicidade que  $D(z)$  não tem raízes múltiplas (uma vez que problemas numéricos não estão em questão aqui, podemos sempre considerar o caso de raízes múltiplas como o limite de um caso de raízes simples). A expansão de  $H(z)$  em resíduos, assumindo  $\deg[C(z)] < \deg[D(z)]$  como discutido no Item 2.6.4, tem, portanto, a forma

$$H(z) = \sum_{i=1}^M \frac{r_i}{z - p_i}, \quad (\text{A.2})$$

de onde resulta que

$$\frac{1}{z} H(z^{-1}) H(z) = \frac{1}{z} \sum_{i=1}^M \sum_{j=1}^M \frac{r_i}{z - p_i} \frac{r_j}{z^{-1} - p_j}$$

$$\begin{aligned}
&= \sum_{i=1}^M \sum_{j=1}^M \frac{r_i}{z - p_i} \frac{r_j}{1 - p_j z} \\
&= - \sum_{i=1}^M \sum_{j=1}^M \frac{r_i}{z - p_i} \frac{r_j/p_j}{z - 1/p_j}.
\end{aligned} \tag{A.3}$$

Cada termo da somatória acima, por sua vez, pode ser expandido em resíduos, resultando em

$$\begin{aligned}
\frac{1}{z} H(z^{-1}) H(z) &= \sum_{i=1}^M \sum_{j=1}^M \frac{r_i r_j / p_j}{p_i - 1/p_j} \left[ \frac{1}{z - 1/p_j} - \frac{1}{z - p_i} \right] \\
&= \sum_{i=1}^M \sum_{j=1}^M \frac{r_i r_j}{p_i p_j - 1} \left[ \frac{1}{z - 1/p_j} - \frac{1}{z - p_i} \right],
\end{aligned} \tag{A.4}$$

que também pode ser escrito como

$$\frac{1}{z} H(z^{-1}) H(z) = \sum_{i=1}^M \sum_{j=1}^M \frac{r_i r_j}{p_i p_j - 1} \left[ \frac{1}{z - 1/p_i} - \frac{1}{z - p_j} \right]. \tag{A.5}$$

Os termos da somatória dupla podem ainda ser separados considerando os casos  $r_i = r_j^*$  e  $r_i \neq r_j^*$ . Usando a forma (A.4) quando  $r_i = r_j^*$  e a forma (A.5) quando  $r_i \neq r_j^*$  temos

$$\begin{aligned}
\frac{1}{z} H(z^{-1}) H(z) &= \sum_{i=1}^M \frac{|r_i|^2}{|p_i|^2 - 1} \left[ \frac{1}{z - 1/p_i} - \frac{1}{z - p_i^*} \right] \\
&\quad + \sum_{i=1}^M \sum_{j=1}^M \frac{r_i r_j}{p_i p_j - 1} \left[ \frac{1}{z - 1/p_i} - \frac{1}{z - p_j} \right],
\end{aligned} \tag{A.6}$$

Supondo  $H(z)$  causal e estável, resulta que no operando da integral em (A.1) os polos dentro do percurso de integração são os polos  $1/p_i$  na expressão acima. Temos então,

$$\frac{1}{2\pi j} \oint_{|z|=1} H(z^{-1}) H(z) \frac{1}{D(z)} \frac{dz}{z} = \sum_{i=1}^M \left[ \frac{|r_i|^2}{|p_i|^2 - 1} + \sum_{j=1}^M \frac{r_i r_j}{p_i p_j - 1} \right] \frac{1}{D(1/p_i)}. \tag{A.7}$$

Consideremos agora que  $H(z)$  tem  $m$  polos com raio  $r$  e que  $r$  se aproxima de 1. Definindo

$$\gamma = \frac{\sum_{i=1}^m |r_i|^2 \operatorname{Re}\left[\frac{1}{D(1/p_i)}\right]}{\sum_{i=1}^m |r_i|^2}, \quad (\text{A.8})$$

temos que, se as seguintes condições forem atendidas:

1. a relação entre os resíduos  $r_i$  dos  $m$  polos de raio  $r$  e os resíduos dos demais polos não tende a zero quando  $r$  tende a 1,
2.  $\gamma$  não tende a zero quando  $r$  tende a um,

então

$$\sum_{i=1}^M \left[ \frac{|r_i|^2}{|p_i|^2 - 1} + \sum_{\substack{j=1 \\ r_i \neq r_j}}^M \frac{r_i r_j}{p_i p_j - 1} \right] \frac{1}{D(1/p_i)} \xrightarrow{r \rightarrow 1} \gamma \sum_{i=1}^m \frac{|r_i|^2}{r^2 - 1}, \quad (\text{A.9})$$

pois para os demais termos das somatórias em (A.7)  $|p_i|^2$  ou  $p_i p_j$  não tendem a 1 quando  $r$  tende a um. Se eliminarmos o termo  $1/D(z)$  em (A.1) teremos  $\gamma = 1$ , de onde segue que podemos escrever

$$\left| \left\langle \frac{C(z)}{D^2(z)}, \frac{C(z)}{D(z)} \right\rangle \right| \xrightarrow{r \rightarrow 1} |\gamma| \left\langle \frac{C(z)}{D(z)}, \frac{C(z)}{D(z)} \right\rangle, \quad \text{se } \lim_{r \rightarrow 1} |\gamma| \neq 0. \quad (\text{A.10})$$

A validade desta expressão pode ser facilmente verificada para exemplos numéricos. A condição  $\lim_{r \rightarrow 1} |\gamma| \neq 0$ , em termos práticos, resulta em pouca perda de generalidade: pela expressão em (A.8) podemos ver que os termos  $\operatorname{Re}[1/D(1/p_i)]$  tendem a infinito quando  $r$  tende a um; para  $m = 1$  e  $m = 2$ , portanto, sempre teremos  $|\gamma| \xrightarrow{r \rightarrow 1} \infty$ ; para  $m > 2$ , mesmo que exista um conjunto de parâmetros  $\mathbf{w}_0$  tal que  $\lim_{r \rightarrow 1} |\gamma| = 0$ , qualquer variação em relação a este valor fará com que  $|\gamma| \xrightarrow{r \rightarrow 1} \infty$ , pois (A.8) é uma soma de termos cujo módulo tende a infinito. Com base nisto, a ressalva  $\lim_{r \rightarrow 1} |\gamma| \neq 0$  não é explicitada ao utilizar (A.10).

## A.2 Dedução de (2.161)

Considerando que  $H(z)$  tem  $m$  polos com raio  $r$  e ângulo  $\pm\theta_i$ ,  $i = 1, 2, \dots, m$ ,  $\theta_i \geq 0$ , podemos escrever inicialmente

$$\begin{aligned}
\langle H(z), H(z) \rangle &= \frac{1}{2\pi} \int_0^{2\pi} |H(e^{j\omega})|^2 d\omega \\
&= \sum_{i=1}^m \frac{n_i}{\pi} \int_{\theta_i - \varepsilon_i}^{\theta_i + \varepsilon_i} |H(e^{j\omega})|^2 d\omega + \mathcal{O}(H), \tag{A.11}
\end{aligned}$$

para quaisquer  $\varepsilon_i$  tal que  $0 < \varepsilon_i < \theta_i$ , onde  $n_i = 1$  se  $\theta_i = 0$  (polo real) e  $n_i = 2$  se  $\theta_i > 0$  (par de polos complexos conjugados), e  $\mathcal{O}(H)$  reúne os demais trechos da integral. Da mesma maneira podemos escrever

$$\begin{aligned}
\left\langle \frac{H(z)}{D(z)}, \frac{H(z)}{D(z)} \right\rangle &= \frac{1}{2\pi} \int_0^{2\pi} |H(e^{j\omega})|^2 \left| \frac{1}{D(e^{j\omega})} \right|^2 d\omega \\
&= \sum_{i=1}^m \frac{n_i}{\pi} \int_{\theta_i - \varepsilon_i}^{\theta_i + \varepsilon_i} |H(e^{j\omega})|^2 \left| \frac{1}{D(e^{j\omega})} \right|^2 d\omega + \mathcal{O}'(H, D). \tag{A.12}
\end{aligned}$$

Quando  $r$  tende a 1 apenas os termos das somatórias em (A.11) e (A.12) tendem a infinito, e, com (2.152), (2.154) e (A.10) temos

$$\begin{aligned}
\lim_{r \rightarrow 1} \frac{\sigma_2}{\sigma_1} &= \lim_{r \rightarrow 1} \frac{|\gamma| \sum_{i=1}^m \frac{n_i}{\pi} \int_{\theta_i - \varepsilon_i}^{\theta_i + \varepsilon_i} |H(e^{j\omega})|^2 d\omega}{\sum_{i=1}^m \frac{n_i}{\pi} \int_{\theta_i - \varepsilon_i}^{\theta_i + \varepsilon_i} |H(e^{j\omega})|^2 \left| \frac{1}{D(e^{j\omega})} \right|^2 d\omega} \\
&\leq \lim_{r \rightarrow 1} \frac{|\gamma| \sum_{i=1}^m \frac{n_i}{\pi} \int_{\theta_i - \varepsilon_i}^{\theta_i + \varepsilon_i} |H(e^{j\omega})|^2 d\omega}{\sum_{i=1}^m \frac{n_i}{\pi} \min_{|\omega - \theta_i| < \varepsilon_i} \left| \frac{1}{D(e^{j\omega})} \right|^2 \int_{\theta_i - \varepsilon_i}^{\theta_i + \varepsilon_i} |H(e^{j\omega})|^2 d\omega} \\
&\leq \lim_{r \rightarrow 1} \frac{|\gamma| \sum_{i=1}^m \frac{n_i}{\pi} \int_{\theta_i - \varepsilon_i}^{\theta_i + \varepsilon_i} |H(e^{j\omega})|^2 d\omega}{\min_i \min_{|\omega - \theta_i| < \varepsilon_i} \left| \frac{1}{D(e^{j\omega})} \right|^2 \sum_{i=1}^m \frac{n_i}{\pi} \int_{\theta_i - \varepsilon_i}^{\theta_i + \varepsilon_i} |H(e^{j\omega})|^2 d\omega}
\end{aligned}$$

$$= \lim_{r \rightarrow 1} \frac{|\gamma|}{\min_i \min_{|\omega - \theta_i| < \varepsilon_i} \left| \frac{1}{D(e^{j\omega})} \right|^2} \quad (\text{A.13})$$

Da expressão (A.8), agora, temos que

$$\begin{aligned} |\gamma| &\leq \frac{\sum_{i=1}^m |r_i|^2 \left| \operatorname{Re} \left[ \frac{1}{D(1/p_i)} \right] \right|}{\sum_{i=1}^m |r_i|^2} \\ &\leq \max_i \left| \operatorname{Re} \left[ \frac{1}{D(1/p_i)} \right] \right| \\ &= \left| \operatorname{Re} \left[ \frac{1}{D(1/p_j)} \right] \right| \\ &\leq \left| \frac{1}{D(1/p_j)} \right|, \end{aligned} \quad (\text{A.14})$$

para algum  $j$ ,  $1 \leq j \leq m$ . Com esta expressão e (A.13) podemos escrever

$$\lim_{r \rightarrow 1} \frac{\sigma_2}{\sigma_1} \leq \lim_{r \rightarrow 1} \frac{\left| \frac{1}{D(1/p_j)} \right|}{\min_i \min_{|\omega - \theta_i| < \varepsilon_i} \left| \frac{1}{D(e^{j\omega})} \right|^2}. \quad (\text{A.15})$$

Para qualquer valor de  $\varepsilon_j$ , onde  $j$  é o mesmo de (A.14), uma vez que, para qualquer  $i$ ,  $|1/D(e^{j\theta_i})|$  tende a infinito quando  $r$  tende a 1, sempre podemos escolher os demais  $\varepsilon_i$  (isto é  $i \neq j$ ) de tal modo menores que  $\varepsilon_j$  que

$$\lim_{r \rightarrow 1} \min_i \min_{|\omega - \theta_i| < \varepsilon_i} \left| \frac{1}{D(e^{j\omega})} \right|^2 = \min_{|\omega - \theta_j| < \varepsilon_j} \left| \frac{1}{D(e^{j\omega})} \right|^2, \quad (\text{A.16})$$

de onde segue

$$\lim_{r \rightarrow 1} \frac{\sigma_2}{\sigma_1} \leq \lim_{r \rightarrow 1} \frac{\left| \frac{1}{D(1/p_j)} \right|}{\min_{|\omega - \theta_j| < \varepsilon_j} \left| \frac{1}{D(e^{j\omega})} \right|^2}, \quad (\text{A.17})$$

para qualquer  $\varepsilon_j$  tal que  $0 < \varepsilon_j < \theta_j$ . Agora, uma vez que as raízes de  $D(z)$  estão fora do círculo unitário e  $1/p_j$  está dentro do círculo unitário, para qualquer  $r$  podemos escolher um



$\varepsilon_j$  suficientemente pequeno de modo que

$$\lim_{r \rightarrow 1} \frac{\left| \frac{1}{D(1/p_j)} \right|}{\min_{|\omega - \theta_j| < \varepsilon_j} \left| \frac{1}{D(e^{j\omega})} \right|} \leq 1, \quad (\text{A.18})$$

e, portanto,

$$\begin{aligned} \lim_{r \rightarrow 1} \frac{\sigma_2}{\sigma_1} &\leq \lim_{r \rightarrow 1} \frac{\left| \frac{1}{D(1/p_j)} \right|}{\left| \frac{1}{D(1/p_j)} \right|^2} \\ &= \lim_{r \rightarrow 1} \frac{1}{\left| \frac{1}{D(1/p_j)} \right|} \\ &= 0. \end{aligned} \quad (\text{A.19})$$

# Apêndice B

## Lista dos principais símbolos empregados

A seguir está uma relação dos principais símbolos usados no trabalho. Estão indicados também o item e a página do trabalho no qual cada símbolo é definido.

$A(z) = 1 + a_1z + \dots + a_Mz^M$  : polinômio usado como denominador de  $\hat{H}(z)$ , Item 2.2.2, p. 31;

$\bar{A}(z) = z^M A(z^{-1})$  : polinômio obtido invertendo a ordem dos coeficientes de  $A(z)$ , Item 3.4.2, p. 145;

$A_P(z^p) = 1 + a_{P,1}z^p + \dots + a_{P,M}z^{pM}$  : polinômio usado como denominador de  $\hat{H}(z)$  na forma polifásica, Item 4.2.2, p. 189;

$A_P$  : matriz de estados da descrição no espaço de estados de  $H_l(z)$ , Item 4.1.1, p. 180;

$\mathbf{b}_l$  : vetor da descrição no espaço de estados de  $H_l(z)$ , Item 4.1.1, p. 180;

$B(z) = b_0 + b_1z + \dots + b_Mz^M$  : polinômio usado como numerador de  $\hat{H}(z)$ ; Item 2.2.2, p. 31;

$B_P(z) = b_{P,0} + b_{P,1}z + \dots + b_{P,pM}z^{pM}$  : polinômio usado como numerador de  $\hat{H}(z)$  na forma polifásica, Item 4.2.2, p. 189;

$C(z) = c_0 + c_1z + \dots + c_{M_c}z^{M_c}$  : polinômio usado como numerador de  $H(z)$ , Item 4.1.1, p. 179;

$C_l(z)$  : numerador de  $H_l(z)$ , Item 4.1.1, p. 181;

$C_P(z) = c_{P,0} + c_{P,1}z + \dots + c_{P,pM_c}z^{pM_c}$  : polinômio usado como numerador de  $H(z)$  na forma polifásica, Item 4.1.1, p. 181;

$\mathcal{C}$  : matriz de controlabilidade, Item 3.1.1, p. 128;

$\mathcal{C}(z)$  : vetor de funções de controlabilidade, Item 2.2.2, p. 31;

$D(z) = 1 + d_1z + \dots + d_{M_d}z^{M_d}$  : Polinômio usado como denominador de  $\hat{H}(z)$ , Item 4.1.1, p. 179;

$D_P(z^p) = 1 + d_{P,1}z^p + \dots + d_{P,M_d}z^{pM_d}$  : Polinômio usado como denominador de  $\hat{H}(z)$  na forma polifásica, Item 4.1.1, p. 181;

$\delta, \bar{\delta}$  : valores medido e estimado da velocidade de convergência local (em db/iteração), Item 2.7.1, p. 104;

$e(n) = y(n) - \hat{y}(n)$  : erro de saída, Item 1.4.1, p. 13;

$\varepsilon(n) = y'(n) - \hat{y}(n) = y(n) - \hat{y}(n), \zeta(n) \equiv 0$  : erro de aproximação, Item 1.4.1, p. 12;

$\epsilon(n) = P(z)e(n)$  : erro de saída filtrado, Item 2.2.1, p. 29 ;

$\phi_1(n), \phi_2(n)$  : vetores usados na definição de  $\mathbf{R}(\mathbf{w})$ , Item 2.2.1, p. 32 e p. 33, respectivamente;

$\phi_{P,1}(n), \phi_{P,2}(n)$  : vetores usados na definição de  $\mathbf{R}_P(\mathbf{w})$ , Item 4.3.1, p. 196 e p. 197, respectivamente;

$\varphi(n)$  : vetor dos modos normais da aproximação local do algoritmo de adaptação, Item 2.2.4, p. 42;

$\varphi_m$  : elemento de  $\varphi(n)$  associado a  $\lambda_m$ , Item 2.2.4, p. 46;

$F(z) = Q(z) - zR(z)H(z)$  : função auxiliar, Item 2.6.1, p. 71;

$F_P(z) = Q_P(z) - z^p R_P(z^p)H(z)$  : função auxiliar, Item 4.3.4, p. 202;

$\|g(z)\|^2$  : superfície de erro reduzida, Item 3.2.3, p. 135;

$h_{l,k}$  :  $k$ -ésimo termo da resposta ao pulso unitário da  $l$ -ésima fase da expansão polifásica de  $H(z)$ , Item 4.1.1, p. 180;

$H(z)$  : função de transferência do sistema sendo modelado, Item 1.3, p. 9;

$\hat{H}(z)$  : função de transferência do filtro adaptativo, Item 1.4.1, p. 13;

$H_l(z)$  : função de transferência da fase  $l$  da expansão polifásica de  $H(z)$ , Item 4.1.1, p. 179;

$\mathbf{H}, \mathbf{H}_w$  : matriz Hessiana do erro de saída, Item 3.2.1, p. 137;

$k_1 = \lim_{r \rightarrow 1} 1 / \|D(z)\|^2$  : limite superior para  $\sigma_{\min}$ , Item 2.6.4, p. 78;

$\lambda_i$  : autovalor de  $\mathbf{R}(\mathbf{w}^*)$ , Item 2.2.3, p. 40;

$\lambda_m$  : autovalor de  $\mathbf{R}(\mathbf{w}^*)$  com menor parte real, Item 2.2.4, p. 46;

$\Lambda$  : matriz dos autovalores de  $\mathbf{R}(\mathbf{w}^*)$ , Item 2.2.4, p. 42;

$\mu$  : passo de adaptação do algoritmo, Item 2.2.1, p. 29;

$m$  : número de polos de  $H(z)$  próximos à circunferência unitária, Item 2.6.4 p. 78;

$\nu_i$  : raiz de  $D(z)$ , Item 4.1.1, p. 181;

$\nabla_{w,i}^e(n, \mathbf{w}) = \partial e(n, \mathbf{w}) / \partial w_i$  : derivada em relação ao parâmetro  $w_i$  do erro de saída, considerando os parâmetros de  $\hat{H}(z)$  fixados em  $\mathbf{w}$ , Item 2.3.1, p. 50;

$\nabla_a(n)$  : sinal usado para compor  $\boldsymbol{\psi}(n)$ ;  $\nabla_a(n) = \hat{y}(n)/A(z, n)$  no caso do algoritmo RG, Item 2.3.2, p. 53;  $\nabla_a(n) = y(n)/A(z, n)$  no caso do algoritmo SMM, Item 2.4.1, p. 61;

$\nabla_{P,a}(n)$  : sinal usado para compor  $\boldsymbol{\psi}_P(n)$ ;  $\nabla_{P,a}(n) = \hat{y}(n)/A_P(z^P, n)$  no caso do algoritmo RG, Item 4.2.3, p. 190;  $\nabla_{P,a}(n) = y(n)/A_P(z^P, n)$  no caso do algoritmo SMM, Item 4.2.3, p. 191;

$\nabla_b(n) = u(n)/A(z, n)$  : sinal usado para compor  $\boldsymbol{\psi}(n)$ , Item 2.3.2, p. 53;

$\nabla_{P,b}(n) = u(n)/A_P(z^P, n)$  : sinal usado para compor  $\boldsymbol{\psi}_P(n)$ , Item 4.2.3, p. 190;

$\mathbf{p}_m, |\mathbf{p}_m|$  : vetor de  $\mathbf{P}$  associado a  $\lambda_m$  ; o vetor dado pelo módulo dos elementos de  $\mathbf{p}_m$ , Item 2.2.4, p. 47 e p. 48, respectivamente;

- $P(z)$  : filtro do erro de saída, Item 2.2.1, p. 29;
- $\mathbf{P}$  : matriz dos autovetores de  $\mathbf{R}(\mathbf{w}^*)$ , Item 2.2.4, p. 42;
- $Q(z)$  : função auxiliar, Item 2.6.1, p. 71;
- $Q_P(z)$  : função auxiliar, Item 4.3.4, p. 202;
- $R(z)$  : função auxiliar, Item 2.6.1, p. 71;
- $R_P(z)$  : função auxiliar, Item 4.3.4, p. 202;
- $\mathbf{R}(\mathbf{w}), \mathbf{R}'(\mathbf{w})$  : matriz de estados da ODE, Item 2.2.1, p. 33;
- $\mathbf{R}_P(\mathbf{w}), \mathbf{R}'_P(\mathbf{w})$  : matriz de estados da ODE, Item 4.3.1, p. 197;
- $r$  : raio de polo de  $H(z)$ , Item 2.6.4, p. 77;
- $\psi(n)$  : vetor de adaptação do algoritmo, Item 2.2.1, p. 29;
- $\psi_P(n)$  : vetor de adaptação do algoritmo para estrutura polifásica, Item 4.2.3, p. 192;
- $\mathcal{S}_u(z) = \sum_{k=-\infty}^{\infty} E[u(n)u(n+k)]z^k$  : densidade espectral de potência de  $u(n)$ , Item 1.5.1, p. 20;
- $\sigma_{\max}, \sigma_{\min}$  : valores singulares máximo e mínimo de  $\mathbf{R}(\mathbf{w}^*)$ , Item 2.6.1, p. 71;
- $\sigma_1, \sigma_3$  : limites inferiores para  $\sigma_{\max}$ , Item 2.6.4, p. 75;
- $\sigma_2$  : limite inferior para  $\sigma_{\max}$ , Item 2.6.4, p. 76;
- $\sigma_4$  : limite inferior para  $\sigma_{\max}$ , Item 2.6.5, p. 80;
- $\sigma_5$  : limite inferior para  $\sigma_{\max}$ , Item 2.6.5, p. 81;
- $\sigma_6$  : limite inferior para  $\sigma_{\max}$ , Item 2.6.5, p. 82;
- $\theta(n)$  : vetor de parâmetros da aproximação local do algoritmo de adaptação, Item 2.2.3, p. 37;
- $u(n)$  : sinal de entrada do sistema sendo modelado  $H(z)$  e do filtro adaptativo  $\hat{H}(z)$ , Item 1.3, p. 9;

$v(n) = F(z)u(n)$  : sinal auxiliar, Item 2.6.1, p. 71;

$v_1, v_2, v_3$  : velocidades de crescimento dos limites inferiores de  $\sigma_{\max}$ , Item 2.6.4, p. 77;

$V(z)$  : complemento passa-tudo de  $C(z)$ , Item 3.1.2, p. 130;

$w(n)$  : vetor de parâmetros de  $\widehat{H}(z)$  Item 2.2.1, p. 29;

$w_P(n)$  : vetor de parâmetros de  $\widehat{H}(z)$  na forma polifásica, Item 4.2.3, p. 192;

$w^*$  : vetor de parâmetros de  $H(z)$ , Item 2.2.1, p. 30;

$w_P^*$  : vetor de parâmetros da expansão polifásica de  $H(z)$ , Item 4.3.4, p. 203;

$\tilde{w}(n) = w^* - w(n)$  : erro de parâmetros, Item 2.2.1, p. 32;

$w(t)$  : vetor de parâmetros determinístico da ODE usado para aproximar  $w(n)$ , Item 2.2.1, p. 29;

$y(n) = H(z)u(n) + \zeta(n)$  : sinal de saída do sistema modelado acrescido de ruído de medida  $\zeta(n)$ , Item 1.3, p. 9;

$y'(n) = H(z)u(n)$  : sinal de saída do sistema modelado não acrescido de ruído de medida, Item 2.2.2, p. 31;

$\hat{y}(n) = \widehat{H}(z)u(n)$  : sinal de saída do filtro adaptativo, Item 1.4.1, p. 13;

$\zeta(n)$  : ruído de medida que é acrescido à saída  $y'(n) = H(z)u(n)$  do sistema sendo modelado, Item 1.3, p. 9;

$\zeta_k(z)$  : função usada na análise SVD de  $\|g(z)\|^2$ , Item 3.4.1, p. 142;

# Bibliografia

- [1] P.A. Regalia, *Adaptive IIR Filtering in Signal Processing and Control*, Marcel Dekker, New York, 1995.
- [2] T. Kailath, A.H. Sayed, B. Hassibi, *State-Space Estimation Theory* (pre-print)
- [3] J.J. Shynk, "Adaptive IIR Filtering ", *IEEE Acoustics, Speech and Signal Processing Magazine*, pp 4-21, Abril 1989
- [4] S.D. Stearns, "Error Surfaces of Recursive Adaptive Filters", *IEEE Transactions on Circuits and Systems*, pp 603-606, Junho 1981.
- [5] T. Söderström, P. Stoica, "Some Properties of the Output Error Method", *Automatica*, vol. 18, pp 93-99, 1982.
- [6] T. Söderström, "On the Uniqueness of Maximum Likelihood Identification, *Automatica*, vol. 11, pp 193-197, 1975.
- [7] H. Fan, M. Nayeri, "On Error Surfaces of Sufficient Order Adaptive IIR Filters: Proofs and Counterexamples to a Unimodality Conjecture ", *IEEE Transactions on Acoustics, Speech and Signal Processing*, pp 1436-1442, Setembro 1989.
- [8] A. V. Oppenheim, R.W. Schaffer, *Discrete-time signal processing*, Prentice-Hall International, 1989.
- [9] K. E. Steiglitz, L. E. McBride, "A technique for the identification of linear systems ", *IEEE Transactions on Automatic Control*, vol.10 pp 461-464, Outubro 1965
- [10] P. Stoica, T. Söderström, "The Steiglitz-McBride identification algorithm revisited—convergence analysis and accuracy aspects", *IEEE Transactions on Automatic Control*, vol. 26, pp 712-717, Junho 1981

- [11] M. Gerken, Notas de aula do curso "PEE 709-Filtros digitais sem perdas ", 1991
- [12] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Englewood Cliffs, 1991
- [13] A. Benveniste et al., "Analysis of Stochastic Approximation Schemes with Discontinuous and Dependent Forcing Terms with Applications to Data Communication Algorithms", *IEEE Transactions on Automatic Control*, pp 1042-1058, Dezembro 1980.
- [14] H. Fan, "Application of Benveniste's Convergence Results in the Study of Adaptive IIR Filtering Algorithms", *IEEE Transactions on Information Theory*, pp 692-709, Julho 1988
- [15] B. Widrow, S. D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, 1985
- [16] N. Kalouptsidis, S. Theodoridis, ed., *Adaptive System Identification and Signal Processing Algorithms*, Prentice-Hall, London, 1993
- [17] L. Ljung, T. Söderström, *Theory and Practice of Recursive Identification*, MIT Press, Cambridge, 1983
- [18] L. Ljung, *System Identification, Theory For The User*, Prentice-Hall, Englewood Cliffs, 1987
- [19] L. Ljung, "On Positive Real Transfer Functions and the Convergence of Some Recursive Schemes", *IEEE Transactions on Automatic Control*, pp 539-551, Agosto 1977.
- [20] C. F. N. Cowan, "Channel Equalization ", em N. Kalouptsidis, S. Theodoridis, ed., *Adaptive System Identification and Signal Processing Algorithms*, Prentice-Hall, London, 1993
- [21] S. U. H. Qureshi, "Adaptive Equalization ", em K. Feher, ed., *Advanced Digital Communications*, Prentice-Hall, Englewood Cliffs, 1987
- [22] F. Ling, "Echo Cancellation ", em N. Kalouptsidis, S. Theodoridis, ed., *Adaptive System Identification and Signal Processing Algorithms*, Prentice-Hall, London, 1993
- [23] D. G. Messerschmitt, "Echo Cancellation in Speech and Data Transmission", em K. Feher, ed., *Advanced Digital Communications*, Prentice-Hall, Englewood Cliffs, 1987



- [24] S. J. Elliot, P. A. Nelson, "Active Noise Control", *IEEE Signal Processing Magazine*, vol. 10, n<sup>o</sup> 4, pp 12-35, Abril 1993
- [25] M. Nayeri, W. K. Jenkins, "Alternate Realizations to Adaptive IIR Filters and Properties of Their Performance Surfaces", *IEEE Transactions on Circuits and Systems*, pp 485-496, Abril 1989
- [26] H. Fan, "A Structural View of Asymptotic Convergence Speed of Adaptive IIR Filtering Algorithms: Part I- Infinite Precision Implementation", *IEEE Transactions on Acoustics, Speech and Signal Processing*, pp 1493-1517, Abril 1993
- [27] P. M. S. Burt, M. Gerken, "A Polyphase IIR Adaptive Filter: Error Surface Analysis and Application", a ser apresentado no *IEEE ICASSP-97*
- [28] C. R. Johnson, Jr., "A Convergence Proof for a Hyperstable Adaptive Recursive Filter", *IEEE Transactions on Information Theory*, pp 745-749, Novembro 1979
- [29] *SVD and Signal Processing-Algorithms, Applications and Architectures*, Ed. E. F. Deprettere, North-Holland, Amsterdam, 1988
- [30] U. Forsen, "Adaptive Filters with Arbitrary Structures", Licenciata Thesis, Royal Institute of Technology, Stockholm, 1990
- [31] P. Lancaster e M. Tismenetsky, *The Theory of Matrices*, 2<sup>a</sup> ed., Academic Press, 1985
- [32] M. Hirsch e S. Smale, *Differential Equations, Dynamical Systems, and Linear Algebra*, Academic Press, 1974.