

RODRIGO BORGES BARCELOS

**O EFEITO DE INICIALIZAÇÃO NO PSEUDO
ALGORITMO DE PROJEÇÕES AFINS:
ESTUDO E ANÁLISE ESTOCÁSTICA**

Dissertação de mestrado submetida à Banca Examinadora do Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Santa Catarina, como requisito parcial para a obtenção do Grau de Mestre em Engenharia Elétrica, sob a orientação do Prof. Dr. Márcio Holsbach Costa e coorientação do Prof. Dr. Sérgio José Melo de Almeida.

Florianópolis
2013

Catálogo na fonte pela Biblioteca Universitária da
Universidade Federal de Santa Catarina.

Barcelos, Rodrigo Borges

O efeito de inicialização no pseudo algoritmo de projeções afins : estudo e análise estocástica / Rodrigo Borges Barcelos ; orientador, Márcio Holsbach Costa ; co-orientador, Sérgio José Melo de Almeida. - Florianópolis, SC, 2013.
156 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico. Programa de Pós-Graduação em Engenharia Elétrica.

Inclui referências

1. Engenharia Elétrica. 2. Filtro adaptativo. 3. Pseudo algoritmo de projeções afins. 4. Análise estocástica. 5. Identificação de sistemas. I. Costa, Márcio Holsbach. II. Almeida, Sérgio José Melo de. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia Elétrica. IV. Título.

RODRIGO BORGES BARCELOS

**O EFEITO DE INICIALIZAÇÃO NO PSEUDO ALGORITMO DE
PROJEÇÕES AFINS: ESTUDO E ANÁLISE ESTOCÁSTICA**

Esta Dissertação foi julgada adequada para obtenção do título de Mestre em Engenharia Elétrica, área de concentração Comunicações e Processamento de Sinais, e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia Elétrica.

Florianópolis, 10 de Maio de 2013.

Prof. Patrick Kuo-Peng, Dr.
Coordenador do Curso
Universidade Federal de Santa Catarina

Prof. Márcio Holsbach Costa, Dr.
Orientador
Universidade Federal de Santa Catarina

Banca Examinadora:

Prof. Sérgio José Melo de Almeida, Dr.
Coorientador
Universidade Católica de Pelotas

Prof. José Carlos Moreira Bermudez, Dr.
Universidade Federal de Santa Catarina

Prof. Eduardo Luiz Ortiz Batista, Dr.
Universidade Federal de Santa Catarina

Prof. Cássio Guimarães Lopes, Dr.
Universidade de São Paulo

AGRADECIMENTOS

Este trabalho encerra mais um ciclo importante em minha vida, ao mesmo tempo em que abre portas a um novo mundo surpreendente à minha frente. A cada passo dado, sinto que me proporcionou amadurecimento, não só científico, mas com a redescoberta constante dos meios de interação e integração em ambiente acadêmico e social. E aliado a tudo isso, a importância de auto-conhecimento e reflexões sobre a importância de se manter consciente e livre. Dessa forma, se fazem necessários alguns agradecimentos especiais.

Agradeço inicialmente aos servidores da UFSC que contribuíram de alguma forma com o meu aprendizado, seja pelo auxílio técnico-administrativo, serviços de restaurante e biblioteca, limpeza ou segurança. Dentre eles destaque: Josiane Kirchner, Elton, Marcelo Siqueira, Walter Gontijo e Wilson Costa.

Aos professores do programa de pós-graduação que contribuíram com conhecimento e experiência, em especial ao professor Raimes Moraes que me auxiliou em um momento decisivo do meu trajeto no mestrado. Não posso esquecer-me de agradecer também aos professores Joceli Mayer, do LPDS, e Rui Seara e Sidnei Noceti Filho, do LINSE, por me receberem de forma tão hospitaleira em seus laboratórios. Aos professores do curso de extensão em língua francesa da UFSC: Carolina, Diego, Augusto e Monique. *Je serai éternellement reconnaissant.*

Aos colegas de laboratório pela possibilidade de aprendizado científico e interpessoal durante esse período no mestrado: Vini, Marcos, Daniel, Renata, Tales, Bina, Matsuo, Eduardo, Antoniou e Franciele.

Às pessoas especiais que reconheci na “ilha da magia” (sim, porque segundo o poeta, “a gente não faz amigos, reconhece-os”), e que de uma forma ou outra me apoiaram e me incentivaram: Carol, Henrique Popeye, Alline, Fabi, Deyvid e Leonardo, Cha e Thiago, Jonas, Rafael, Rodrigo, Monique, Cacá, Paulo e Lisi. Aos amigos que aqui reencontrei: Gilmar (e os amigos do futebol), Maria Olandina e Gabi. Aquelas que me apoiaram e inspiraram de outros “pagos”: Lorena, Marília, Pri, Nine e Germana. E principalmente aos grandes amigos da minha terra natal, os quais mantenho sempre junto no coração.

Ao professor José Carlos Moreira Bermudez por sua disposição e importantes contribuições neste trabalho.

Ao meu coorientador Sérgio José Melo de Almeida pelos conselhos, amizade e pela valiosa ajuda ao longo do trabalho.

Ao meu orientador Márcio Holsbach Costa, minha profunda admiração e respeito pelo mestre e pessoa.

A mais nova integrante da família, minha afilhada Isabella, por representar a continuidade e trazer alegrias às nossas vidas. Ao meu afilhado Lucas, que continue traçando um caminho brilhante. Ao Rogers e à Carol, que me propiciaram um grande momento de testemunhar a celebração de seu amor, que continue cada vez mais forte.

E finalmente, aos meus pais, minha irmã e a tia Tereza por todo apoio e amor incondicional, não só durante este trabalho, mas durante toda a minha vida.

*One must learn by doing the thing;
For though you think you know it,
You have no certainty, until you try*
(Sophocles, século 5 A.C.)

[...] ensinar não é transferir conhecimentos, conteúdos, nem forrar a ação pela qual um sujeito criador dá forma, estilo ou alma a um corpo indeciso e acomodado. Não há docência sem discência, as duas se explicam e seus sujeitos, apesar das diferenças que os conotam, não se reduzem à condição de objeto, um do outro. Quem ensina aprende ao ensinar e quem aprende ensina ao aprender.

(Paulo Freire, 1996)

RESUMO

Este trabalho apresenta um novo modelo analítico para o comportamento do *pseudo algoritmo de projeções afins* (PAP), uma variação de baixo custo computacional do algoritmo de projeções afins convencional (APA), introduzindo um mecanismo até então desconhecido que ocorre nas primeiras iterações do processo de adaptação, denominado de *Efeito de Inicialização*. Dessa forma, contrariamente ao apresentado na literatura específica, a condição de passo unitário é necessária, porém não suficiente para garantir a equivalência no comportamento de ambos os algoritmos (PAP e APA). Esse efeito ocorre devido à característica escalar do erro instantâneo utilizado no algoritmo PAP e afeta o processo de ortogonalização na atualização dos coeficientes. Uma modelagem matemática para esse mecanismo é derivada neste trabalho, e posteriormente empregada para obtenção de novas equações determinísticas (recursivas e fechadas) para o comportamento médio dos coeficientes, erro quadrático médio e momentos de segunda ordem do vetor de erro dos coeficientes, considerando um sinal de excitação de entrada autoregressivo. Equações que descrevem quantitativamente o regime permanente também são derivadas. Finalmente, simulações Monte Carlo corroboram com a excelente acurácia dos modelos analíticos propostos, e os resultados teóricos apresentados demonstram que o Efeito de Inicialização pode acarretar em um impacto significativo no comportamento de convergência do PAP em relação ao APA.

Palavras-chave: Filtro Adaptativo, Pseudo Algoritmo de Projeções Afins, Análise Estocástica, Identificação de Sistemas.

ABSTRACT

This work presents a new analytical model for the behavior of the pseudo affine projection (PAP) adaptive algorithm, a low cost computational variation of the conventional affine projection algorithm (APA), introducing a previously unknown mechanism that occurs at the first iterations of the adaptation process, called *Initialization Effect*. Unlike described in the specific literature, the condition of the unity step-size is necessary, but it is not sufficient to ensure the equivalence of both algorithms (PAP and APA). This effect occurs due to the scalar structure of the instantaneous error used in the PAP algorithm and it affects the correct orthogonalization process of the weights. A mathematical model for this mechanism is provided in this work, and subsequently it is used to obtain new deterministic equations (recursive and closed-form) to the mean weight behavior, mean squared error and second-order moments of the weight error vector for autoregressive inputs. Also, equations that describe quantitatively the steady-state behavior are also derived. Finally, Monte Carlo simulations show improvements in the accuracy of the proposed analytical models and the presented theoretical results indicate that the Initialization Effect can lead to a significant impact on the behavior of the algorithm convergence.

Keywords: Adaptive Filter, Pseudo Affine Projection Algorithm, Stochastic analysis, System Identification.

LISTA DE FIGURAS

Figura 1 – Realização transversal de um filtro FIR ($N = 4$).....	33
Figura 2 – Diagrama de blocos de estimação linear.....	34
Figura 3 – Exemplo de função de custo paraboloide (Figura 6.4 de (MANOLAKIS; INGLE e KOGON, 2005)).	36
Figura 4 – Diagrama de blocos de um filtro adaptativo supervisionado.	39
Figura 5 – Classes de aplicações de filtros adaptativos supervisionados: (a) identificação de sistema; (b) modelamento inverso; (c) predição de sinal; (d) cancelamento de interferência multi-sensor.....	40
Figura 6 – Diagrama de blocos de um problema de identificação de sistema....	42
Figura 7 – Modelo de (a) gerador e (b) analisador de processo (Figura 2.2 de (HAYKIN, 1996)).....	46
Figura 8 – Resposta impulsiva de um canal de eco (ITU-T G.168).....	110
Figura 9 - Resposta impulsiva utilizada em Almeida (2004).....	110
Figura 10 – Comportamento do EQM para um processo AR(2); $N = 128$ e $P = 2$. (a) PAP em vermelho; (b) APA em preto; (c) equação (3.48) em azul; (d) modelo teórico de Almeida (2005) em ciano e (e) modelo proposto em magenta.....	112
Figura 11 – Comportamento do EQM para um processo AR(1); $N = 64$ e $P = 1$. (a) PAP em vermelho; (b) APA em preto; (c) equação (3.48) em azul; (d) modelo teórico de Almeida (2005) em ciano e (e) modelo proposto em magenta.....	112
Figura 12 – Comportamento médio do 8° e do 32° coeficiente para o Exemplo 6.3.1. AR(1); $N = 64$ e $P = 1$. (a)(d) simulação Monte Carlo (preto); (b)(e) modelo proposto em (4.14) (azul); e (c)(f) coeficiente da planta (vermelho).	113
Figura 13 – Comportamento médio do 10° coeficiente de $\mathbf{v}(n)$ para o Exemplo 6.3.1 em escala logarítmica. AR(1); $N = 64$ e $P = 1$. (a) simulação Monte Carlo (preto); (b) modelo em Almeida (2004) (vermelho) e (c) modelo proposto em (4.11) (azul).....	114
Figura 14 – Comportamento médio do 5° e 37° coeficiente para o Exemplo 6.3.2. AR(2); $N = 128$ e $P = 2$. (a)(d) simulação Monte Carlo (preto); (b)(e) modelo proposto em (4.14) (azul); e (c)(f) coeficiente da planta (vermelho).	115
Figura 15 – Comportamento médio do 30° e 42° coeficiente de $\mathbf{v}(n)$ para o Exemplo 6.3.2 em escala logarítmica. AR(2); $N = 128$ e $P = 2$. (a)(d) simulação Monte Carlo (preto); (b)(e) modelo em Almeida (2004) (vermelho) e (c)(f) modelo proposto em (4.11) (azul).	116
Figura 16 – Comportamento médio do 5° coeficiente de $\mathbf{v}(n)$ para o filtro do Exemplo 6.3.3 em escala linear; $\mathbf{R}_{\hat{\mathbf{v}}}(n)$ estimado numericamente; AR(30); $N = 64$ e $P = 30$. (a) simulação Monte Carlo (preto); (b) modelo em Almeida (2004) (vermelho) e (c) modelo proposto em (4.14) (azul).....	117
Figura 17 – Erro quadrático médio para o Exemplo 6.4.1. AR(1); $N = 64$; $P = 1$. (a) Simulação Monte Carlo (preto); (b) modelo em Almeida (2004)	

(vermelho), (c) modelo proposto em (5.35) (azul), (d) regime permanente segundo (5.98); e (e) erro mínimo.	119
Figura 18 – Erro quadrático médio para o Exemplo 6.4.2. AR(2); $N = 64$; $P = 2$. (a) Simulação Monte Carlo (preto); (b) modelo em Almeida (2004) (vermelho), (c) modelo proposto em (5.35) (azul); (d) regime permanente segundo (5.98); e (e) erro mínimo.	120
Figura 19 – Erro quadrático médio para o Exemplo 6.4.3. AR(2); $N = 64$. simulação Monte Carlo: (a) $P = 2$ (preto) e (c) $P = 8$ (magenta); modelo proposto em (5.35): (b) $P = 2$ (azul) e (d) $P = 8$ (vermelho).....	121
Figura 20 – Erro quadrático médio para o Exemplo 6.4.4. AR(15); $N = 64$; $P = 15$. (a) Simulação Monte Carlo (preto); (b) modelo em Almeida (2004) (vermelho); (c) modelo proposto em (5.35) (azul); (d) regime permanente segundo (5.98); e (e) erro mínimo.	122
Figura 21 – Erro quadrático médio para o Exemplo 6.4.5. AR(15); $\mathbf{w}(0) = \mathbf{w}^o$; $N = 64$, $P = 15$. (a) Simulação Monte Carlo (preto); (b) modelo de (ALMEIDA, 2004) (vermelho) e (c) modelo proposto (azul).	123
Figura 22 – Erro quadrático médio para o Exemplo 6.4.5. AR(15); $\mathbf{w}(0) = \mathbf{w}^o$; $N = 128$, $P = 15$. (a) Simulação Monte Carlo (preto); (b) modelo de (ALMEIDA, 2004) (vermelho); (c) modelo proposto (azul); (d) regime permanente segundo (5.98); e (e) erro mínimo.....	124
Figura 23 – Erro quadrático médio para o Exemplo 6.4.6. AR(30); $N = 128$; $P = 30$. (a) Simulação Monte Carlo (preto); (b) modelo em Almeida (2004) (vermelho); (c) modelo proposto em (5.35) (azul); (d) regime permanente segundo (5.98); e (e) erro mínimo.	125
Figura 24 – Erro quadrático médio para o Exemplo 6.4.6. AR(30); $N = 64$; $P = 30$. (a) Simulação Monte Carlo (preto); (b) modelo em Almeida (2004) (vermelho); (c) modelo proposto em (5.35) (azul); (d) regime permanente segundo (5.98); e (e) erro mínimo.	126
Figura 25 – Erro quadrático médio para o Exemplo 6.4.6. AR(30); $N = 40$; $P = 30$. (a) Simulação Monte Carlo (preto); (b) modelo em Almeida (2004) (vermelho); (c) modelo proposto em (5.35) (azul); (d) regime permanente segundo (5.98); e (e) erro mínimo.	126
Figura 26 – Interpretação vetorial da estimação pelo EMQ para $N = 3$ e $M = 2$ (Baseada na figura 8.5 de (MANOLAKIS; INGLE e KOGON, 2005)). ...	136
Figura 27 – Interpretação geométrica do algoritmo NLMS (Figura 2.2 de (ALMEIDA, 2004)).....	143
Figura 28 – Interpretação geométrica do novo algoritmo (Figura 2.3 de (ALMEIDA, 2004)).....	145
Figura 29 – Vetor base (a) e deslocados (b e c) do algoritmo para determinação dos elementos da matriz Ψ	151
Figura 30 – Pontos percentuais superior e inferior da distribuição qui-quadrada.	154
Figura 31 – Curvas Características Operacionais para $\alpha = 0,01$	156
Figura 32 – Curvas Características Operacionais para $\alpha = 0,05$	156

LISTA DE TABELAS

Tabela 1- Custo computacional estimado por iteração do APA para sinais de número real, em termos do número de multiplicações e adições reais.....	53
Tabela 2- Custo computacional estimado por iteração do PAP para sinais de número real, em termos do número de multiplicações e adições reais.....	53
Tabela 3- Número aproximado de operações do APA e PAP por iteração, para $P=\{1,2,3\}$ e considerando-se $N \gg P$	54
Tabela 4- Exemplos de (3.18) para $0 \leq n \leq 3$ e $0 < i \leq P$	62
Tabela 5- Relação iterativa de igualdade de $u^T(n-i)v(n)$ em (3.24) para o PAP. 64	
Tabela 6- Relação iterativa de igualdade de $u^T(n-i)v(n)$ em (3.51) para o APA.69	
Tabela 7- Pontos percentuais da distribuição qui-quadrada.....	154

LISTA DE ABREVIATURAS E SIGLAS

$E\{.\}$	Valor esperado.
$\text{tr}\{.\}$	Operação linear de traço.
$\min\{.,.\}$	Função que resulta no menor valor entre os argumentos.
APA	Algoritmo de Projeções Afins.
AR	Autoregressivo.
DQM	Desvio Quadrático Médio.
EMQ	Estimador dos Mínimos Quadrados.
EQM	Erro Quadrático Médio.
EQMex	Erro Quadrático Médio em Excesso.
FBLP	<i>Forward and Backward Linear Prediction.</i>
FDP	Função Densidade de Probabilidade.
FIR	Resposta ao Impulso Finita (<i>Finite Impulse Response</i>).
IC	Intervalo de Confiança.
IIR	Resposta ao Impulso Infinita (<i>Infinite Impulse Response</i>).
i.i.d.	Independente e indenticamente distribuída.
ITU	<i>International Telecommunications Union.</i>
LMS	Algoritmo adaptativo <i>Least Mean-Square</i> .
NLMS	Algoritmo adaptativo LMS normalizado.
PAP	Pseudo Algoritmo de Projeções Afins.
PDS	Processamento Digital de Sinais.
RLS	Algoritmo adaptativo <i>Recursive Least Squares</i> .
SNR	Relação Sinal-Ruído (<i>Signal-to-Noise Ratio</i>).
WSS	Estacionário no sentido amplo.

LISTA DE SÍMBOLOS

n	Iterações igualmente espaçadas no tempo discreto.
N	Ordem da planta e número de elementos $u(n)$ no regressor $\mathbf{u}(n)$.
P	Ordem do algoritmo APA($P+1$) e número de regressores em $\mathbf{U}(n)$.
K	Número de regressores atual e passados. ($K = P+1$).
M	Número de coeficientes do processo AR.
G	Graus de liberdade.
$z(n)$	Inovação do processo AR.
$u(n)$	Saída do processo AR e sinal de excitação do sistema.
$\mathbf{u}(n)$	Vetor de excitação de entrada (regressor) ($N \times 1$).
\mathbf{R}_u	Matriz de correlação do regressor $\mathbf{u}(n)$ ($N \times N$).
\mathbf{R}_{uu}	Matriz de correlação do regressor $\mathbf{u}(n)$ ($(N+P) \times (N+P)$).
\mathbf{p}_{du}	Vetor de correlação cruzada entre $d(n)$ e $\mathbf{u}(n)$ ($N \times 1$).
$\mathbf{U}(n)$	Matriz de P vetores regressores passados ($N \times P$).
$\mathbf{U}^+(n)$	Pseudo inversão de $\mathbf{U}(n)$ ($N \times P$).
$\mathbf{U}_u(n)$	Matriz de regressor atual e P passados ($N \times (P+1)$).
$\mathbf{U}_u^+(n)$	Pseudo inversão de $\mathbf{U}_u(n)$ ($N \times (P+1)$).
$\mathbf{P}_U(n)$	Operador de projeção no subespaço gerado pelas colunas de $\mathbf{U}(n)$.
$\mathbf{P}_\perp(n)$	Operador de projeção complementar ortogonal ao subespaço gerado pelas colunas de $\mathbf{U}(n)$.
$\mathbf{w}(n)$	Vetor de coeficientes ($N \times 1$).
\mathbf{w}^o	Vetor de coeficientes ótimos da planta ($N \times 1$).
$\mathbf{v}(n)$	Vetor de erro dos coeficientes ($N \times 1$).
$\mathbf{K}(n)$	Matriz dos momentos de segunda ordem de $\mathbf{v}(n)$.
$k(n)$	Traço de $\mathbf{K}(n)$.
$y(n)$	Saída do filtro adaptativo.
$d(n)$	Sinal desejado (saída da planta corrompida por ruído).
$e(n)$	Erro escalar da diferença entre a saída do filtro adaptativo e sinal desejado.
$r(n)$	Distúrbio aditivo da saída da planta.
$\mathbf{d}(n)$	Vetor de sinais desejados $d(n)$ ($K \times 1$).
$\mathbf{e}(n)$	Vetor de erro escalar $e(n)$ ($K \times 1$).
$\mathbf{r}(n)$	Vetor de distúrbio aditivo do modelo $r(n)$ ($K \times 1$).

$\mathbf{r}(n-1)$	Vetor de distúrbios aditivos $r(n)$ passados ($P \times 1$).
σ_z^2	Variância do sinal de inovação $z(n)$.
σ_u^2	Variância do sinal de excitação $u(n)$.
σ_r^2	Variância do ruído.
μ	Passo de adaptação do algoritmo.
\mathbf{a}	Vetor de coeficientes do processo AR ($Q \times 1$).
$\hat{\mathbf{a}}(n)$	Estimação dos mínimos quadrados do vetor \mathbf{a} ($P \times 1$).
$r_o(n)$	Sequência de ruído filtrado.
$\phi(n)$	Erro de estimação dos mínimos quadrados de $\mathbf{u}(n)$ e vetor de direção de atualização dos coeficientes ($N \times 1$).
$\mathbf{R}_{\phi\phi}$	Matriz de correlação de $\phi(n)$.
σ_ϕ^2	Variância de $\phi(n)$.
$u_{-1}(n)$	Função degrau unitário.
$\delta(n)$	Delta de Kronecker.
$\mathbf{Q}(-n+1)$	Matriz de função degrau inversa ($P \times P$).
$\mathbf{Q}_o(n)$	Matriz de função de delta de Kronecker ($P+1 \times P$).
$\mathbf{q}(n)$	Vetor de correção do efeito de inicialização ($P \times 1$).
$\boldsymbol{\kappa}(n)$	Vetor associado ao Efeito de Inicialização.
$\tilde{\mathbf{u}}(n)$	Preditor linear de $\mathbf{u}(n)$ sob $\mathbf{Q}(-n+1)$. ($N \times 1$).
$\mathbf{R}_{\phi\tilde{\mathbf{u}}}(n)$	Matriz de correlação entre $\phi(n)$ e $\tilde{\mathbf{u}}(n)$ ($N \times N$).
$\mathbf{R}_{\tilde{\mathbf{u}}\tilde{\mathbf{u}}}(n)$	Matriz de correlação de $\tilde{\mathbf{u}}(n)$ ($N \times N$).
\mathbf{I}_N	Matriz identidade ($N \times N$).
\mathbf{I}_P	Matriz identidade ($P \times P$).
ξ	Erro quadrático médio em regime.
ε	Fator de regularização para problemas de inversão.
β_i	i -ésima constante da equação fechada de $\mathbf{v}(n)$.
γ_i	i -ésima constante dos termos da matriz $\mathbf{K}(n)$.
\mathbf{g}	Memória do Efeito de Inicialização em $\mathbf{v}(n)$.
τ	Memória do Efeito de Inicialização para $k(n)$.
$\mathbf{c}(n)$	Utilizado na derivação da equação fechada para $\mathbf{v}(n)$.
y_i e \mathbf{y}_i	Escalar e vetor usado na fatoração de $\mathbf{K}(n)$.
η	Fator de multiplicação no EQM em regime.
χ	Símbolo para distribuição chi-quadrada.
α	Probabilidade de erro tipo I (Anexo E).
β	Probabilidade de erro tipo II (Anexo E).
λ	Razão entre o valor real e hipotético (Anexo E).
R	Tamanho da amostra utilizada no (Anexo E).

SUMÁRIO

1 INTRODUÇÃO	25
1.1 Motivação	27
1.2 Objetivos do Trabalho	27
1.3 Estrutura da Dissertação	28
1.4 Trabalhos Publicados	29
2 FUNDAMENTAÇÃO TEÓRICA.....	31
2.1 Filtragem Adaptativa Linear Supervisionada.....	31
2.1.1 Algoritmos para Filtragem Adaptativa Linear Supervisionada	41
2.1.1.1 Algoritmos LMS e NLMS	42
2.2 O Processo Autoregressivo	44
2.3 O Algoritmo de Projeções Afins e o Pseudo APA.....	47
2.3.1 O APA como a Solução de um Sistema Linear Subdeterminado.....	47
2.3.2 O Pseudo Algoritmo de Projeções Afins - PAP	49
2.3.3 Propriedades do Algoritmo de Projeções Afins Convencional.....	52
2.3.4 Custo computacional dos algoritmos APA e PAP	53
3 DEFINIÇÃO DO PROBLEMA	55
3.1 Introdução ao Problema	55
3.2 Considerações para a Análise do PAP	55
3.2.1 Hipóteses Simplificadoras e Estatísticas	56
3.3 O Efeito de Inicialização do PAP	58
3.4 A Relação entre APA e PAP.....	65
3.5 Conclusão	70
4 COMPORTAMENTO MÉDIO DOS COEFICIENTES.....	71
4.1 Análise do Comportamento Médio dos Coeficientes.....	71
4.2 Forma Fechada para o Vetor de Erro dos Coeficientes.....	75
4.3 Conclusão	78
5 ERRO QUADRÁTICO MÉDIO.....	79
5.1 Erro Quadrático Médio em Regime Transiente	79
5.2 Momentos de Segunda Ordem em Regime Transiente	88
5.2.1 Avaliação dos Momentos de Segunda Ordem	88
5.2.2 Traço da Equação Recursiva dos Momentos de Segunda Ordem	102
5.2.3 Forma Fechada para $k(n)$.....	103
5.3 Comportamento em Regime Permanente	105
5.3.1 Desvio Quadrático Médio.....	106
5.3.2 Erro Quadrático Médio em Regime Permanente.....	106
5.3.3 Erro Quadrático Médio em Excesso.....	107
5.4 Conclusão	108

6 RESULTADOS E DISCUSSÕES	109
6.1 Condições de Teste	109
6.2 O Efeito de Inicialização	111
6.3 Comportamento Médio dos Coeficientes.....	113
6.4 Erro Quadrático Médio	118
6.5 Conclusão	128
7 CONCLUSÕES E TRABALHO FUTURO.....	129
7.1 Propostas para Continuação do Trabalho	130
REFERÊNCIAS.....	131
APÊNDICE A – ESTIMAÇÃO DOS PARÂMETROS DE UM PROCESSO AUTOREGRESSIVO ATRAVÉS DE PREDITOR LINEAR PELO MÉTODO DOS MÍNIMOS QUADRADOS	134
APÊNDICE B – MATRIZ DE CORRELAÇÃO $R_{\phi_{\tilde{u}}}(n)$.....	140
APÊNDICE C – MATRIZ DE CORRELAÇÃO $R_{\tilde{u}}(n)$.....	142
ANEXO A – INTERPRETAÇÃO GEOMÉTRICA DO APA	143
ANEXO B – DETERMINAÇÃO PARA $E\{y_1y_2y_3y_4\}$.....	146
ANEXO C – DETERMINAÇÃO PARA $E\{y_1y_2^T w_1w_2^T y_3y_4^T\}$.....	147
ANEXO D – AVALIAÇÃO DA FUNÇÃO DE CORRELAÇÃO DE UM SINAL AUTOREGRESSIVO	149
ANEXO E – INFERÊNCIA ESTATÍSTICA SOBRE A VARIÂNCIA DE UMA POPULAÇÃO.....	152

1 INTRODUÇÃO

Filtros adaptativos lineares são amplamente utilizados em diferentes áreas do conhecimento. Na engenharia, eles são empregados em diversas aplicações de tempo-real como, por exemplo, radar, sonar, controle ativo de ruído, aparelhos auditivos, e cancelamento de eco. O estudo de filtragem adaptativa teve grande impulso com o desenvolvimento do algoritmo *Least Mean-Square* (LMS) em 1959. Apesar de bastante difundido por seu baixo custo computacional, esse algoritmo apresenta uma convergência lenta para sinais de entrada fortemente correlacionados, no caso estacionário. Além disso, variações na potência de entrada podem acarretar significativas variações indesejadas no ajuste dos coeficientes. Para contornar o segundo problema, o algoritmo LMS normalizado (NLMS) realiza uma normalização da entrada através de um mecanismo de controle automático de ganho intrínseco à sua estrutura. Entretanto, sua convergência também pode ser comprometida pela correlação do sinal de entrada. Uma alternativa é a utilização do algoritmo *Recursive Least Squares* (RLS), que apresenta uma maior robustez à dispersão de autovalores da matriz de correlação de entrada. Porém, apresenta um elevado custo computacional e a possibilidade de instabilidade numérica.

O algoritmo de projeções afins (APA) surgiu como uma solução atrativa para reduzir o efeito de correlação do sinal de entrada sobre o processo de adaptação sem graves problemas associados à instabilidade numérica. A melhoria no desempenho do regime transitório resulta de uma maior complexidade computacional e de um aumento no erro quadrático médio em excesso em relação ao NLMS. Em virtude dos avanços tecnológicos e da redução progressiva das limitações à complexidade na implementação de filtros adaptativos, a família APA é hoje empregada em diversas aplicações (GAY e TAVATHIA, 1995) (GONZALEZ et al., 2012). Dentre os algoritmos da família APA, dois se destacam pelo custo computacional reduzido: o *fast affine projection algorithm* (GAY e TAVATHIA, 1995) e o pseudo algoritmo de projeções afins (PAP) (RUPP, 1998).

Em 1984, Ozeki e Umeda (1984) apresentaram uma demonstração geométrica que resultou em um algoritmo de projeções afins de segunda ordem e constataram que para passo unitário, ele

atualizava o vetor de coeficientes na direção ortogonal aos últimos vetores de entrada. Posteriormente, Markus Rupp (1998) demonstrou que o mecanismo dos mínimos quadrados embutido na atualização dos coeficientes desse algoritmo descorrelaciona o sinal de entrada, acelerando a convergência e fazendo com que o algoritmo seja atrativo para aplicações com sinais fortemente correlacionados. Em contrapartida, essa propriedade pode levar a um aumento na potência do ruído de fundo em regime permanente, em comparação com o algoritmo NLMS. Bouitelle et al. (1999), por sua vez, estendeu o algoritmo de Ozeki e Umeda para a aplicação de cancelamento de eco, denominando-o de pseudo algoritmo de projeções afins. Apesar de não haver evidências de alguma relação na derivação desses trabalhos, o algoritmo analisado por Rupp (1998) e o apresentado por Bouitelle et al. (1999) são equivalentes.

Em virtude da crescente utilização da família APA, diversos trabalhos vêm sendo publicados sobre o seu comportamento médio, em regime transitório e permanente, estabilidade e custo computacional (BERSHAD; LINEBARGER e MCLAUGHLIN, 2001) (ALMEIDA et al., 2005) (SHIN e SAYED, 2004) (RUPP, 2011). Porém, o completo entendimento dessa família ainda é um grande desafio em virtude do conjunto subdeterminado de restrições embutido em suas equações, que torna sua análise significativamente mais complexa. Shin e Sayed (2004) propuseram um modelo para o APA convencional nos regimes transitório e permanente, a partir do argumento de conservação de energia. Apesar de o modelo para desempenho no transiente ser aplicável a qualquer tipo de sinal de excitação, é necessário conhecer *a priori* as suas estatísticas e, portanto, nenhuma forma fechada é disponibilizada.

Baseado na análise determinística de Rupp (1998), Bershada, Linebarger e Mclaughlin (2001) propuseram um modelo analítico para o erro quadrático médio do APA no caso de sinais autoregressivos (AR). Além disso, considerou-se o uso de passo unitário (máxima velocidade de convergência), situação em que o APA e PAP apresentariam o mesmo desempenho, segundo Bouitelle, Scalart e Corazza (1999). Este modelo foi posteriormente incrementado por Almeida et al. (2005), resultando em equações recursivas para a descrição do comportamento médio e quadrático médio do APA, assumindo-se um número elevado de coeficientes. Essa modelagem resultou em um aumento na acurácia das predições em relação à Bershada, Linebarger e Mclaughlin (2001).

1.1 Motivação

O uso extensivo da teoria desenvolvida por Almeida et al. (2005) evidenciou imprecisões do modelo nas iterações iniciais do pseudo algoritmo de projeções afins para certas configurações de parâmetros. Posteriormente, resultados preliminares, publicados em Costa et al. (2012a), demonstraram a necessidade de extensão em uma relação determinística apresentada em Rupp (1998), empregada para todos os estágios do processo de adaptação nos modelos desenvolvidos por Bershada, Linebarger e McLaughlin (2001) e Almeida et al. (2005). Essa extensão visa modelar o efeito gerado pela inicialização arbitrária do algoritmo em relação aos coeficientes ótimos, e que acarreta em uma discrepância no comportamento do PAP em relação ao APA, mesmo sob a restrição de passo unitário.

1.2 Objetivos do Trabalho

Os principais objetivos deste trabalho são:

- Estudar o efeito presente na inicialização do PAP e sua relação com o APA;
- Modelar matematicamente o mecanismo determinístico associado a esse efeito na inicialização do PAP;
- Obter equações determinísticas recursivas e na forma fechada para as figuras de mérito de comportamento médio dos coeficientes e de erro quadrático médio, visando incrementar a acurácia do modelo de Almeida (2004) no regime transitório;
- Estudar o impacto do Efeito de Inicialização durante os períodos de transiente e regime permanente;
- Apresentar resultados de simulações que corroborem a validade e acurácia do novo modelo.

1.3 Estrutura da Dissertação

O Capítulo 2 aborda os fundamentos teóricos necessários ao desenvolvimento deste trabalho, revisitando a filtragem adaptativa linear, o processo autoregressivo, e o algoritmo de projeções afins e sua variante de baixo custo computacional (PAP).

Posteriormente, o Capítulo 3 dedica-se à definição do problema. Inicialmente, são introduzidas as limitações, hipóteses e relações determinísticas necessárias para a análise estocástica do PAP. Com a apresentação do problema, é derivada uma nova relação determinística que permite modelar com mais acurácia o comportamento do PAP. A Seção 3.4 também descreve o Efeito de Inicialização e sua contribuição na relação entre os algoritmos APA e PAP.

O comportamento do PAP à luz da nova expressão matemática é o assunto dos próximos dois capítulos: o Capítulo 4 descreve o modelo para o comportamento médio dos coeficientes, resultando em uma equação determinística recursiva e uma equação fechada que possibilitam a compreensão das consequências do Efeito de Inicialização ao longo da evolução dos coeficientes; o Capítulo 5 trata do erro quadrático médio e dos momentos de segunda ordem do vetor de erro dos coeficientes do PAP em regime transiente e permanente.

Os resultados de simulações dos modelos teóricos desenvolvidos são apresentados no Capítulo 6, onde são comparados com o modelo obtido em Almeida (2005), utilizando como referência a simulação Monte Carlo do PAP. Discussões sobre os resultados obtidos também são apresentadas nesse capítulo. Finalmente, o Capítulo 7 apresenta as conclusões finais e sugestões para a continuidade deste trabalho.

A seguinte notação é utilizada neste documento: grandezas escalares são denotadas por letras maiúsculas e minúsculas em itálico, vetores são denotados por letras minúsculas em negrito e matrizes por letras maiúsculas em negrito.

1.4 Trabalhos Publicados

Artigos publicados em congresso nacional e internacional até o presente momento, e cujos conteúdos fazem parte desta dissertação:

1. COSTA, M. H.; ALMEIDA, S. J. M.; BERMUDEZ, J. C. M.; BARCELOS, R. B.; New Insights into the Weight Behaviour of the Affine Projection Algorithm. In: Proc. European Signal Process. Conf. Bucharest: 2012a. p. 2610-2614.
2. COSTA, M. H.; ALMEIDA, S. J. M.; BERMUDEZ, J. C. M.; BARCELOS, R. B.; Efeito de Inicialização no Modelo do Comportamento dos Coeficientes do Algoritmo de Projeções Afins. In: XXX Simpósio Brasileiro de Telecomunicações (SBrT'12). Brasília: 2012b.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentados conceitos que visam fundamentar o entendimento deste trabalho. Ele está organizado da seguinte forma: na Seção 2.1 são introduzidos conceitos de filtragem adaptativa linear supervisionada. A Seção 2.2 apresenta o processo autoregressivo (AR), visando embasar a estimação de parâmetros AR através de um preditor linear pelo método dos mínimos quadrados, presente no Apêndice A, peça fundamental do mecanismo de adaptação do pseudo algoritmo de projeções afins para entrada autoregressiva. Finalmente, na Seção 2.3, o algoritmo de projeções afins e sua versão de baixo custo computacional - o pseudo algoritmo de projeções afins - são detalhados com base nos conceitos anteriormente apresentados.

2.1 Filtragem Adaptativa Linear Supervisionada

O processamento adaptativo de sinais teve um grande impulso a partir do final da década de 60 através do desenvolvimento do algoritmo adaptativo *Least Mean-Squares* (LMS). O termo *adaptativo* refere-se a um sistema cuja função é ajustar iterativamente seus parâmetros internos (geralmente em tempo real) com o objetivo de cumprir determinados critérios de desempenho aceitáveis pela aplicação, que dependem do estado do sistema, assim como dos elementos que o compõem (FARHANG-BOROJENY, 1999, p. 1).

Segundo Costa (2001, p.2, apud WIDROW e STEARNS (1985)), os sistemas adaptativos possuem todas ou algumas das seguintes características:

- adaptação automática à medida que ocorre a modificação do ambiente e/ou alterações do sistema (auto-otimização);
- podem ser treinados para desenvolver uma tarefa específica de filtragem ou decisão, ou seja, podem ser programados através de um processo de treinamento (autoprogramáveis);
- em decorrência do item anterior, não necessitam procedimentos elaborados de síntese, são basicamente autoprotetáveis;

- podem extrapolar o espaço de conhecimento e lidar com novas situações após o treinamento com um pequeno conjunto de padrões de entrada (auto-aprendizado)
- até certo ponto podem reparar a si mesmos, ou seja, podem adaptar-se em regiões próximas da ótima mesmo quando sujeitos a certos tipos de defeitos ou limitações (auto-regeneráveis);
- geralmente são mais complexos e difíceis de analisar que sistemas não-adaptativos, mas oferecem a possibilidade de um desempenho substancialmente melhor quando as características do ambiente são desconhecidas ou variantes no tempo.

Em sua essência, os filtros adaptativos são sistemas variantes no tempo e não-lineares, visto que dependem do sinal de entrada e não obedecem ao *Princípio da Superposição* (HAYKIN, 1996, p. 3). Entretanto, um filtro adaptativo é dito linear se a sua saída é uma combinação linear das observações na entrada. Sistemas adaptativos com essa característica são mais fáceis de tratar matematicamente.

Estruturalmente, filtros lineares podem ser divididos segundo a sua resposta ao impulso em resposta finita ao impulso (FIR, do inglês *Finite Impulse Response*) e resposta infinita ao impulso (IIR, do inglês *Infinite Impulse Response*). Filtros IIR geralmente possuem uma seletividade maior em relação aos filtros FIR, e conseqüentemente necessitam de um número menor de coeficientes para uma mesma função de transferência, devido à localização dos pólos (ANTONIOU, 2005, p. 554). Entretanto, existe a possibilidade de ocorrência de mínimos locais na superfície de desempenho, baixa velocidade de convergência e maior possibilidade de instabilidade durante o processo de adaptação (ANTONIOU, 2005, p. 871).

Em contrapartida, os filtros FIR possuem resposta ao impulso de comprimento finito e geralmente não apresentam recursividade. São amplamente utilizados em virtude das suas características intrínsecas de estabilidade (os pólos estão na origem, portanto dentro da circunferência de raio unitário do plano Z) e facilidade de implementação. Além disso, apresentam uma menor sensibilidade aos erros de arredondamento, em virtude da ausência de recursividade, e podem ter fase linear e não mínima, e atraso de grupo constante (SHENOI, 2006, p. 267).

A definição da estrutura de realização é um dos requisitos mais importantes da etapa de especificação de um filtro seletivo ou adaptativo. Os filtros lineares podem assumir diferentes tipos de estruturas de realização a partir de uma equação de diferenças, como por exemplo as formas direta I e II. Além disso, em algumas aplicações como conformação de feixe (*beamforming*) e cancelamento de ruído, a saída pode ser obtida através de uma combinação linear entre diferentes sinais de excitação em paralelo, recebendo a denominação de combinador linear (FARHANG-BOROUJENY, 1999, p. 4). A Figura 1 apresenta a estrutura direta I de um filtro FIR linear, causal e invariante no tempo.

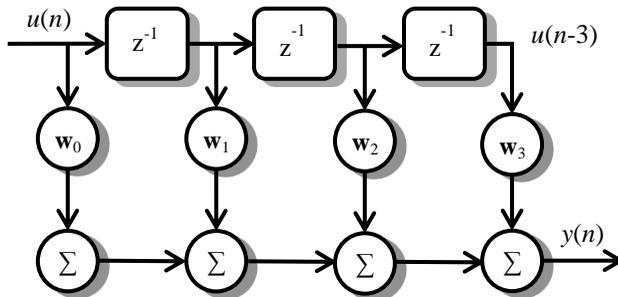


Figura 1 – Realização transversal de um filtro FIR ($N = 4$).

A estrutura de realização direta, também denominada transversal ou em linha de retardo (*tapped-delay line*), envolve a combinação de três operações básicas (HAYKIN, 1996, p. 204):

1. o armazenamento, representado pela cascata de $N-1$ atrasos de uma amostra de um sinal de excitação $u(n)$;
2. o produto interno escalar entre as amostras do sinal de entrada e as da resposta ao impulso do sistema (coeficientes);
3. a operação de adição, que consiste na soma dos resultados das multiplicações definidas anteriormente para produzir a saída do filtro $y(n)$.

Dessa forma, a estrutura de realização transversal da Figura 1 é expressa pela equação de diferenças não recursiva (pólos na origem),

$$y(n) = \sum_{k=0}^{N-1} u(n-k)w_k \quad (2.1)$$

e representada na forma vetorial pelo produto interno

$$y(n) = \mathbf{u}^T(n)\mathbf{w} \quad (2.2)$$

sendo que $y(n)$ é a saída do filtro, w_k são os parâmetros internos que modelam a resposta ao impulso do filtro, $\mathbf{w} = [w_0 \ w_1 \ \dots \ w_{N-1}]^T$ é o vetor de parâmetros e o vetor de sinal de excitação de entrada $\mathbf{u}(n)$ consiste em uma janela das N últimas amostras do sinal de entrada, de forma que $\mathbf{u}(n) = [u(n) \ u(n-1) \ \dots \ u(n-N)]^T$. Esse último vetor geralmente é denominado de *regressor* na literatura específica.

Neste trabalho, denomina-se de *estrutura FIR transversal*, a estrutura de realização transversal de um filtro FIR linear e causal, conforme Figura 1, podendo ser variante ou invariante no tempo, dependendo apenas do contexto em que for aplicado (filtragem adaptativa ou seletiva, respectivamente).

A Figura 2 apresenta um problema clássico de filtragem utilizando uma estrutura transversal com coeficientes constantes no qual se deseja minimizar a diferença entre dois processos aleatórios, o sinal na saída do filtro $y(n)$ e o desejado $d(n)$. O erro instantâneo entre ambos pode ser expresso por

$$e(n) = d(n) - y(n). \quad (2.3)$$

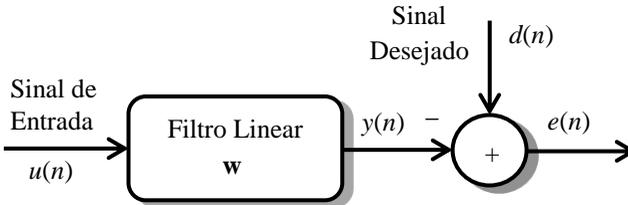


Figura 2 – Diagrama de blocos de estimação linear.

O processo de minimização pode ser realizado sobre os valores instantâneos ou em uma função destes, usualmente denominada *função de custo* ou *superfície de desempenho*. Uma alternativa bastante utilizada, em virtude das suas características de inexistência de mínimos locais e tratabilidade matemática, é o erro quadrático médio (EQM). O EQM pode ser obtido substituindo-se (2.2) em (2.3), elevando o resultado ao quadrado e tomando o seu valor esperado, resultando em:

$$\begin{aligned} E \{e^2(n)\} &= E \{d^2(n) - 2d(n)y(n) + y^2(n)\} \\ &= E \{d^2(n)\} - 2E \{d(n)y(n)\} + E \{y^2(n)\} \\ &= E \{d^2(n)\} - 2E \{d(n)\mathbf{u}^T(n)\} \mathbf{w} + \mathbf{w}^T E \{\mathbf{u}(n)\mathbf{u}^T(n)\} \mathbf{w}. \end{aligned} \quad (2.4)$$

Os valores esperados presentes em (2.4) são definidos pelas características estatísticas do sinal desejado e de entrada. Entretanto, médias temporais podem ser usadas para a estimação de $E\{d(n)\mathbf{u}^T(n)\}$ e $E\{\mathbf{u}(n)\mathbf{u}^T(n)\}$, desde que $d(n)$ e $\mathbf{u}(n)$ sejam sinais estacionários no sentido amplo (WSS) e ergódicos. A propriedade de ergodicidade é definida para sinais aleatórios que apresentem média temporal idêntica à média do conjunto de medições, ou *ensemble* (FARHANG-BOROUJENY, 1999, p. 46). Assim,

$$E\{e^2(n)\} = \sigma_d^2 - 2\mathbf{p}_{du}^T \mathbf{w} + \mathbf{w}^T \mathbf{R}_u \mathbf{w} \quad (2.5)$$

em que $\sigma_d^2 = E\{d^2\}$,

$$\mathbf{R}_u = E\{\mathbf{u}(n)\mathbf{u}^T(n)\} \quad (2.6)$$

é a matriz de correlação do vetor de entrada e

$$\mathbf{p}_{du} = E\{d(n)\mathbf{u}(n)\} \quad (2.7)$$

é o vetor de correlação cruzada entre o vetor de entrada e a resposta desejada.

A matriz de correlação \mathbf{R}_u desempenha um papel expressivo na análise de filtros lineares seletivos e adaptativos, como por exemplo, aferir a velocidade de convergência e estabilidade. Segundo Haykin (1996, p. 101), três propriedades de \mathbf{R}_u podem ser definidas em um processo estocástico com estacionaridade assintótica no sentido amplo:

- *simetria*: para o caso de sinais reais, a matriz é idêntica à sua transposta, $\mathbf{R}_u = \mathbf{R}_u^T$;
- *estrutura Toeplitz*: todos elementos em qualquer diagonal paralela à diagonal principal possuem valores constantes;
- *ser positiva semi-definida*: desde que sejam satisfeitos os requisitos de estacionaridade assintótica, a matriz de correlação será positiva semi-definida, $\mathbf{v}^T \mathbf{R}_u \mathbf{v} \geq 0$. Na prática pode ser considerada positiva definida (HAYKIN, 2002).

O uso do critério de erro quadrático médio como função de custo para o projeto do filtro linear seletivo apresentado na Figura 2 requer o conhecimento *a priori* das estatísticas conjuntas de segunda ordem dos sinais WSS de excitação e desejado. Através de (2.5) verifica-se que $E\{e^2(n)\}$ é uma função quadrática dos coeficientes do filtro (HAYKIN, 1996, p. 207). Outras funções de custo podem ser utilizadas como, por exemplo, o valor absoluto ou a *hit-or-miss*, porém podem resultar em filtros (•estimadores) mais complexos e de difícil realização. As

seguintes vantagens podem ser listadas para a utilização da superfície de desempenho segundo o EQM:

- a função de custo é um parabolóide com N graus de liberdade e apresenta apenas um único mínimo global, sem ocorrência de mínimos locais, já que corresponde a uma superfície quadrática convexa devido à matriz de correlação ser considerada positiva definida (Figura 3);
- o filtro resultante é ótimo se os sinais $\mathbf{u}(n)$ e $d(n)$ apresentarem distribuição Gaussiana e forem ergódicos e estacionários no sentido amplo.

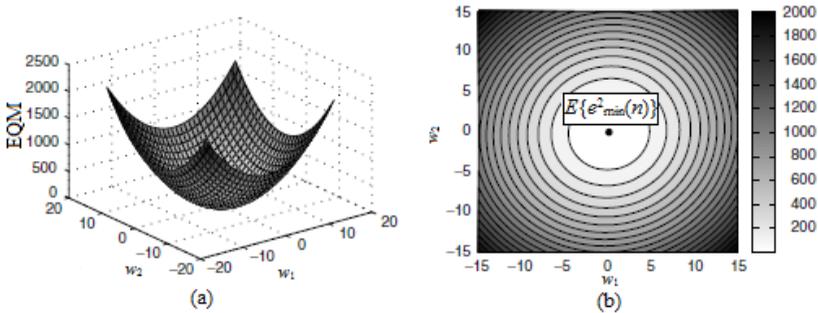


Figura 3 – Exemplo de função de custo parabolóide (Figura 6.4 de (MANOLAKIS; INGLE e KOGON, 2005)).

A solução para o conjunto ótimo de coeficientes segundo o critério do mínimo erro quadrático médio é denominada solução de Wiener-Hopf (HAYKIN, 1996, p. 206) em decorrência do trabalho do matemático americano Norbert Wiener (1894-1964), posteriormente reformulado para o tempo discreto por Norman Levinson (1912-1975). No ponto de mínimo dessa superfície, a função de custo atinge seu menor valor, denotado por $E\{e^2_{\min}(n)\}$. O vetor de gradiente neste ponto, que consiste na derivada parcial de primeira ordem com respeito aos coeficientes, é zero para todos os elementos do vetor,

$$\nabla E\{e^2_{\min}(n)\} = \mathbf{0}. \quad (2.8)$$

Aplicando-se a derivada parcial em (2.5), resulta em

$$\nabla E\{e^2(n)\} = \frac{\partial E\{e^2(n)\}}{\partial \mathbf{w}} = -2\mathbf{p}_{du} + 2\mathbf{R}_u \mathbf{w} \quad (2.9)$$

e inserindo as restrições de (2.8), obtém-se a equação de Wiener-Hopf para os coeficientes ótimos, dada por

$$\mathbf{p}_{du} = \mathbf{R}_u \mathbf{w}^o. \quad (2.10)$$

Pré-multiplicando a equação (2.10) por \mathbf{R}_u^{-1} , tem-se que

$$\mathbf{w}^o = \mathbf{R}_u^{-1} \mathbf{p}_{du} \quad (2.11)$$

sendo necessário que \mathbf{R}_u seja não singular (possua inversa).

A equação (2.11) é obtida considerando o caso particular de estrutura FIR transversal. A forma geral para as equações de Wiener-Hopf é obtida através do *princípio da ortogonalidade*, que estabelece que o filtro é dito ótimo no sentido do erro quadrático médio, quando o vetor de coeficientes \mathbf{w} é tal que o erro $e(n)$ é estatisticamente ortogonal ao vetor de entrada (HAYKIN, 1996, p. 197), ou seja,

$$E \{e(n)u(n-k)\} \Big|_{0 \leq k < N} = 0 \quad (2.12)$$

Apesar de ótima no sentido quadrático médio, a solução de Wiener-Hopf apresenta alguns inconvenientes:

- possui custo computacional elevado para um número grande de coeficientes, o que pode inviabilizar algumas aplicações em tempo real;
- pode resultar em problemas de estabilidade, devido à inversão da matriz de correlação;
- em ambiente não-estacionário, a teoria de filtragem ótima geralmente é sub-ótima, devido à dificuldade intrínseca de conhecimento *a priori* das propriedades estatísticas dos sinais envolvidos no sistema.

De forma a evitar os problemas mencionados, diferentes estratégias iterativas podem ser empregadas. Entre os métodos iterativos de otimização existentes, o de *gradiente de descida mais íngreme* (*steepest-descent*) é bastante difundido por sua simplicidade, e consiste em um procedimento recursivo para encontrar o valor de mínimo de uma função de custo.

Considerando-se a estrutura FIR transversal da Figura 1 e a função de custo de (2.5), a equação de atualização dos coeficientes no instante posterior ($n+1$) é obtida através da seguinte relação (HAYKIN, 1996, p. 342):

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{1}{2} \mu \left[-\nabla E \{e^2(n)\} \right] \quad (2.13)$$

em que μ é o passo de ajuste do algoritmo e o valor $\frac{1}{2}$ visa a simplificação posterior de uma constante associada ao gradiente $\nabla E\{e^2(n)\}$. Substituindo-se a equação (2.9) em (2.13), obtém-se

$$\begin{aligned}\mathbf{w}(n+1) &= \mathbf{w}(n) + \mu \left[\mathbf{p}_{du} - \mathbf{R}_u \mathbf{w}(n) \right] \\ &= \left[\mathbf{I}_N - \mu \mathbf{R}_u \right] \mathbf{w}(n) + \mu \mathbf{p}_{du}.\end{aligned}\quad (2.14)$$

Nota-se que a equação (2.14) utiliza as estatísticas dos sinais de entrada, mas sem a necessidade de inversão de \mathbf{R}_u .

Na implementação, o projetista deve indicar a condição inicial do vetor de coeficientes $\mathbf{w}(0)$ em um ponto qualquer da superfície. O algoritmo utiliza o resultado do gradiente deste ponto para calcular o novo vetor de coeficientes na direção da solução ótima $E\{e^2_{\min}(n)\}$ (e oposta ao gradiente). Desde que sejam cumpridos os requisitos de estabilidade do passo, este procedimento é repetido até chegar à solução ótima. Nas iterações iniciais, o algoritmo efetua uma maior redução na função de erro por iteração. Entretanto, conforme os elementos do gradiente tornam-se menores, devido à proximidade em relação à solução ótima, o processo de adaptação tende a ficar consideravelmente mais lento. Se há uma grande dispersão de autovalores na matriz de correlação do sinal de entrada, então o ponto de solução tende a seguir uma trajetória em zig-zag no espaço dos parâmetros e o desempenho do algoritmo tende a se deteriorar (ANTONIOU, 2005, p. 868).

Uma alternativa para reduzir o número de iterações necessárias para que o algoritmo convirja até o ponto de mínimo é o emprego do *método de Newton*. A equação de atualização dos coeficientes resultante é obtida através de uma aproximação da superfície de desempenho apresentada em (2.5), resultando em (SAYED, 2003, p. 191):

$$\begin{aligned}\mathbf{w}(n+1) &= \mathbf{w}(n) + \mu \mathbf{R}_u^{-1} \left[\mathbf{p}_{du} - \mathbf{R}_u \mathbf{w}(n) \right] \\ &= \left[1 - \mu \right] \mathbf{w}(n) + \mu \mathbf{R}_u^{-1} \mathbf{p}_{du}.\end{aligned}\quad (2.15)$$

Considerando-se a utilização do passo unitário em (2.15), tem-se que apenas uma iteração é necessária para que o algoritmo chegue à solução ótima. Salienta-se que essa melhoria na velocidade de convergência apresenta a necessidade de inversão da matriz de correlação \mathbf{R}_u , aumentando a complexidade computacional e possibilitando problemas de instabilidade. As equações (2.14) e (2.15) podem ser vistas como modelos para gerar vários algoritmos adaptativos a partir de aproximações estocásticas de \mathbf{R}_u e \mathbf{p}_{du} .

De forma a superar a necessidade do conhecimento *a priori* das estatísticas dos sinais envolvidos. Os filtros adaptativos supervisionados aparecem como uma solução atrativa, realizando o aprendizado iterativo das características estatísticas dos sinais envolvidos. Classifica-se como *supervisionado*, o filtro adaptativo que tem acesso ao sinal desejado $d(n)$ e utiliza-o como referência. Conforme a estrutura básica apresentada na Figura 4, o filtro adaptativo supervisionado consiste na combinação de dois processos: um processo de filtragem variante no tempo e uma estratégia de adaptação dos parâmetros do filtro.

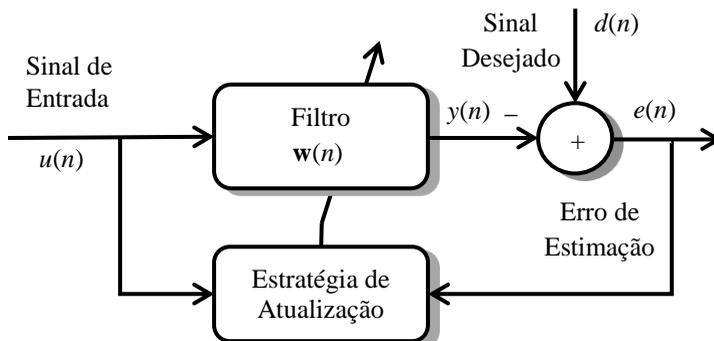


Figura 4 – Diagrama de blocos de um filtro adaptativo supervisionado.

Existem basicamente quatro classes de aplicações para filtros adaptativos supervisionados (MANOLAKIS; INGLE e KOGON, 2005, p. 17):

1. *Identificação (ou modelamento) de sistema*: o filtro adaptativo é usado para modelar uma planta desconhecida através de critérios específicos. Conforme a Figura 5.a, o sinal de excitação $u(n)$ é aplicado na entrada da planta e do filtro adaptativo. O sinal de excitação de entrada é livre de ruído, enquanto o sinal desejado $d(n)$ é corrompido por interferências descorrelacionadas com o sinal de excitação. O objetivo é ajustar os coeficientes da estrutura utilizada de forma que a saída do filtro se aproxime da resposta desejada de acordo com o critério de desempenho definido. Entre as aplicações possíveis encontram-se o cancelamento de eco, o modelamento de canal e a identificação de sistemas de controle (MANOLAKIS; INGLE e KOGON, 2005, p. 17);

2. *Modelamento inverso*: o objetivo é a obtenção do recíproco da função de transferência de uma planta desconhecida. Uma versão atrasada do sinal de entrada da planta é utilizada como resposta desejada para a filtragem adaptativa (Figura 5.b). Exemplos de aplicações que aplicam o modelamento inverso são: a equalização de canal telefônico, visando reduzir a interferência intersimbólica, e a decorrelação preditiva;
3. *Predição de sinal*: o objetivo desta classe é realizar previsões a respeito da evolução de um processo aleatório com base no seu conhecimento passado (Figura 5.c). Exemplos de aplicações empregadas são o *adaptive differential pulse code modulation* (ADPCM) e a codificação linear preditiva (LPC);
4. *Cancelamento de interferência (ruído)*: tem como objetivo realizar o cancelamento ou redução de um ruído adicionado a um sinal de interesse (sendo a composição destes sinais denominada de sinal primário), a partir de um sinal de referência correlacionado ao ruído (Figura 5.d). Alguns autores também denominam essa classe de melhoramento de sinal (*signal enhancement*). Pode ser empregado no cancelamento de ruído e em antenas inteligentes (*beamforming*) multi-sensoriais.

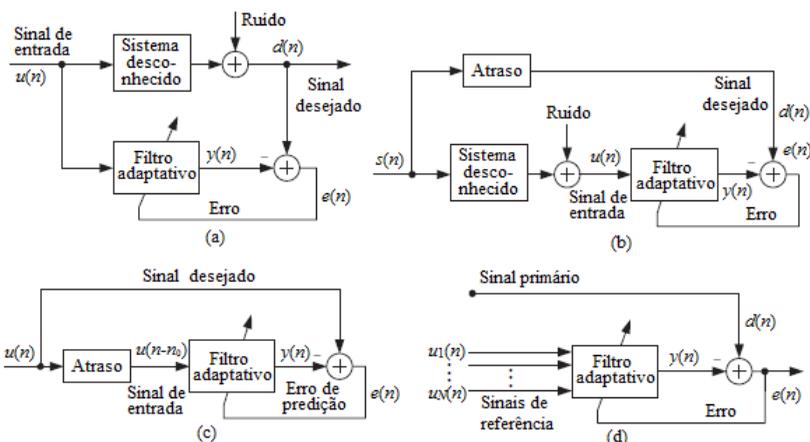


Figura 5 – Classes de aplicações de filtros adaptativos supervisionados: (a) identificação de sistema; (b) modelamento inverso; (c) predição de sinal; (d) cancelamento de interferência multi-sensor.

2.1.1 Algoritmos para Filtragem Adaptativa Linear Supervisionada

Os algoritmos adaptativos para filtros lineares supervisionados possibilitam o ajuste iterativo dos parâmetros do filtro em direção à solução ótima através da minimização da estimação instantânea da superfície de desempenho. Dessa forma, sua complexidade é menor, porém sendo função da estrutura e do critério de desempenho utilizado. Os algoritmos mais difundidos são geralmente derivados a partir de dois métodos (FARHANG-BOROJENY, 1999, p. 6) (HAYKIN, 1996, p. 9):

- **MÉTODO DO GRADIENTE ESTOCÁSTICO:**

Uma forma geralmente empregada na derivação de algoritmos adaptativos para filtros lineares supervisionados é a modificação de métodos iterativos de otimização, como o método do gradiente de descida mais íngreme. Dessa forma, possibilitam o ajuste iterativo dos parâmetros do filtro na direção da solução ótima, através de aproximações empregadas na estimação instantânea da superfície de desempenho. O termo *gradiente estocástico* visa distingui-los do método de otimização pelo gradiente de descida mais íngreme (HAYKIN, 1996, p. 365). O procedimento de atualização nesse caso consiste em:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \begin{pmatrix} \text{Direção de} \\ \text{atualização} \end{pmatrix} \begin{pmatrix} \text{Ponderação da} \\ \text{atualização} \end{pmatrix} \quad (2.16)$$

em que o símbolo μ denota o passo do algoritmo e permite controlar o salto na busca pela solução ótima. O algoritmo *Least Mean-Square* (LMS) é o mais difundido.

- **MÉTODO DOS MÍNIMOS QUADRADOS:**

Enquanto a solução empregada pelo método do gradiente estocástico se baseia em uma formulação estocástica para o problema de filtragem, a solução pelo método dos mínimos quadrados representa uma perspectiva determinística, através da soma dos quadrados do erro das estimativas. Algoritmos baseados nesse método convergem mais rapidamente que o LMS para um sistema invariante no tempo e sinais estacionários, e não são demasiadamente sensíveis à densidade espectral do sinal de excitação. Porém, apresentam uma maior complexidade computacional e podem tornar-se numericamente instáveis (FARHANG-BOROJENY, 1999, p. 8). Um exemplo de algoritmo baseado nesse método é o *Recursive Least-Squares* (RLS), que também pode ser derivado de forma estocástica.

2.1.1.1 Algoritmos LMS e NLMS

Em decorrência da ampla gama de aplicações dos algoritmos LMS e NLMS e de seu uso como padrão na comparação de desempenho com outros filtros adaptativos, apresenta-se a seguir uma discussão sobre ambos, permitindo introduzir o NLMS na análise do APA.

O algoritmo LMS foi desenvolvido por Widrow e Hoff em 1959 na Universidade de Stanford, sendo o mais difundido algoritmo adaptativo devido ao seu baixo custo computacional. Pertence à classe de algoritmos baseados no método do gradiente estocástico.

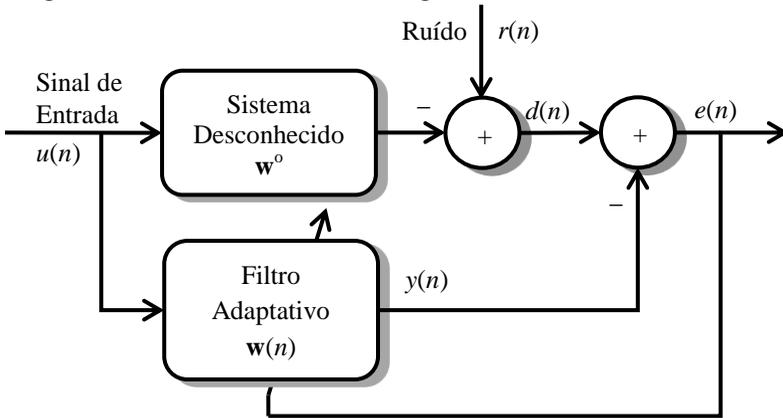


Figura 6 – Diagrama de blocos de um problema de identificação de sistema.

Dado o diagrama de blocos do problema de identificação de sistema apresentado na Figura 6, o sinal desejado $d(n)$ é representado pelo produto interno da resposta ao impulso do sistema desconhecido $\mathbf{w}^o = [w_0^o \ w_1^o \ \dots \ w_{N-1}^o]^T$ (de tal forma que $\mathbf{w}(n) = \mathbf{w}^o$ é ótimo no sentido de Wiener) com o regressor de entrada $\mathbf{u}(n) = [u(n) \ u(n-1) \ \dots \ u(n-N+1)]^T$, adicionado por um distúrbio $r(n)$, de tal forma que

$$d(n) = \mathbf{u}^T(n)\mathbf{w}^o + r(n). \quad (2.17)$$

Este distúrbio é representado por um processo aleatório do tipo ruído branco, de média zero e variância σ_r^2 , independente do sinal de entrada.

O sinal $y(n)$ corresponde à saída de um filtro FIR transversal variante no tempo expresso por

$$y(n) = \mathbf{u}^T(n)\mathbf{w}(n) \quad (2.18)$$

em que $\mathbf{w}(n) = [w_0(n) \ w_1(n) \ \dots \ w_{N-1}(n)]^T$ é o vetor de coeficientes, representando a resposta ao impulso do filtro. A estrutura do filtro linear

é similar à apresentada na Figura 1, porém com coeficientes variantes no tempo. Como já descrito anteriormente, o erro instantâneo de estimação do processo de filtragem é obtido através da diferença entre o sinal de saída do filtro adaptativo e o sinal desejado,

$$e(n) = d(n) - y(n). \quad (2.19)$$

Dessa forma, a função de custo segundo o critério do erro quadrático médio para o sistema da Figura 6 é obtida substituindo-se (2.17) e (2.18) em (2.19), elevando-se ao quadrado e tomando-se o valor esperado de ambos os lados da expressão, resultando em:

$$E \{e^2(n)\} = \sigma_r^2 + E \left\{ \left(\mathbf{u}^T(n) [\mathbf{w}(n) - \mathbf{w}^o] \right)^2 \right\}. \quad (2.20)$$

O algoritmo LMS realiza a adaptação dos coeficientes através de uma estimativa do gradiente da superfície de desempenho de (2.20). Nesse sentido, ao invés do erro quadrático médio, é utilizado o erro instantâneo,

$$\nabla E \{e^2(n)\} \cong \nabla e^2(n). \quad (2.21)$$

e aplicando-se o gradiente no erro instantâneo de (2.19), resulta em

$$\nabla e^2(n) = 2e(n) \frac{\partial e(n)}{\partial \mathbf{w}(n)} = -2e(n) \mathbf{u}(n) \quad (2.22)$$

Substituindo-se (2.22) em (2.13), tem-se a seguinte equação de atualização dos coeficientes:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \mathbf{u}(n) e(n). \quad (2.23)$$

Dessa forma, o LMS pode ser visto como uma aproximação do algoritmo de descida mais íngreme, usando apenas os sinais de entrada disponíveis no sistema a cada iteração. Uma limitação conhecida do LMS é que variações na potência de $u(n)$ levam a variações significativas no ajuste dos coeficientes, podendo acarretar em instabilidade. A amenização desse problema é provida pelo algoritmo NLMS.

Em Sayed (2003, p. 225), o algoritmo LMS normalizado (NLMS) é apresentado como uma aproximação instantânea do método de Newton. Utilizando-se as seguintes aproximações:

$$\mathbf{R}_u \cong \mathbf{u}(n) \mathbf{u}^T(n) \quad (2.24)$$

$$\mathbf{p}_{du} \cong \mathbf{u}(n) d(n) \quad (2.25)$$

em (2.15), e aplicando-se um fator de regularização ε para evitar alguma instabilidade na inversão da matriz $\mathbf{u}(n) \mathbf{u}^T(n)$, tem-se que

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \left[\varepsilon \mathbf{I}_N + \mathbf{u}(n)\mathbf{u}^T(n) \right]^{-1} \mathbf{u}(n)e(n). \quad (2.26)$$

A recursão dos coeficientes requer a inversão de $[\varepsilon \mathbf{I}_N + \mathbf{u}(n)\mathbf{u}^T(n)]$. Porém, como essa matriz é uma modificação de posto um de um múltiplo de uma matriz identidade, sua inversa tem uma estrutura similar (SAYED, 2003, p. 225). Aplicando-se o lema de inversão de matrizes expresso por

$$\left[\mathbf{A} + \mathbf{BCD} \right]^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B} \left[\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B} \right]^{-1} \mathbf{DA}^{-1} \quad (2.27)$$

tem-se que

$$\left[\varepsilon \mathbf{I}_N + \mathbf{u}(n)\mathbf{u}^T(n) \right]^{-1} = \varepsilon^{-1} \mathbf{I}_N - \frac{\varepsilon^{-2}}{1 + \varepsilon^{-1} \mathbf{u}^T(n)\mathbf{u}(n)} \mathbf{u}(n)\mathbf{u}^T(n) \quad (2.28)$$

em que a expressão no lado direito de (2.28) é uma modificação de posto um de $\varepsilon^{-1} \mathbf{I}_N$. Multiplicando ambos os lados por $\mathbf{u}(n)$, resulta em

$$\left[\varepsilon \mathbf{I}_N + \mathbf{u}(n)\mathbf{u}^T(n) \right]^{-1} \mathbf{u}(n) = \frac{\mathbf{u}(n)}{\varepsilon + \mathbf{u}^T(n)\mathbf{u}(n)} \quad (2.29)$$

e assim, a recursão para os coeficientes do algoritmo NLMS é dada por

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \frac{\mathbf{u}(n)}{\varepsilon + \mathbf{u}^T(n)\mathbf{u}(n)} e(n). \quad (2.30)$$

Dessa forma, a normalização de $\mathbf{u}(n)$ é efetuada através do denominador do algoritmo NLMS. Uma abordagem alternativa para sua derivação através de otimização local pelo método dos mínimos quadrados é apresentada posteriormente, na Seção 2.3, consistindo em uma análise introdutória para o algoritmo de projeções afins.

2.2 O Processo Autoregressivo

O processo autoregressivo (AR) é um tipo de processo aleatório largamente empregado para modelar ou estimar uma série temporal associada a algum fenômeno, sendo empregado em diferentes áreas, como engenharia, biologia, economia, entre outras. Consiste basicamente na saída de um filtro recursivo IIR causal, estável, do tipo *all-poles*, expresso pela equação,

$$u(n) = \sum_{k=1}^M a_k u(n-k) + z(n) \quad (2.31)$$

sendo a_k os coeficientes do filtro, $z(n)$ um processo aleatório branco com média zero e variância σ_z^2 , definido como inovação (FARHANG-

BOROUJENY, 1999, p. 386). A denominação autoregressiva deve-se à combinação linear finita de suas amostras passadas $u(n-k)$, $k = 1, \dots, M$.

Uma condição necessária e suficiente para garantir a estacionaridade assintótica no sentido amplo (WSS) de um processo AR é que todos os pólos do filtro estejam contidos dentro da circunferência de raio unitário do plano Z (HAYKIN, 1996, p. 116). Para o caso particular de $M = 1$, tem-se que (MANOLAKIS; INGLE e KOGON, 2005, p. 167):

- quando $a_1 = 0$, $u(n) = z(n)$, resultando em um processo aleatório branco;
- quando $a_1 = 1$, o processo $u(n)$ tende à um processo *random walk*. Conforme a tende ao valor unitário, o processo torna-se cada vez mais correlacionado.

Segundo Haykin (1996, p. 110), o processo AR pode assumir duas configurações, sendo caracterizadas pela forma como a entrada e a saída estão dispostas:

- *Gerador de processo (síntese)*: aplicando-se um processo aleatório branco $z(n)$ na entrada do filtro IIR *all-poles* da Figura 7.a, assumindo-se um determinado conjunto de coeficientes que configure um sistema estável, a saída do sistema $u(n)$ será um processo AR. Uma aplicação de um gerador é fornecer a estimativa espectral de potência de um processo randômico, sendo que um modelo com ordem entre 5 e 10 é suficiente para obter uma estimativa aceitável da densidade de espectro de potência para a grande maioria dos processos encontrados na natureza (FARHANG-BOROUJENY, 1999, p. 387);
- *Analizador de processo (análise)*: Consiste no recíproco da configuração anterior. Aplicando-se o processo AR mencionado no item anterior na entrada do filtro FIR *all-zeros* da Figura 7.b produz-se na saída um processo aleatório branco $z(n)$. Os zeros (coeficientes) deste filtro localizam-se exatamente no mesmo local que os pólos do filtro gerador. O analisador pode ser utilizado como filtro de predição linear de erro com fase mínima e, portanto, é inerentemente estável

(HAYKIN, 1996, p. 273). Uma aplicação possível consiste na estimação espectral usada para modelar processos aleatórios correlacionados, visando encontrar os parâmetros deste processo. Por exemplo, pode ser usado na modelagem do trato vocal para reproduzir o processo natural de produção de fala.

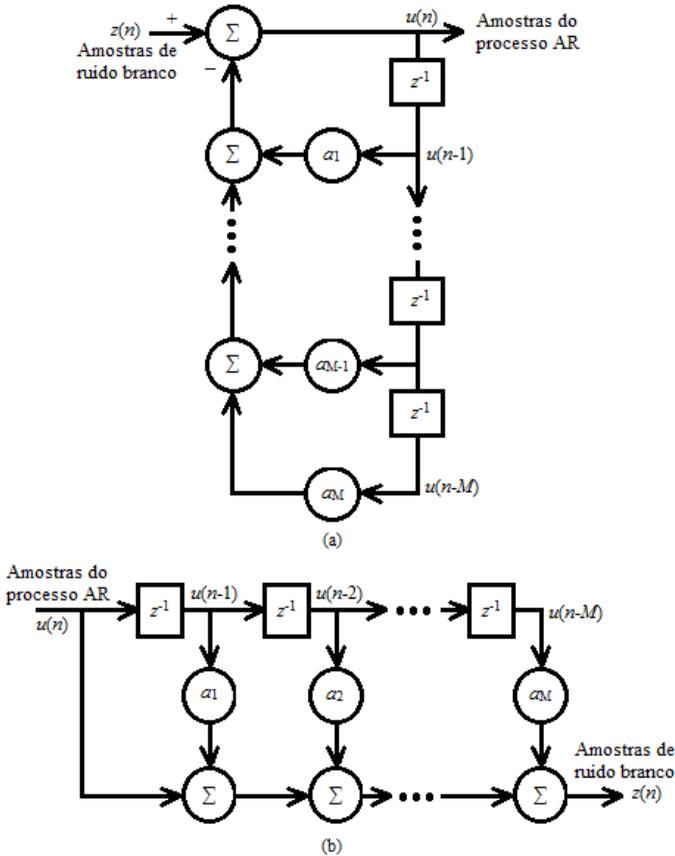


Figura 7 – Modelo de (a) gerador e (b) analisador de processo (Figura 2.2 de (HAYKIN, 1996)).

O modelo do sinal de excitação de entrada deste trabalho é um processo gerador AR, enquanto a estimativa dos parâmetros do pseudo algoritmo de projeções afins (PAP) é realizada através de um analisador pelo método dos mínimos quadrados.

Diversos métodos de estimação dos coeficientes de um processo AR podem ser encontrados na literatura específica. Dentre eles, pode-se destacar o *método de Yule-Walker* e o *algoritmo de máxima entropia (método de Burg)*. O método *forward and backward linear prediction (FBLP)*, por sua vez, é baseado nos preditores lineares *forward linear prediction* e *backward linear prediction*, que permite a estimação de parâmetros através de um preditor linear pelo método dos mínimos quadrados (HAYKIN, 1996, p. 506). Esse último apresenta similaridade com o mecanismo intrínseco de decorrelação do pseudo algoritmo de projeções afins. No Apêndice A é apresentada uma derivação para a predição linear do PAP, baseado no problema de filtragem apresentado por Haykin (1996).

2.3 O Algoritmo de Projeções Afins e o Pseudo APA

Essa seção apresenta dois importantes integrantes da família de algoritmos de projeções afins: o algoritmo de projeções afins (APA) convencional e uma versão de baixo custo computacional denominada de pseudo algoritmo de projeções afins (PAP). A formulação do APA é derivada a partir da solução de um sistema subdeterminado de equações, conforme descrito em Ozeki e Umeda (1984, p. 22), enquanto o PAP é obtido com base na estimação dos parâmetros de um processo autoregressivo, conforme Ozeki e Umeda (1984, p. 21), (BOUTEILLE; SCALART e CORAZZA, 1999) e Rupp (1998, p. 772).

2.3.1 O APA como a Solução de um Sistema Linear Subdeterminado

Considerando-se o problema de filtragem adaptativa da Figura 6, e substituindo (2.18) em (2.19), tem-se a expressão para o erro instantâneo do APA:

$$e_{\text{APA}}(n) = d(n) - \mathbf{u}^T(n) \mathbf{w}_{\text{APA}}(n). \quad (2.32)$$

Dessa forma, deseja-se determinar o vetor de coeficientes $\mathbf{w}_{\text{APA}}(n+1)$ *a posteriori* que anule o erro $e_{\text{APA}}(n)$ causado pelo vetor de coeficientes *a priori*, ou seja, que resulta em

$$\mathbf{u}^T(n) \mathbf{w}_{\text{APA}}(n+1) = d(n). \quad (2.33)$$

Como descrito por Almeida:

Este é um sistema de uma única equação com N incógnitas $w_0(n+1), w_1(n+1), \dots, w_{N-1}(n+1)$. Tal sistema tem infinitas soluções, caracterizando um sistema subdeterminado. (ALMEIDA, 2004, p. 31).

Dentro do contexto de infinitas soluções, podem-se agregar restrições que imponham características desejáveis ao vetor de solução. Uma possibilidade é a escolha da solução com a menor norma euclidiana da distância entre vetores de coeficientes consecutivos, $\|\mathbf{w}_{\text{APA}}(n+1) - \mathbf{w}_{\text{APA}}(n)\|^2$, seguindo o princípio do mínimo distúrbio (ALMEIDA, 2004, p. 32). Outra possibilidade, assumindo-se a disponibilidade de $P+1$ vetores de entrada passados, é expandir a restrição do erro nulo gerado pelo vetor de coeficientes *a posteriori* para uma janela de tempo mais ampla, como por exemplo:

$$\begin{cases} \mathbf{u}^T(n) \mathbf{w}_{\text{APA}}(n+1) = d_0(n) \\ \mathbf{u}^T(n-1) \mathbf{w}_{\text{APA}}(n+1) = d_1(n) \\ \vdots \\ \mathbf{u}^T(n-P) \mathbf{w}_{\text{APA}}(n+1) = d_P(n). \end{cases} \quad (2.34)$$

Assumindo-se que ambas as características são desejáveis, pode-se expandir a equação (2.33) como

$$\mathbf{e}_p(n) = \mathbf{d}(n) - \mathbf{U}_u^T(n) \mathbf{w}_{\text{APA}}(n+1) \quad (2.35)$$

em que $\mathbf{U}_u(n) = [\mathbf{u}(n) \ \mathbf{u}(n-1) \ \dots \ \mathbf{u}(n-P)]$ é uma matriz formada por $P+1$ regressores, $\mathbf{d}(n) = [d_0(n) \ d_1(n) \ \dots \ d_P(n)]$ é um vetor de $P+1$ amostras do sinal desejado na iteração (n) . Dessa forma, tem-se que $\mathbf{e}_p(n) = \mathbf{0}_{P+1 \times 1}$, um vetor coluna de $P+1$ zeros. Definindo-se um vetor de erro $\mathbf{e}(n) = [e_{\text{APA}_0}(n) \ e_{\text{APA}_1}(n) \ \dots \ e_{\text{APA}_P}(n)]^T$, tem-se que o erro gerado pelo vetor de coeficientes *a priori* é dado por

$$\mathbf{e}(n) = \mathbf{d}(n) - \mathbf{U}_u^T(n) \mathbf{w}_{\text{APA}}(n). \quad (2.36)$$

Seguindo o procedimento adotado por Almeida (2004, p. 33), a utilização do princípio do mínimo distúrbio através do *Método dos Multiplicadores de Lagrange*, sob o conjunto de restrições apresentado, resulta em

$$\mathbf{w}_{\text{APA}}(n+1) = \mathbf{w}_{\text{APA}}(n) + \mathbf{U}_u(n) \left[\mathbf{U}_u^T(n) \mathbf{U}_u(n) \right]^{-1} \mathbf{e}(n). \quad (2.37)$$

De forma similar aos algoritmos LMS e NLMS, é desejável a inclusão de um passo de adaptação, μ , com a finalidade de controlar o processo de convergência. Assim, a equação de atualização do algoritmo de projeções afins convencional (APA) é definida como:

$$\mathbf{w}_{\text{APA}}(n+1) = \mathbf{w}_{\text{APA}}(n) + \mu \mathbf{U}_u(n) \left[\mathbf{U}_u^T(n) \mathbf{U}_u(n) \right]^{-1} \mathbf{e}(n). \quad (2.38)$$

Almeida (2004, p. 35) cita que “o algoritmo NLMS pode ser considerado um caso particular do algoritmo APA, quando é utilizado apenas um vetor de entrada ($P = 0$)”. Dessa forma, se $\mathbf{U}_u(n) = [\mathbf{u}(n)]$, tem-se a equação de atualização dos coeficientes do algoritmo NLMS,

$$\mathbf{w}_{\text{NLMS}}(n+1) = \mathbf{w}_{\text{NLMS}}(n) + \mu \frac{\mathbf{u}(n)}{\mathbf{u}^T(n)\mathbf{u}(n)} e_{\text{NLMS}}(n) \quad (2.39)$$

e, portanto, $\mathbf{w}_{\text{APA}}(n) = \mathbf{w}_{\text{NLMS}}(n)$ e $e_{\text{APA}}(n) = e_{\text{NLMS}}(n)$. Uma interpretação geométrica é apresentada no Anexo A, demonstrando o procedimento de ortogonalização do APA para $P = 1$ e a relação com o NLMS.

Devido à reutilização de regressores e do sinal desejado, os algoritmos da família de projeções afins são também denominados por *algoritmos de reuso de dados* (SAYED, 2003). Para interessados em uma interpretação alternativa do processo de ortogonalização do APA, a Subseção 2.2.1 de (ALMEIDA, 2004, p. 36) apresenta tal abordagem através do método de Gram-Schmidt. Finalmente, a equação de recursão dos coeficientes do algoritmo APA (2.37) pode ser derivada também através do gradiente estocástico utilizando o método de Newton de (2.15), conforme demonstração em Sayed (2003, p. 238).

2.3.2 O Pseudo Algoritmo de Projeções Afins - PAP

Em Ozeki e Umeda (1984, p. 21) é apresentado um novo algoritmo adaptativo que introduz uma inovação geométrica em relação ao algoritmo NLMS (*learning identification*).

Conforme descrito em Almeida (2004, p. 41), assumindo-se o algoritmo de projeções afins com passo unitário e $P = 1$, a direção de atualização do vetor de coeficientes pode ser obtida subtraindo-se, do vetor de entrada $\mathbf{u}(n)$, a sua projeção no vetor de entrada anterior $\mathbf{u}(n-1)$. Este processo resulta no vetor $\boldsymbol{\phi}(n)$, que é ortogonal a $\mathbf{u}(n-1)$, e pode ser definido por:

$$\boldsymbol{\phi}(n) = \mathbf{u}(n) - \mathbf{P}_u(n)\mathbf{u}(n) \quad (2.40)$$

em que $\mathbf{P}_u(n)$ é o operador de projeção no subespaço formado pelo vetor $\mathbf{u}(n-1)$, ou seja

$$\mathbf{P}_u(n) = \frac{\mathbf{u}(n-1)\mathbf{u}^T(n-1)}{\mathbf{u}^T(n-1)\mathbf{u}(n-1)}. \quad (2.41)$$

Assim, pode-se observar que o ajuste $\Delta\mathbf{w}(n) = \mathbf{w}(n+1) - \mathbf{w}(n)$ é realizado na direção do vetor $\boldsymbol{\phi}(n)$, e não mais na direção de $\mathbf{u}(n)$ Anexo A, de forma que

$$\begin{cases} \mathbf{u}^T(n)\mathbf{w}(n+1) = d(n) \\ \mathbf{w}(n+1) - \mathbf{w}(n) = \rho\boldsymbol{\phi}(n). \end{cases} \quad (2.42)$$

Multiplicando-se a segunda equação de (2.42) por $\boldsymbol{\phi}^T(n)$ e isolando ρ , obtém-se

$$\rho = \frac{\boldsymbol{\phi}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} [\mathbf{w}(n+1) - \mathbf{w}(n)]. \quad (2.43)$$

Pós-multiplicando a transposta de (2.40) pelos vetores *a posteriori* $\mathbf{w}(n+1)$ e *a priori* $\mathbf{w}(n)$, tem-se que

$$\begin{cases} \boldsymbol{\phi}^T(n)\mathbf{w}(n+1) = \mathbf{u}^T(n)\mathbf{w}(n+1) - \mathbf{u}^T(n)\mathbf{P}'_u(n)\mathbf{w}(n+1) \\ \boldsymbol{\phi}^T(n)\mathbf{w}(n) = \mathbf{u}^T(n)\mathbf{w}(n) - \mathbf{u}^T(n)\mathbf{P}'_u(n)\mathbf{w}(n) \end{cases} \quad (2.44)$$

em que utilizando (2.44) em (2.43) e assumindo-se que $\mathbf{P}'_u(n)[\mathbf{w}(n+1) - \mathbf{w}(n)] = \rho\mathbf{P}'_u(n)\boldsymbol{\phi}(n) = \mathbf{0}_{N \times 1}$, obtém-se

$$\rho = \frac{\mathbf{u}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} [\mathbf{w}(n+1) - \mathbf{w}(n)]. \quad (2.45)$$

Substituindo-se $\mathbf{u}^T(n)\mathbf{w}(n+1) = d(n)$ e $d(n) - \mathbf{u}^T(n)\mathbf{w}(n) = e(n)$ em (2.45), resulta em

$$\rho = \frac{e(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)}. \quad (2.46)$$

Finalmente, substituindo-se (2.46) na segunda equação de (2.42), obtém-se a equação de atualização do vetor dos coeficientes $\mathbf{w}(n+1)$ para o PAP com passo unitário e $P = 1$, expresso por

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\boldsymbol{\phi}(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} e(n). \quad (2.47)$$

As simulações realizadas por Ozeki e Umeda (1984) demonstraram que para um sinal de excitação AR de primeira ordem, o algoritmo proposto apresentou uma significativa melhora de desempenho quando comparado ao algoritmo NLMS. Posteriormente, Rupp (1998) expandiu esses resultados, concluindo que "se $\mathbf{u}(n)$ é um processo AR de ordem P , $\hat{\mathbf{a}}(n)$ é uma estimativa dos seus coeficientes AR, e $\boldsymbol{\phi}(n)$ é um vetor cujos elementos são estimativas de um processo

aleatório branco” (RUPP, 1998, tradução nossa). Isso se deve à estrutura de um analisador AR.

A equação (2.41) pode ser generalizada de tal forma que o vetor $\phi(n)$ seja ortogonal não apenas a um vetor passado, mas a P vetores do sinal de entrada Apêndice A de forma que

$$\mathbf{P}_U(n) = \mathbf{U}(n) \left[\mathbf{U}^T(n) \mathbf{U}(n) \right]^{-1} \mathbf{U}^T(n) = \mathbf{U}^+(n) \mathbf{U}^T(n) \quad (2.48)$$

em que $\mathbf{U}(n) = [\mathbf{u}(n-1) \ \mathbf{u}(n-2) \ \dots \ \mathbf{u}(n-P)]$ e

$$\mathbf{U}^+(n) = \mathbf{U}(n) \left[\mathbf{U}^T(n) \mathbf{U}(n) \right]^{-1} \quad (2.49)$$

é a pseudo inversa de Moore-Penrose de $\mathbf{U}(n)$.

A formulação para o vetor $\phi(n)$, como erro de estimativa de um preditor linear Apêndice A, é obtida substituindo-se (2.48) em (2.40), resultando em

$$\phi(n) = \mathbf{u}(n) - \mathbf{U}(n) \hat{\mathbf{a}}(n) \quad (2.50)$$

em que $\hat{\mathbf{a}}(n)$ é a estimativa dos mínimos quadrados do processo autoregressivo, conforme equação (A.15),

$$\hat{\mathbf{a}}(n) = \left[\mathbf{U}^T(n) \mathbf{U}(n) \right]^{-1} \mathbf{U}^T(n) \mathbf{u}(n) \quad (2.51)$$

sendo que $\mathbf{U}^T(n) \mathbf{U}(n)$ deve ter posto completo.

Substituindo-se um vetor coluna de N amostras do processo AR (2.31) em (2.51) obtém-se

$$\hat{\mathbf{a}}(n) = \mathbf{a} + \left[\mathbf{U}^T(n) \mathbf{U}(n) \right]^{-1} \mathbf{U}^T(n) \mathbf{z}(n) \quad (2.52)$$

em que \mathbf{a} são os coeficientes do filtro utilizado no processo AR e onde observa-se que $\mathbf{z}(n)$ possui dependência algébrica com os P vetores passados $\mathbf{u}(n-i)$, para $i = 1, \dots, P$, que compõem $\mathbf{U}(n)$. E, finalmente, substituindo (2.31) e (2.52) em (2.50), tem-se que

$$\phi(n) = \left[\mathbf{I} - \mathbf{P}_U(n) \right] \mathbf{z}(n) = \mathbf{P}_\perp(n) \mathbf{z}(n) \quad (2.53)$$

em que o operador de projeção $\mathbf{P}_U(n)$, definido em (2.48), atua no subespaço formado pelas colunas de $\mathbf{U}(n)$ e $\mathbf{P}_\perp(n)$ é a matriz de projeção no subespaço ortogonal complementar (APÊNDICE A). Então,

[...] o vetor $\mathbf{z}(n)$ pode ser descomposto como $\mathbf{z}_U(n) + \mathbf{z}_\perp(n)$, sendo $\mathbf{z}_U(n) = \mathbf{P}_U(n) \mathbf{z}(n)$ e $\mathbf{z}_\perp(n) = \mathbf{P}_\perp(n) \mathbf{z}(n)$. Dessa forma, apenas $\mathbf{z}_U(n)$ é algebricamente dependente de $\mathbf{U}(n)$. Além disso, desde que $\{\mathbf{z}(n)\}$ seja uma sequência de ruído branco, a potência de $\mathbf{z}(n)$ é igualmente distribuída

por suas N dimensões. Assim, apenas a potência da parcela $\mathbf{z}(n)$ que é projetada no subespaço P -dimensional definido por $\mathbf{U}(n)$ é responsável pela dependência entre $\mathbf{z}(n)$ e $\mathbf{U}(n)$. (ALMEIDA, 2004, p. 60)

Dessa forma, para sinais de entrada autoregressivos (AR) e passo unitário, o algoritmo PAP consiste na seguinte sequência de operações:

1. atualização da matriz $\mathbf{U}(n)$:

$$\mathbf{U}(n) = [\mathbf{u}(n-1) \ \mathbf{u}(n-2) \ \cdots \ \mathbf{u}(n-P)]$$

2. estimação do processo AR por mínimos quadrados:

$$\hat{\mathbf{a}}(n) = [\mathbf{U}^T(n)\mathbf{U}(n)]^{-1} \mathbf{U}^T(n)\mathbf{u}(n)$$

3. determinação da direção de atualização dos coeficientes:

$$\boldsymbol{\phi}(n) = \mathbf{u}(n) - \mathbf{U}(n)\hat{\mathbf{a}}(n)$$

4. avaliação do erro de estimação:

$$e(n) = d(n) - y(n)$$

5. atualização do vetor de coeficientes:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\boldsymbol{\phi}(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} e(n).$$

2.3.3 Propriedades do Algoritmo de Projeções Afins Convencional

De acordo com Rupp (1998) e Almeida (2004, p. 49), as principais propriedades do APA são:

1. descorrelaciona as amostras do sinal de excitação (propriedade de pré-branqueamento), através de um processo interno de branqueamento, reduzindo, desta forma, o tempo de convergência (RUPP, 1998);
2. o ruído de medida é modificado por uma filtragem linear, aumentando sua influência sobre o erro em excesso (RUPP, 1998);
3. a velocidade de convergência é independente da potência do sinal de excitação (ALMEIDA, 2004, p. 35) devido à normalização do vetor de entrada na equação de atualização dos coeficientes.

2.3.4 Custo computacional dos algoritmos APA e PAP

O custo computacional do algoritmo APA é apresentado na Tabela 1, sendo uma função da ordem do algoritmo (P ou $K = P+1$) e do número de coeficientes (N), e considerando-se sinais representados no domínio dos números reais. As informações contidas nessa tabela se baseiam na análise do custo computacional do algoritmo ε -APA, uma versão mais estável do APA, descrito em Sayed (SAYED, 2003).

Tabela 1- Custo computacional estimado por iteração do APA para sinais de número real, em termos do número de multiplicações e adições reais.

	Multiplicações	Somas
$\mathbf{U}_u^T(n)\mathbf{U}_u(n)$	K^2N	$K^2(N-1)$
$[\mathbf{U}_u^T(n)\mathbf{U}_u(n)]^{-1}$	K^3	K^3
$\mathbf{y}(n) = \mathbf{U}_u^T(n)\mathbf{w}(n)$	KN	$K(N-1)$
$\mathbf{e}(n) = \mathbf{d}(n) - \mathbf{y}(n)$	----	K
$[\mathbf{U}_u^T(n)\mathbf{U}_u(n)]^{-1}\mathbf{e}(n)$	K^2	$K(K-1)$
$\mathbf{U}_u(n)[\mathbf{U}_u^T(n)\mathbf{U}_u(n)]^{-1}\mathbf{e}(n)$	KN	$N(K-1)$
$\mathbf{w}(n+1)$	----	N
TOTAL	$(K^2+2K)N+K^3+K$	$(K^2+2K)N+K^3$

De forma análoga, o custo computacional do algoritmo PAP é apresentado na Tabela 2.

Tabela 2- Custo computacional estimado por iteração do PAP para sinais de número real, em termos do número de multiplicações e adições reais.

	Multiplicações	Somas
$\mathbf{U}^T(n)\mathbf{U}(n)$	P^2N	$P^2(N-1)$
$[\mathbf{U}^T(n)\mathbf{U}(n)]^{-1}$	P^3	P^3
$\mathbf{U}^T(n)\mathbf{u}(n)$	PN	$P(N-1)$
$\hat{\mathbf{a}}(n)$	P^2	$P(P-1)$
$\mathbf{U}(n)\hat{\mathbf{a}}(n)$	PN	$N(P-1)$
$\boldsymbol{\phi}(n) = \mathbf{u}(n) - \mathbf{U}(n)\hat{\mathbf{a}}(n)$	----	N
$\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)$	N	$N-1$
$\mathbf{y}(n) = \mathbf{u}^T(n)\mathbf{w}(n)$	N	$N-1$
$\mathbf{e}(n) = \mathbf{d}(n) - \mathbf{y}(n)$	----	1
$\boldsymbol{\phi}(n).\mathbf{e}(n) / [\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]$	N	----
$\mathbf{w}(n+1)$	----	N
TOTAL	$(P^2+2P+3)N+P^3+P^2$	$(P^2+2P+3)N+P^3-2P-1$

A análise realizada em ambas as tabelas considera que a inversão da matriz $K \times K$ e $P \times P$ necessita de $O(K^3)$ e $O(P^3)$ operações, respectivamente.

Como em casos de aplicações práticas em tempo real, $N \gg P$, e geralmente P possui um valor baixo, tem-se que para $P = \{1,2,3\}$, e $K = P+1$, o seguinte número de operações é necessário para cada algoritmo, a cada iteração.

Tabela 3- Número aproximado de operações do APA e PAP por iteração, para $P = \{1,2,3\}$ e considerando-se $N \gg P$.

P	APA - $O[(K^2+2K)N]$	PAP - $O[(P^2+2P+3)N]$
1	$8N$	$6N$
2	$15N$	$11N$
3	$24N$	$18N$

Dessa forma, a redução obtida no número de operações (multiplicações e somas) a cada iteração pelo PAP é de aproximadamente 25% , em relação ao APA. Salienta-se que conforme a ordem aumenta essa diferença entre os dois algoritmos é menos significativa. Dessa forma, no presente trabalho, o algoritmo PAP é descrito como uma solução de baixo custo computacional, conforme apresentado em Gonzalez et al. (2012).

3 DEFINIÇÃO DO PROBLEMA

Este capítulo estabelece inicialmente as hipóteses simplificadoras e estatísticas utilizadas nos modelos anteriores de Bershad; Linebarger e McLaughlin (2001) e Almeida (2004). Em seguida, é apresentada a descrição do problema e a formulação matemática proposta para sua solução, resultando em uma nova relação determinística mais acurada para o comportamento do pseudo algoritmo de projeções afins.

3.1 Introdução ao Problema

Em Bershad; Linebarger e McLaughlin (2001) e Almeida (2004) foram apresentados modelos analíticos para o comportamento médio dos coeficientes e erro quadrático médio do pseudo algoritmo de projeções afins. Em virtude da utilização de passo unitário, esses modelos foram assumidos válidos também para o algoritmo de projeções afins, e apresentaram excelente acurácia em relação à simulação Monte Carlo para todos os casos analisados. Entretanto, o uso extensivo desses modelos revelou imperfeições na previsão do comportamento do algoritmo para certas combinações de parâmetros (COSTA et al., 2012a). Nessas condições, verificou-se adicionalmente uma discrepância entre o comportamento do APA e do PAP, contrariando as afirmações de Bouteille; Scalart e Corazza (1999) sobre a equivalência de desempenho no caso de passo unitário. Essas observações levaram ao questionamento do mérito das considerações e aproximações utilizadas na obtenção dos modelos analíticos anteriores, apresentadas a seguir.

3.2 Considerações para a Análise do PAP

Devido à complexidade de análise do PAP, é necessário estabelecer simplificações, hipóteses e aproximações adequadas para que sejam derivados modelos teóricos que permitam o entendimento do comportamento do algoritmo nas diversas situações de aplicação. Esta seção apresenta as considerações utilizadas nos trabalhos anteriores e que colaboram nesse sentido.

A seguir são apresentados os requisitos fundamentais para que o novo modelo possa ser derivado:

- Passo de adaptação unitário, $\mu = 1$;
- O sinal de excitação $u(n)$ é assumido perfeitamente modelado por um processo autoregressivo (AR). Esse requisito é associado ao mecanismo de predição linear do algoritmo PAP derivado na Seção 2.3 e demonstrado no Apêndice A;
- O distúrbio da planta $r(n)$ é representado por um processo aleatório de média zero, independente e identicamente distribuído (i.i.d.);
- A ordem do PAP e do processo autoregressivo que modela o sinal de entrada são idênticas ($P = M$);
- A planta e o filtro adaptativo possuem o mesmo número N de coeficientes;
- A estrutura do filtro adaptativo é FIR transversal;
- Sinais possuem amplitude representada no domínio dos números reais.

3.2.1 Hipóteses Simplificadoras e Estatísticas

As hipóteses simplificadoras a seguir foram utilizadas no modelo derivado em Almeida (2004, p. 52):

1. O sinal de excitação do processo AR $z(n)$ é estacionário, branco, Gaussiano, com variância σ_z^2 e média zero;
2. O sinal autoregressivo $u(n)$ é estacionário, Gaussiano, com variância σ_u^2 e média zero;
3. Assume-se que os vetores $\mathbf{z}(n-k)$ e $\mathbf{w}(n-p)$ são estatisticamente independentes em quaisquer k e p ;
4. Assume-se que os vetores $\mathbf{z}(n-k)$ e $\mathbf{z}(n-p)$ são estatisticamente independentes para k diferente de p ;
5. Assume-se que os vetores de direção $\phi(n-k)$ e $\phi(n-p)$ são independentes para k diferente de p .

Com relação à hipótese número dois:

[...] é frequentemente utilizada na análise estatística de algoritmos adaptativos, na forma da teoria da independência (MAZO, 1979). Na teoria da independência considera-se que os vetores de entrada (regressores) são independentes, o que não é verdadeiro em aplicações que utilizam estruturas com linha de retardo. Nestes casos, o vetor de entrada é preenchido amostra a amostra, existindo apenas a diferença de um elemento entre dois vetores consecutivos. Entretanto, a teoria da independência leva a resultados satisfatórios em diversas situações de aplicação prática dos algoritmos estocásticos. (ALMEIDA, 2004, p. 52)

Adicionalmente, são assumidas as seguintes hipóteses e propriedades estatísticas definidas em Bershad; Linebarger e McLaughlin (2001) e Almeida (2004, p. 59):

- **H1:** *O número de coeficientes do filtro adaptativo é grande o suficiente para considerar que $N \gg P$, ou seja, o número de elementos do vetor é muito maior que o número de vetores passados que formam a matriz $\mathbf{U}(n)$. Usualmente é o caso para sistemas com resposta ao impulso longa, uma vez que P tende a ser limitado pela complexidade do algoritmo (ALMEIDA, 2004, p. 60);*
- **H2:** *Assume-se que a dependência estatística entre o vetor $\mathbf{z}(n)$ e a matriz $\mathbf{U}(n)$ pode ser desprezada, com base na hipótese **H1** e nas conclusões apresentadas na citação após a equação (2.53). Esta hipótese é suportada em Bershad; Linebarger e McLaughlin (2001) e Almeida (2004, p. 60);*
- **H3:** *Os vetores $\phi(n)$ e $\mathbf{w}(n)$ são considerados estatisticamente independentes. Esta hipótese é similar à hipótese da independência (MAZO, 1979), quando aplicada a um sinal branco utilizando um filtro adaptativo em linha de retardo (ALMEIDA, 2004, p. 60);*

- **H4:** O vetor $\phi(n)$ é aleatório, Gaussiano e de média zero. A equação (2.53) mostra que cada componente $\phi(n-i)$ do vetor $\phi(n)$ é determinada pela expressão $\sum_{j=1}^N P_{\perp ij} z(n-j+1)$. Assumindo-se a hipótese **H1** e tendo-se que o vetor de inovações $\mathbf{z}(n)$ é um processo aleatório do tipo ruído branco (um analisador AR), as variáveis aleatórias desse somatório são independentes. Dessa forma, pelo teorema do limite central (MAZO, 1979), a distribuição de $\phi(n)$ tende para uma Gaussiana para valores elevados de N . Como $\mathbf{z}(n)$ é de média zero, $\phi(n)$ também será média zero (ALMEIDA, 2004, p. 61).

Da hipótese **H3**, tem-se ainda que os elementos do vetor $\phi(n) = [\mathbf{I}_N - \mathbf{P}_U(n)]\mathbf{z}(n) = \mathbf{P}_\perp(n)\mathbf{z}(n)$ em (2.53) são estimativas de uma sequência de ruído branco $\mathbf{z}(n) = \mathbf{z}_\perp(n) + \mathbf{z}_u(n)$ em que $\mathbf{z}_u(n) = \mathbf{P}_U(n)\mathbf{z}(n)$ e $\mathbf{z}_\perp(n) = \mathbf{P}_\perp(n)\mathbf{z}(n)$. Sendo $\mathbf{z}(n)$ um processo aleatório branco, a potência de $\mathbf{z}(n)$ é igualmente distribuída por suas N dimensões e $\mathbf{z}_\perp(n)$ é um vetor com energia apenas em $N-P$ dimensões do espaço N -dimensional, enquanto $\mathbf{z}_u(n)$ contribui com energia nas P dimensões restantes. Assim, a matriz de correlação do vetor $\phi(n)$ é expressa pela igualdade $E\{\phi(n)\phi^T(n)\} = E\{\mathbf{z}_\perp(n)\mathbf{z}_\perp^T(n)\}$, resultando em (ALMEIDA, 2004, p. 61):

$$\mathbf{R}_{\phi\phi}(n) \triangleq E\{\phi(n)\phi^T(n)\} \approx \frac{N-P}{N} \sigma_z^2 \mathbf{I}_N. \quad (3.1)$$

3.3 O Efeito de Inicialização do PAP

Conforme descrito anteriormente, os modelos teóricos de Bershad; Linebarger e Mclaughlin (2001) e Almeida (2004) foram fundamentados - adicionalmente às hipóteses da Seção 3.2 - pelas relações determinísticas propostas em Rupp (1998). Essa seção apresenta uma nova derivação para essa relação, considerando todos os instantes do comportamento do pseudo algoritmo de projeções afins.

Utilizando-se as equações (2.17), (2.18) e (2.19) do filtro adaptativo, e (2.47) e (2.50) utilizadas na atualização dos coeficientes do algoritmo, respectivamente,

$$\begin{cases} d(n) = \mathbf{u}^T(n) \mathbf{w}^o + r(n) \\ y(n) = \mathbf{u}^T(n) \mathbf{w}(n) \\ e(n) = d(n) - y(n) \\ \mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\boldsymbol{\phi}(n)}{\boldsymbol{\phi}^T(n) \boldsymbol{\phi}(n)} e(n) \\ \boldsymbol{\phi}(n) = \mathbf{u}(n) - \mathbf{U}(n) \hat{\mathbf{a}}(n) \end{cases}$$

e definindo-se o vetor de erro dos coeficientes como

$$\mathbf{v}(n) = \mathbf{w}(n) - \mathbf{w}^o \quad (3.2)$$

sendo que \mathbf{w}^o é o vetor de coeficientes ótimos da planta no sentido de Wiener, chega-se à seguinte expressão determinística para a recursão do vetor de erro dos coeficientes (ALMEIDA, 2004, p. 56):

$$\mathbf{v}(n+1) = \mathbf{v}(n) - \frac{\boldsymbol{\phi}(n) \mathbf{u}^T(n)}{\boldsymbol{\phi}^T(n) \boldsymbol{\phi}(n)} \mathbf{v}(n) + \frac{\boldsymbol{\phi}(n) r(n)}{\boldsymbol{\phi}^T(n) \boldsymbol{\phi}(n)}. \quad (3.3)$$

Isolando-se $\mathbf{u}(n)$ em (2.47) e substituindo a sua transposta em (3.3), obtém-se que

$$\begin{aligned} \mathbf{v}(n+1) &= \mathbf{v}(n) - \frac{\boldsymbol{\phi}(n) \boldsymbol{\phi}^T(n)}{\boldsymbol{\phi}^T(n) \boldsymbol{\phi}(n)} \mathbf{v}(n) \\ &\quad - \frac{\boldsymbol{\phi}(n) \hat{\mathbf{a}}^T(n)}{\boldsymbol{\phi}^T(n) \boldsymbol{\phi}(n)} \mathbf{U}^T(n) \mathbf{v}(n) + \frac{r(n) \boldsymbol{\phi}(n)}{\boldsymbol{\phi}^T(n) \boldsymbol{\phi}(n)}. \end{aligned} \quad (3.4)$$

Em Rupp (1998, p. 772) é apresentado o seguinte conjunto de relações determinísticas, obtido através da pré-multiplicação de $\boldsymbol{\phi}^T(n)$, $\mathbf{u}^T(n)$ e $\mathbf{U}^T(n)$ em (3.3), resultando em

$$\boldsymbol{\phi}^T(n) \mathbf{v}(n+1) = \boldsymbol{\phi}^T(n) \mathbf{v}(n) - \mathbf{u}(n) \mathbf{v}(n) + r(n) \quad (3.5)$$

$$\mathbf{u}^T(n) \mathbf{v}(n) = r(n) \quad (3.6)$$

$$\mathbf{U}^T(n) \mathbf{v}(n+1) = \mathbf{U}^T(n) \mathbf{v}(n) \quad (3.7)$$

sendo que a derivação é apresentada na Subseção 3.4 de Almeida (2004). Substituindo (3.6) recursivamente em (3.7), chega-se a uma relação determinística dada por

$$\mathbf{U}^T(n) \mathbf{v}(n) = \mathbf{r}(n-1) \quad (3.8)$$

em que $\mathbf{r}(n-1) = [r(n-1) \ r(n-2) \ \dots \ r(n-P)]^T$ é o vetor de amostras passadas de ruído.

Assim, aplicando (3.8) em (3.4) tem-se que

$$\mathbf{v}(n+1) = \mathbf{v}(n) - \frac{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \mathbf{v}(n) + \frac{\boldsymbol{\phi}(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} r_a(n) \quad (3.9)$$

sendo definida uma sequência de ruído filtrado como (RUPP, 1998)

$$r_a(n) = r(n) - \hat{\mathbf{a}}^T(n)\mathbf{r}(n-1). \quad (3.10)$$

Dessa forma, Rupp apresenta as propriedades básicas de pré-branqueamento, filtragem do ruído e controle de convergência (versão normalizada do vetor de entrada) do algoritmo desenvolvido por Ozeki e Umeda (1984).

Com base neste trabalho, os modelos desenvolvidos por Bershad; Linebarger e McLaughlin (2001) e Almeida (2004) foram desenvolvidos a partir da equação (3.9). Entretanto, indícios de imprecisões foram detectados no transiente inicial do modelo obtido. Dessa forma, verificou-se que para $\mathbf{U}^T(n)\mathbf{v}(n+1)|_{n=-1} = \mathbf{U}^T(-1)[\mathbf{w}(0) - \mathbf{w}^o]$ (antes da primeira atualização dos coeficientes), a associação entre as equações (3.7) e (3.8), que resultaria em $\mathbf{r}(-2)$, ocasiona uma desigualdade. A partir dessa constatação, e considerando a recursividade na adaptação dos coeficientes, fez-se necessário a expansão de (3.8) nos primeiros instantes do processo de adaptação, conforme apresentado a seguir.

Diferentemente do procedimento realizado para obter (3.4), o qual não leva em consideração o instante anterior à primeira atualização dos coeficientes, são utilizadas condições para todo n , isto é,

$$\mathbf{v}(n+1) = \begin{cases} = \mathbf{w}(0) - \mathbf{w}^o & \text{para } n < 0 \\ = \mathbf{v}(n) - \frac{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \mathbf{v}(n) & \\ - \frac{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \mathbf{U}^T(n)\mathbf{v}(n) + \frac{r(n)\boldsymbol{\phi}(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} & \text{para } n \geq 0 \end{cases} \quad (3.11)$$

sendo que $\mathbf{w}(0)$ são as condições iniciais do vetor de coeficientes.

Inicialmente, nesta seção, o índice i será empregado para indicar os vetores coluna da matriz $\mathbf{U}(n)$, sendo $1 \leq i \leq P$. Para $n \leq 0$, antes que o algoritmo realize a primeira adaptação, tem-se que as linhas de $\mathbf{U}^T(n)\mathbf{v}(n)$ são expressas por

$$\mathbf{u}^T(n-i)\mathbf{v}(n) = \mathbf{u}^T(n-i)[\mathbf{w}(0) - \mathbf{w}^o] \quad \text{para } n \leq 0 \quad (3.12)$$

Para $n > 0$, a avaliação de $\mathbf{U}^T(n)\mathbf{v}(n)$ é obtida pré-multiplicando (3.11) por $\mathbf{u}^T(n)$, o que resulta em

$$\mathbf{u}^T(n)\mathbf{v}(n+1) = \begin{cases} = \mathbf{u}^T(n) [\mathbf{w}(0) - \mathbf{w}^0] & \text{para } n < 0 \\ = \mathbf{u}^T(n)\mathbf{v}(n) - \frac{\mathbf{u}^T(n)\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)}\mathbf{v}(n) \\ - \frac{\mathbf{u}^T(n)\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)}\mathbf{U}^T(n)\mathbf{v}(n) \\ + \frac{r(n)\mathbf{u}^T(n)\boldsymbol{\phi}(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} & \text{para } n \geq 0 \end{cases} \quad (3.13)$$

sendo que isolando $\mathbf{u}(n)$ em (2.50) e substituindo sua transposta em (3.13), leva à

$$\mathbf{u}^T(n)\mathbf{v}(n+1) = \begin{cases} = \mathbf{u}^T(n) [\mathbf{w}(0) - \mathbf{w}^0] & \text{para } n < 0 \\ = \boldsymbol{\phi}^T(n)\mathbf{v}(n) + \hat{\mathbf{a}}^T(n)\mathbf{U}^T(n)\mathbf{v}(n) \\ - \boldsymbol{\phi}^T(n)\mathbf{v}(n) \\ - \frac{\hat{\mathbf{a}}^T(n)\mathbf{U}^T(n)\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)}\mathbf{v}(n) \\ - \hat{\mathbf{a}}^T(n)\mathbf{U}^T(n)\mathbf{v}(n) \\ - \frac{\hat{\mathbf{a}}^T(n)\mathbf{U}^T(n)\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)}\mathbf{U}^T(n)\mathbf{v}(n) \\ + r(n) + \frac{r(n)\hat{\mathbf{a}}^T(n)\mathbf{U}^T(n)\boldsymbol{\phi}(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} & \text{para } n \geq 0. \end{cases} \quad (3.14)$$

Aplicando a ortogonalidade vetorial, $\mathbf{U}^T(n)\boldsymbol{\phi}(n) = \mathbf{0}_{P \times 1}$ (APÊNDICE A), em (3.14), resulta em

$$\mathbf{u}^T(n)\mathbf{v}(n+1) = \begin{cases} \mathbf{u}^T(n) [\mathbf{w}(0) - \mathbf{w}^0] & \text{para } n < 0 \\ r(n) & \text{para } n \geq 0. \end{cases} \quad (3.15)$$

A seguir, a equação (3.11) é pré-multiplicada por $\mathbf{U}^T(n)$, o que resulta em

$$\mathbf{U}^T(n)\mathbf{v}(n+1) = \begin{cases} = \mathbf{U}^T(n) [\mathbf{w}(0) - \mathbf{w}^o] & \text{para } n < 0 \\ = \mathbf{U}^T(n)\mathbf{v}(n) - \frac{\mathbf{U}^T(n)\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \mathbf{v}(n) \\ - \frac{\mathbf{U}^T(n)\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \mathbf{U}^T(n)\mathbf{v}(n) & \text{para } n \geq 0. \\ + \frac{r(n)\mathbf{U}^T(n)\boldsymbol{\phi}(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \end{cases} \quad (3.16)$$

Aplicando-se novamente a condição de ortogonalidade entre $\boldsymbol{\phi}(n)$ e o subespaço formado pelas colunas de $\mathbf{U}(n)$, tem-se que

$$\mathbf{U}^T(n)\mathbf{v}(n+1) = \begin{cases} = \mathbf{U}^T(n) [\mathbf{w}(0) - \mathbf{w}^o] & \text{para } n < 0 \\ = \mathbf{U}^T(n)\mathbf{v}(n) & \text{para } n \geq 0. \end{cases} \quad (3.17)$$

Os elementos de (3.17) para $n \geq 0$ e $0 < i \leq P$, podem ser expressos como

$$\mathbf{u}^T(n-i)\mathbf{v}(n+1) = \mathbf{u}^T(n-i)\mathbf{v}(n) \quad \text{para } n \geq 0, 1 \leq i \leq P. \quad (3.18)$$

Exemplos de (3.18) para os instantes $0 \leq n \leq 3$ e $0 < i \leq P$ são apresentados na Tabela 4.

Tabela 4- Exemplos de (3.18) para $0 \leq n \leq 3$ e $0 < i \leq P$.

		$n = 0$	$n = 1$	$n = 2$	$n = 3$
P	$i = 1$	$\mathbf{u}^T(-1)\mathbf{v}(1)$ $= \mathbf{u}^T(-1)\mathbf{v}(0)$	$\mathbf{u}^T(0)\mathbf{v}(2)$ $= \mathbf{u}^T(0)\mathbf{v}(1)$	$\mathbf{u}^T(1)\mathbf{v}(3)$ $= \mathbf{u}^T(1)\mathbf{v}(2)$	$\mathbf{u}^T(2)\mathbf{v}(4)$ $= \mathbf{u}^T(2)\mathbf{v}(3)$
	$i = 2$	$\mathbf{u}^T(-2)\mathbf{v}(1)$ $= \mathbf{u}^T(-2)\mathbf{v}(0)$	$\mathbf{u}^T(-1)\mathbf{v}(2)$ $= \mathbf{u}^T(-1)\mathbf{v}(1)$	$\mathbf{u}^T(0)\mathbf{v}(3)$ $= \mathbf{u}^T(0)\mathbf{v}(2)$	$\mathbf{u}^T(1)\mathbf{v}(4)$ $= \mathbf{u}^T(1)\mathbf{v}(3)$
	$i = 3$	$\mathbf{u}^T(-3)\mathbf{v}(1)$ $= \mathbf{u}^T(-3)\mathbf{v}(0)$	$\mathbf{u}^T(-2)\mathbf{v}(2)$ $= \mathbf{u}^T(-2)\mathbf{v}(1)$	$\mathbf{u}^T(-1)\mathbf{v}(3)$ $= \mathbf{u}^T(-1)\mathbf{v}(2)$	$\mathbf{u}^T(0)\mathbf{v}(4)$ $= \mathbf{u}^T(0)\mathbf{v}(3)$

	$i = P$	$\mathbf{u}^T(-P)\mathbf{v}(1)$ $= \mathbf{u}^T(-P)\mathbf{v}(0)$	$\mathbf{u}^T(1-P)\mathbf{v}(2)$ $= \mathbf{u}^T(1-P)\mathbf{v}(1)$	$\mathbf{u}^T(2-P)\mathbf{v}(3)$ $= \mathbf{u}^T(2-P)\mathbf{v}(2)$	$\mathbf{u}^T(3-P)\mathbf{v}(4)$ $= \mathbf{u}^T(3-P)\mathbf{v}(3)$

Na Tabela 4, verifica-se que os elementos das diagonais (apresentados na mesma cor) constituem uma série de igualdades, de forma que

$$\left\{ \begin{array}{l} \dots \\ \mathbf{u}^T(-1)\mathbf{v}(0) = \mathbf{u}^T(-1)\mathbf{v}(1) = \mathbf{u}^T(-1)\mathbf{v}(2) = \dots \\ \mathbf{u}^T(0)\mathbf{v}(1) = \mathbf{u}^T(0)\mathbf{v}(2) = \mathbf{u}^T(0)\mathbf{v}(3) = \dots \\ \mathbf{u}^T(1)\mathbf{v}(2) = \mathbf{u}^T(1)\mathbf{v}(3) = \mathbf{u}^T(1)\mathbf{v}(4) = \dots \\ \mathbf{u}^T(2)\mathbf{v}(3) = \mathbf{u}^T(2)\mathbf{v}(4) = \mathbf{u}^T(2)\mathbf{v}(5) = \dots \\ \dots \end{array} \right. \text{ para } n > 0 \quad (3.19)$$

e portanto, tem-se que a seguinte igualdade é válida:

$$\mathbf{u}^T(n-i)\mathbf{v}(n) = \begin{cases} \mathbf{u}^T(n-i)[\mathbf{w}(0) - \mathbf{w}^0] & \text{para } n \geq 0 \text{ e } n < i-1 \\ \mathbf{u}^T(n-i)\mathbf{v}(n-i+1) & \text{para } n \geq 0 \text{ e } n \geq i-1. \end{cases} \quad (3.20)$$

Adicionalmente, as linhas de $\mathbf{U}^T(n)\mathbf{v}(n)$ em (3.12) podem ser expressas como

$$\mathbf{u}^T(n-i)\mathbf{v}(n) = \begin{cases} = \mathbf{u}^T(n-i)[\mathbf{w}(0) - \mathbf{w}^0] & \text{para } n \leq 0 \\ = \mathbf{u}^T(n-i)\mathbf{v}(n) & \text{para } n > 0 \end{cases} \quad (3.21)$$

sendo que utilizando (3.20) em (3.21), obtém-se

$$\mathbf{u}^T(n-i)\mathbf{v}(n) = \begin{cases} = \mathbf{u}^T(n-i)[\mathbf{w}(0) - \mathbf{w}^0] & \text{para } n < 0 \\ = \mathbf{u}^T(n-i)[\mathbf{w}(0) - \mathbf{w}^0] & \text{para } n \geq 0 \text{ e } n < i-1 \\ = \mathbf{u}^T(n-i)\mathbf{v}(n-i+1) & \text{para } n \geq 0 \text{ e } n \geq i-1. \end{cases} \quad (3.22)$$

Por outro lado, devido à característica escalar de (3.15), tem-se que substituindo n por $n-i$, resulta em

$$\mathbf{u}^T(n-i)\mathbf{v}(n-i+1) = \begin{cases} \mathbf{u}^T(n-i)[\mathbf{w}(0) - \mathbf{w}^0] & \text{para } n < i \\ r(n-i) & \text{para } n \geq i. \end{cases} \quad (3.23)$$

Finalmente, substituindo (3.23) em (3.22), chega-se à

$$\mathbf{u}^T(n-i)\mathbf{v}(n) = \begin{cases} = \mathbf{u}^T(n-i)[\mathbf{w}(0) - \mathbf{w}^0] & \text{para } n \leq 0 \\ = \mathbf{u}^T(n-i)[\mathbf{w}(0) - \mathbf{w}^0] & \text{para } n > 0 \text{ e } n < i \\ = r(n-i) & \text{para } n > 0 \text{ e } n \geq i \end{cases} \quad (3.24)$$

sendo que essa igualdade é demonstrada graficamente através da Tabela 5 para $P = 4$ e $0 \leq n \leq 4$.

Tabela 5- Relação iterativa de igualdade de $\mathbf{u}^T(n-i)\mathbf{v}(n)$ em (3.24) para o PAP.

		$n = 0$	$n = 1$	$n = 2$	$n = 3$	$n = 4$
P	$i=1$	$\mathbf{u}^T(-1)\mathbf{v}(0)$	$r(0)$	$r(1)$	$r(2)$	$r(3)$
	$i=2$	$\mathbf{u}^T(-2)\mathbf{v}(0)$	$\mathbf{u}^T(-1)\mathbf{v}(0)$	$r(0)$	$r(1)$	$r(2)$
	$i=3$	$\mathbf{u}^T(-3)\mathbf{v}(0)$	$\mathbf{u}^T(-2)\mathbf{v}(0)$	$\mathbf{u}^T(-1)\mathbf{v}(0)$	$r(0)$	$r(1)$
	$i=4$	$\mathbf{u}^T(-4)\mathbf{v}(0)$	$\mathbf{u}^T(-3)\mathbf{v}(0)$	$\mathbf{u}^T(-2)\mathbf{v}(0)$	$\mathbf{u}^T(-1)\mathbf{v}(0)$	$r(0)$

Na Tabela 5, em vermelho, estão marcados os elementos correspondentes à inicialização do algoritmo e em verde os referentes às amostras do ruído de medida. Cada coluna da Tabela 5 corresponde aos elementos do vetor $\mathbf{U}^T(n)\mathbf{v}(n)$ a cada iteração. Assim, a equação (3.24) pode ser expressa por

$$\mathbf{u}^T(n-i)\mathbf{v}(n) = r(n-i) + q_i(n) \quad (3.25)$$

sendo que o elemento $q_i(n)$ visa produzir os elementos em vermelho na Tabela 5 para $n > 0$, e portanto,

$$q_i(n) = \left[\mathbf{u}^T(n-i)\mathbf{v}(0) - r(n-i) \right] u_{-1}(i-n) \quad (3.26)$$

em que $u_{-1}(k)$ é uma função degrau unitário (1 se $k \geq 0$ e 0 para outros valores de k) e $1 \leq i \leq P$. Os elementos $q_i(n)$ podem ser expressos na forma vetorial como

$$\mathbf{q}(n) = \mathbf{Q}(-n+1) \left[\mathbf{U}^T(n)\mathbf{v}(0) - \mathbf{r}(n-1) \right] \quad (3.27)$$

tendo introduzido uma matriz determinística variante no tempo $\mathbf{Q}(-n+1)$, de dimensões $P \times P$, definida por

$$\mathbf{Q}(-n+1) \triangleq \begin{bmatrix} u_{-1}(-n) & 0 & \dots & 0 \\ 0 & u_{-1}(-n+1) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & u_{-1}(-n+P-1) \end{bmatrix}. \quad (3.28)$$

A matriz $\mathbf{Q}(-n+1)$ é diagonal e à medida que o tempo (n) evolui, os elementos de sua diagonal principal tornam-se sucessivamente nulos.

Expressando (3.25) na forma vetorial, resulta em

$$\mathbf{U}^T(n)\mathbf{v}(n) = \mathbf{r}(n-1) + \mathbf{q}(n) \quad (3.29)$$

e substituindo (3.27) em (3.29), tem-se que

$$\begin{aligned} \mathbf{U}^T(n)\mathbf{v}(n) &= \mathbf{r}(n-1) + \mathbf{Q}(-n+1)\mathbf{U}^T(n)\mathbf{v}(0) \\ &\quad - \mathbf{Q}(-n+1)\mathbf{r}(n-1). \end{aligned} \quad (3.30)$$

A equação (3.30) estende o resultado de Rupp (1998) apresentado em (3.8), de forma a considerar todas as etapas do processo de adaptação (transitório e regime permanente) para o pseudo algoritmo de projeções afins. Como pode ser visto através de (3.28), esse efeito ocorre até a iteração $n = P-1$ e conseqüentemente a inicialização do algoritmo influencia o comportamento do algoritmo. O termo $\mathbf{q}(n)$ é uma conseqüência do Efeito de Inicialização, cujo resultado pode ser visualizado através dos itens em vermelho na Tabela 5.

3.4 A Relação entre APA e PAP

Como descrito na introdução deste capítulo, inicialmente uma análise numérica extensiva indicou a existência de uma disparidade na convergência dos algoritmos APA e PAP, contrariando as afirmações de Bouteille; Scalart e Corazza (1999) sobre a equivalência no desempenho para o caso de passo unitário. Esta seção demonstra que a condição de passo unitário é necessária, porém não é suficiente para que ambos os algoritmos apresentem o mesmo comportamento.

Preliminarmente, a relação entre o erro obtido com a atualização *a posteriori* $\mathbf{e}_p(n)$ e *a priori* $\mathbf{e}(n)$ dos coeficientes do algoritmo APA é obtida substituindo (2.38) em (2.35),

$$\begin{aligned} \mathbf{e}_p(n) &= \underbrace{\mathbf{d}(n) - \mathbf{U}_u^T(n)\mathbf{w}_{\text{APA}}(n)}_{\mathbf{e}(n)} \\ &\quad - \mu \mathbf{U}_u^T(n)\mathbf{U}_u(n) \left[\mathbf{U}_u^T(n)\mathbf{U}_u(n) \right]^{-1} \mathbf{e}(n) \\ &= \mathbf{e}(n) - \mu \mathbf{I}_{P+1} \mathbf{e}(n) \end{aligned} \quad (3.31)$$

sendo que $\mathbf{U}_u^T(n)\mathbf{U}_u(n) \left[\mathbf{U}_u^T(n)\mathbf{U}_u(n) \right]^{-1} = \mathbf{I}_{P+1}$ e os dois primeiros termos à direita da primeira igualdade de (3.31) podem ser expressos por (2.36). Dessa forma, (3.31) resulta em

$$\mathbf{e}_p(n) = (1 - \mu) \mathbf{e}(n). \quad (3.32)$$

Considerando-se o uso de passo unitário em (3.32), e definindo-se que $\mathbf{U}_u(n) = [\mathbf{u}(n) \mathbf{U}(n)]$, tem-se que o erro ocorrido em $\mathbf{w}(n+1)$, utilizado nas restrições da minimização dos mínimos quadrados pelo método de Lagrange (Almeida, 2004) na equação (3.38), é expresso por

$$\mathbf{e}_p(n) \Big|_{n \geq 0} = \begin{bmatrix} r(n) - \mathbf{u}^T(n) \mathbf{v}_{\text{APA}}(n+1) \\ \mathbf{r}(n-1) - \mathbf{U}^T(n) \mathbf{v}_{\text{APA}}(n+1) \end{bmatrix} = \mathbf{0}_{(P+1) \times 1} \quad (3.33)$$

em que $\mathbf{0}_{P+1 \times 1}$ é um vetor coluna composto de zeros. Na iteração anterior, tem-se que o erro *a posteriori* em função de $\mathbf{v}_{\text{APA}}(n)$ é obtido por

$$\mathbf{e}_p(n-1) \Big|_{n > 0} = \begin{bmatrix} \mathbf{r}(n-1) - \mathbf{U}^T(n) \mathbf{v}_{\text{APA}}(n) \\ r(n-P-1) - \mathbf{u}^T(n-P-1) \mathbf{v}_{\text{APA}}(n) \end{bmatrix} = \mathbf{0}_{(P+1) \times 1} \quad (3.34)$$

Porém, uma inicialização arbitrária de $\mathbf{w}(0)$, antes da primeira iteração de atualização do algoritmo, provavelmente não irá satisfazer a relação entre o vetor de amostras passadas de ruído e $\mathbf{U}(n)\mathbf{v}(n)$, de forma que

$$\mathbf{e}_p(n-1) \Big|_{n \leq 0} = \begin{bmatrix} \mathbf{r}(n-1) - \mathbf{U}^T(n) [\mathbf{w}_{\text{APA}}(0) - \mathbf{w}^o] \\ r(n-P-1) - \mathbf{u}^T(n-P-1) [\mathbf{w}_{\text{APA}}(0) - \mathbf{w}^o] \end{bmatrix} \quad (3.35)$$

em que $\mathbf{w}_{\text{APA}}(0)$ são as condições iniciais do filtro adaptativo.

De forma similar, expandindo-se (2.36), tem-se que

$$\mathbf{e}(n) = \begin{bmatrix} r(n) - \mathbf{u}^T(n) \mathbf{v}_{\text{APA}}(n) \\ \mathbf{r}(n-1) - \mathbf{U}^T(n) \mathbf{v}_{\text{APA}}(n) \end{bmatrix} = \begin{bmatrix} e_{\text{APA}}(n) \\ \mathbf{r}(n-1) - \mathbf{U}^T(n) \mathbf{v}_{\text{APA}}(n) \end{bmatrix} \quad (3.36)$$

em que $e_{\text{APA}}(n)$ é o primeiro elemento escalar do vetor de erro $\mathbf{e}(n)$ do APA, obtida em (2.32). Desse modo, considerando a relação entre o erro gerado pela atualização dos coeficientes *a priori* e *a posteriori*, e substituindo as linhas 1 a P de (3.34) e (3.35) nas linhas 2 a $P+1$ de (3.36), leva ao seguinte resultado:

$$\mathbf{e}(n) = \begin{cases} = \begin{bmatrix} e_{\text{APA}}(n) \\ \mathbf{r}(n-1) - \mathbf{U}^T(n) [\mathbf{w}_{\text{APA}}(0) - \mathbf{w}^o] \end{bmatrix} & p / n \leq 0 \\ = \begin{bmatrix} e_{\text{APA}}(n) \\ \mathbf{0}_{P \times 1} \end{bmatrix} & p / n > 0 \end{cases} \quad (3.37)$$

em que $\mathbf{0}_{P \times 1}$ é um vetor coluna composto de P zeros. Dessa forma, tem-se que o erro na atualização *a priori* dos coeficientes do APA, a partir de sua primeira iteração, será composto de um único elemento de erro escalar, representando o erro ocasionado a cada nova iteração.

À luz dessa informação, a equação (2.38) pode ser reescrita no instante $n = 0$, utilizando (3.37), resultando em

$$\mathbf{w}_{\text{APA}}(n+1)|_{n=0} = \mathbf{w}_{\text{APA}}(n) + \begin{bmatrix} \mathbf{U}_u^+(n) \begin{bmatrix} 1 \\ \mathbf{0}_{P \times 1} \end{bmatrix} \\ \mathbf{U}_u^+(n) \begin{bmatrix} \mathbf{0}_{1 \times P} \\ \mathbf{1}_P \end{bmatrix} \end{bmatrix} \times \begin{bmatrix} e_{\text{APA}}(n) \\ \mathbf{r}(n-1) - \mathbf{U}(n) [\mathbf{w}_{\text{APA}}(0) - \mathbf{w}^o] \end{bmatrix} \quad (3.38)$$

e após a primeira atualização dos coeficientes por

$$\mathbf{w}_{\text{APA}}(n+1)|_{n>0} = \mathbf{w}_{\text{APA}}(n) + \begin{bmatrix} \mathbf{U}_u^+(n) \begin{bmatrix} 1 \\ \mathbf{0}_{P \times 1} \end{bmatrix} \\ \mathbf{U}_u^+(n) \begin{bmatrix} \mathbf{0}_{1 \times P} \\ \mathbf{1}_P \end{bmatrix} \end{bmatrix} \begin{bmatrix} e_{\text{APA}}(n) \\ \mathbf{0}_{P \times 1} \end{bmatrix} \quad (3.39)$$

em que $\mathbf{0}_{1 \times P}$ é um vetor linha de P zeros, $\mathbf{1}_P$ é uma matriz identidade de dimensões $P \times P$, e $\mathbf{U}_u^+(n)$ é a pseudo-inversa de Moore-Penrose de $\mathbf{U}_u(n)$,

$$\mathbf{U}_u^+(n) = \mathbf{U}_u(n) [\mathbf{U}_u^T(n) \mathbf{U}_u(n)]^{-1}. \quad (3.40)$$

O primeiro elemento no segundo termo à direita de (3.38) e (3.39) $\mathbf{U}_u^+(n) [\mathbf{1} \ \mathbf{0}_{P \times 1}]^T$, é derivado em Almeida (2004), sendo demonstrado à seguir. Pré-multiplicando-se $\boldsymbol{\phi}(n)$ por $\mathbf{U}_u^T(n)$, resulta em

$$\mathbf{U}_u^T(n) \boldsymbol{\phi}(n) = \begin{bmatrix} \mathbf{u}^T(n) \\ \mathbf{U}^T(n) \end{bmatrix} \boldsymbol{\phi}(n) = \begin{bmatrix} \mathbf{u}^T(n) \boldsymbol{\phi}(n) \\ \mathbf{U}^T(n) \boldsymbol{\phi}(n) \end{bmatrix} \quad (3.41)$$

e pré-multiplicando-se $\boldsymbol{\phi}(n)$ por $\mathbf{u}^T(n)$, com o transposto de (2.50), tem-se que

$$\mathbf{u}^T(n) \boldsymbol{\phi}(n) = [\boldsymbol{\phi}^T(n) + \hat{\mathbf{a}}^T(n) \mathbf{U}^T(n)] \boldsymbol{\phi}(n) = \boldsymbol{\phi}^T(n) \boldsymbol{\phi}(n). \quad (3.42)$$

Uma vez que $\mathbf{U}(n) \hat{\mathbf{a}}(n)$ é ortogonal a $\boldsymbol{\phi}(n)$, ou ainda $\mathbf{U}^T(n) \boldsymbol{\phi}(n) = \mathbf{0}_{P \times 1}$, em virtude de $\boldsymbol{\phi}(n)$ ser ortogonal a todos os vetores coluna de $\mathbf{U}(n)$, a substituição de (3.42) em (3.41), resulta em

$$\mathbf{U}_u^T(n) \boldsymbol{\phi}(n) = \begin{bmatrix} \boldsymbol{\phi}^T(n) \boldsymbol{\phi}(n) \\ \mathbf{0}_{P \times 1} \end{bmatrix}. \quad (3.43)$$

Pré-multiplicando ambos os lados de (3.43) por $\mathbf{U}_u^+(n)$, tem-se

$$\mathbf{U}_u^+(n) \mathbf{U}_u^T(n) \boldsymbol{\phi}(n) = \mathbf{U}_u^+(n) \begin{bmatrix} \boldsymbol{\phi}^T(n) \boldsymbol{\phi}(n) \\ \mathbf{0}_{P \times 1} \end{bmatrix}. \quad (3.44)$$

A matriz $\mathbf{U}_u^+(n) \mathbf{U}_u^T(n)$ é a matriz de projeção no espaço das colunas de $\mathbf{U}_u(n)$, que é composta pelos vetores coluna $\mathbf{u}(n)$, $\mathbf{u}(n-1)$, ...

$\mathbf{u}(n-P)$. Como $\boldsymbol{\phi}(n)$ está contido no espaço das colunas de $\mathbf{U}_u(n)$, o lado esquerdo da equação é igual a $\boldsymbol{\phi}(n)$. Assim,

$$\boldsymbol{\phi}(n) = \boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)\mathbf{U}_u^+(n) \begin{bmatrix} 1 \\ \mathbf{0}_{p \times 1} \end{bmatrix} \quad (3.45)$$

logo,

$$\mathbf{U}_u^+(n) \begin{bmatrix} 1 \\ \mathbf{0}_{p \times 1} \end{bmatrix} = \frac{\boldsymbol{\phi}(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)}. \quad (3.46)$$

Dessa forma, substituindo (3.46) em (3.38) e (3.39), resulta em

$$\mathbf{w}_{\text{APA}}(n+1) = \begin{cases} = \mathbf{w}_{\text{APA}}(0) & p / n < 0 \\ = \mathbf{w}_{\text{APA}}(n) + \frac{\boldsymbol{\phi}(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} e_{\text{APA}}(n) & p / n = 0 \\ + \mathbf{U}_u^+(n) \begin{bmatrix} \mathbf{0}_{1 \times p} \\ \mathbf{1}_p \end{bmatrix} [\mathbf{r}(n-1) - \mathbf{U}^T(n) [\mathbf{w}_{\text{APA}}(0) - \mathbf{w}^o]] & \\ = \mathbf{w}_{\text{APA}}(n) + \frac{\boldsymbol{\phi}(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} e_{\text{APA}}(n) & p / n > 0 \end{cases} \quad (3.47)$$

podendo ser reescrita para $n \geq 0$ como

$$\mathbf{w}_{\text{APA}}(n+1)|_{n \geq 0} = \mathbf{w}_{\text{APA}}(n) + \frac{\boldsymbol{\phi}(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} e_{\text{APA}}(n) + \boldsymbol{\kappa}(n) \quad (3.48)$$

sendo a notação $\boldsymbol{\kappa}(n)$ definida por

$$\boldsymbol{\kappa}(n) \triangleq \mathbf{U}_u^+(n) \mathbf{Q}_\delta(n) \{ \mathbf{r}(n-1) - \mathbf{U}^T(n) [\mathbf{w}_{\text{APA}}(0) - \mathbf{w}^o] \} \quad (3.49)$$

e a matriz $\mathbf{Q}_\delta(n)$ por

$$\mathbf{Q}_\delta(n) \triangleq \delta(n) \times \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & 1 \end{bmatrix} \quad (3.50)$$

em que $\delta(n)$ é o delta de Kronecker (1 se $n = 0$ e 0 para outros valores de n). A finalidade da matriz $\mathbf{Q}_\delta(n)$ é similar a da matriz de funções degrau $\mathbf{Q}(-n+1)$: para $n \neq 0$, a matriz $\mathbf{Q}_\delta(n)$ é nula e $\boldsymbol{\kappa}(n)$ é descartado.

Nota-se que, assumindo a estacionariedade dos sinais envolvidos e a invariância no tempo da planta, a diferença entre as equações de atualização dos coeficientes do APA em (3.48) e do PAP em (2.47) é completamente determinada no primeiro instante da recursão dos coeficientes. Entretanto, a influência de $\boldsymbol{\kappa}(0)$ é conduzida até o infinito pela característica recursiva da equação do APA, gerando uma diferença no valor de $e(n)$ e $e_{\text{APA}}(n)$ a cada iteração. Dessa forma, a condição de passo unitário descrita por Bouteille et al. (1999) é necessária para a equivalência entre os algoritmos APA e PAP, porém não é suficiente.

Essa diferença é visível na relação determinística do APA, apresentada a seguir. Considerando-se apenas as linhas 2 à P da matriz na equação (3.36) e a relação (3.37), tem-se que

$$\mathbf{U}^T(n)\mathbf{v}_{\text{APA}}(n)\Big|_{n>0} = \mathbf{r}(n-1) \quad (3.51)$$

A Tabela 6 apresenta o resultado da igualdade (3.51) para o APA, considerando a equação (3.37).

Tabela 6- Relação iterativa de igualdade de $\mathbf{u}^T(n-i)\mathbf{v}(n)$ em (3.51) para o APA.

		$n = 0$	$n = 1$	$n = 2$	$n = 3$	$n = 4$
P	$i=1$	$\mathbf{u}^T(-1)\mathbf{v}(0)$	$r(0)$	$r(1)$	$r(2)$	$r(3)$
	$i=2$	$\mathbf{u}^T(-2)\mathbf{v}(0)$	$r(-1)$	$r(0)$	$r(1)$	$r(2)$
	$i=3$	$\mathbf{u}^T(-3)\mathbf{v}(0)$	$r(-2)$	$r(-1)$	$r(0)$	$r(1)$
	$i=4$	$\mathbf{u}^T(-4)\mathbf{v}(0)$	$r(-3)$	$r(-2)$	$r(-1)$	$r(0)$

Dessa forma, a ausência de $\boldsymbol{\kappa}(0)$ no PAP, conforme equações (2.47) e (3.48), caracteriza o Efeito de Inicialização. Isso se deve à estrutura escalar do erro na recursão dos coeficientes, resultando na diferença encontrada entre a Tabela 6 e a Tabela 5. A igualdade no comportamento entre ambos os algoritmos é uma condição necessária e suficiente, se e somente se, for utilizado passo unitário e a igualdade $\mathbf{U}^T(n)\mathbf{v}(n) = \mathbf{r}(n-1)$ for válida em $n \geq 0$, tendo-se as mesmas condições iniciais $\mathbf{w}(0) = \mathbf{w}_{\text{APA}}(0)$, e portanto, $\mathbf{e}(n) = [e(n) \mathbf{0}_{P \times 1}]^T$ com base na equação (3.37). Associadamente, nesse caso tem-se que $\boldsymbol{\kappa}(0) = \mathbf{0}_{N \times 1}$.

Conceitualmente, a ausência de $\boldsymbol{\kappa}(0)$ imposta pelo PAP, afeta a ortogonalização do vetor de coeficientes no sentido dos mínimos quadrados, que ocorre quando o vetor de erro gerado pela atualização dos coeficientes *a posteriori*, torna-se nulo. Ressalta-se, porém, que $\boldsymbol{\phi}(n)$ é sempre ortogonal às colunas de $\mathbf{U}(n)$ em todas as iterações.

3.5 Conclusão

Este capítulo apresentou inicialmente as considerações de análise que serão utilizadas neste trabalho, restringindo o conjunto de parâmetros para sinais de excitação autoregressivos e passo unitário. Foram utilizadas as mesmas hipóteses estatísticas empregadas nos trabalhos de Bershad; Linebarger e Mclaughlin (2001) e Almeida (2004), exceto pela solução do problema definido que se constitui da extensão na relação determinística (3.8) de Rupp (1998) no período inicial do regime transiente, modelando o Efeito de Inicialização. Enquanto o APA efetua a ortogonalização dos coeficientes no sentido dos mínimos quadrados com um vetor de erro $e(n)$, o PAP não possui tal comportamento em virtude da ausência dos elementos de $\kappa(0)$ devido ao erro escalar $e(n)$.

4 COMPORTAMENTO MÉDIO DOS COEFICIENTES

Este capítulo apresenta um novo modelo teórico para a previsão do comportamento médio dos coeficientes do pseudo algoritmo de projeções afins para o regime transitório e permanente. São utilizadas as mesmas hipóteses empregadas nos modelos de Bershad; Linebarger e Mclaughlin (2001) e Almeida (2004), porém com a extensão da relação determinística de Rupp (1998) apresentada na Seção 3.3, considerando o Efeito de Inicialização do PAP. Outras hipóteses são apresentadas no decorrer deste capítulo visando a tratabilidade matemática do problema. O capítulo se encerra com uma equação na forma fechada para o comportamento médio dos coeficientes.

4.1 Análise do Comportamento Médio dos Coeficientes

O comportamento médio do vetor de erro dos coeficientes pode ser avaliado substituindo-se (3.30) no penúltimo termo do lado direito da equação (3.11), resultando em

$$\mathbf{v}(n+1) = \begin{cases} = \mathbf{w}(0) - \mathbf{w}^0 & \text{para } n < 0 \\ = \mathbf{v}(n) - \frac{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \mathbf{v}(n) \\ + \frac{r(n)\boldsymbol{\phi}(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} - \frac{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \mathbf{r}(n-1) \\ - \frac{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \mathbf{Q}(-n+1)\mathbf{U}^T(n)\mathbf{v}(0) \\ + \frac{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \mathbf{Q}(-n+1)\mathbf{r}(n-1) & \text{para } n \geq 0. \end{cases} \quad (4.1)$$

Considerando-se apenas as iterações para $n \geq 0$, e tomando-se o valor esperado de (4.1) condicionado em $\mathbf{v}(n)$, tem-se que

$$\begin{aligned}
E \left\{ \mathbf{v}(n+1) \middle| \mathbf{v}(n) \right\} &= \mathbf{v}(n) - E \left\{ \frac{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \middle| \mathbf{v}(n) \right\} \mathbf{v}(n) \\
&+ E \left\{ \frac{\boldsymbol{\phi}(n)r(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \middle| \mathbf{v}(n) \right\} - E \left\{ \frac{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{r}(n-1)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \middle| \mathbf{v}(n) \right\} \\
&- E \left\{ \frac{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \mathbf{Q}(-n+1)\mathbf{U}^T(n) \middle| \mathbf{v}(n) \right\} \mathbf{v}(0) \\
&+ E \left\{ \frac{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \mathbf{Q}(-n+1)\mathbf{r}(n-1) \middle| \mathbf{v}(n) \right\}.
\end{aligned} \tag{4.2}$$

Assumindo-se a independência estatística entre o processo aleatório $r(n)$ e o sinal de excitação $u(n)$, o segundo, terceiro e quinto termos de (4.2) são nulos. Considera-se nessa abordagem que $\boldsymbol{\phi}(n)$ e $\hat{\mathbf{a}}(n)$ (funções do processo $u(n)$) são independentes do ruído e de suas amostras passadas que são estatisticamente independentes, identicamente distribuídas (i.i.d) e de média zero. Então,

$$\begin{aligned}
E \left\{ \mathbf{v}(n+1) \middle| \mathbf{v}(n) \right\} &= \mathbf{v}(n) - E \left\{ \frac{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \middle| \mathbf{v}(n) \right\} \mathbf{v}(n) \\
&+ E \left\{ \frac{\boldsymbol{\phi}(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \middle| \mathbf{v}(n) \right\} \underbrace{E \left\{ r(n) \middle| \mathbf{v}(n) \right\}}_0 \\
&- E \left\{ \frac{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \middle| \mathbf{v}(n) \right\} \underbrace{E \left\{ \mathbf{r}(n-1) \middle| \mathbf{v}(n) \right\}}_{\mathbf{0}_{p \times 1}} \\
&- E \left\{ \frac{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \mathbf{Q}(-n+1)\mathbf{U}^T(n) \middle| \mathbf{v}(n) \right\} \mathbf{v}(0) \\
&+ E \left\{ \frac{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \mathbf{Q}(-n+1) \middle| \mathbf{v}(n) \right\} \underbrace{E \left\{ \mathbf{r}(n-1) \middle| \mathbf{v}(n) \right\}}_{\mathbf{0}_{p \times 1}}.
\end{aligned} \tag{4.3}$$

O primeiro valor esperado do lado direito de (4.3) é avaliado assumindo que o numerador e o denominador são fracamente correlacionados. De acordo com Almeida, cada elemento

[...] tem um numerador dado pelo produto $\phi(n-i)\phi(n-j)$ e um denominador dado pelo somatório $\sum_{k=1}^N \phi(n-k)$. Os componentes de $\phi(n)$ no numerador afetam apenas dois dos N termos no denominador. Por consequência, numerador e denominador podem ser assumidos fracamente correlacionados para valores grandes de N . Para entradas ergódicas, esta hipótese é equivalente a aplicar o princípio das médias (SAMSON e REDDY, 1983). O denominador $\phi^T(n)\phi(n)$ tende a variar no tempo lentamente quando comparado com o numerador $\phi(n-i)\phi(n-j)$ para elevados valores de N . (ALMEIDA, 2004, p. 62)

Assim, a seguinte hipótese é postulada:

- **H5:** *Seja a esperança $E\{[\phi^T(n)\phi(n)]^{-1} \cdot \phi(n)\phi^T(n)\}$. O denominador deste produto tende a variar mais lentamente comparado ao numerador para valores elevados de N , ou seja, eles podem ser assumidos fracamente correlacionados.*

e dessa forma, a aproximação resulta em (ALMEIDA, 2004, p. 62):

$$\begin{aligned} & E \left\{ \left[\phi^T(n)\phi(n) \right]^{-1} \phi(n)\phi^T(n) \right\} \\ & \approx E \left\{ \left[\phi^T(n)\phi(n) \right]^{-1} \right\} E \left\{ \phi(n)\phi^T(n) \right\}. \end{aligned} \quad (4.4)$$

Assim, assume-se que a hipótese **H5** é igualmente válida para o valor esperado do quarto termo de (4.3), tendo em vista que seus elementos são compostos pelo sinal de entrada. Dessa forma, (4.3) resulta em

$$\begin{aligned} & E \left\{ \mathbf{v}(n+1) \mid \mathbf{v}(n) \right\} = \mathbf{v}(n) \\ & - E \left\{ \left[\phi^T(n)\phi(n) \right]^{-1} \mid \mathbf{v}(n) \right\} E \left\{ \phi(n)\phi^T(n) \mid \mathbf{v}(n) \right\} \mathbf{v}(n) \\ & - E \left\{ \left[\phi^T(n)\phi(n) \right]^{-1} \mid \mathbf{v}(n) \right\} E \left\{ \phi(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n) \mid \mathbf{v}(n) \right\} \mathbf{v}(0). \end{aligned} \quad (4.5)$$

A equação (4.5) não pode ser diretamente calculada uma vez que a função densidade de probabilidade (FDP) de $\mathbf{v}(n)$ não é conhecida. Entretanto, a probabilidade condicional de (4.2) pode ser removida assumindo as seguintes hipóteses:

- **H6:** Os elementos do vetor de predição linear $\hat{\mathbf{u}}(n) = \mathbf{U}(n)\hat{\mathbf{a}}(n)$ (veja equação (A.2)) podem ser assumidos conjuntamente Gaussianos, devido à hipótese **H4**, e independentes de $\mathbf{v}(n)$ pela teoria da independência (MAZO, 1979). O mesmo pode ser considerado para $\hat{\mathbf{u}}(n) = \mathbf{U}(n)\mathbf{Q}(-n+1)\hat{\mathbf{a}}(n)$, uma vez que $\mathbf{Q}(-n+1)$ é uma matriz determinística sendo não-nula apenas nas P primeiras iterações da atualização dos coeficientes.
- **H7:** Para $N \gg P$, as flutuações do estimador linear $\hat{\mathbf{a}}(n)$ são muito menores do que seu valor médio, assim ele pode ser aproximado à um vetor invariante no tempo \mathbf{a} (BERSHAD; LINEBARGER e MCLAUGHLIN, 2001), desde que $\phi(n)$ tenha média zero (veja também a Propriedade 1 no final do Apêndice A).

Tomando-se o valor esperado de (4.5) e considerando as hipóteses **H6** e **H7**, de forma a retirar o condicionamento, tem-se a seguinte expressão para o comportamento médio dos coeficientes,

$$\begin{aligned}
 E\{\mathbf{v}(n+1)\} &\approx E\{\mathbf{v}(n)\} \\
 &- E\left\{\left[\boldsymbol{\Phi}^T(n)\boldsymbol{\Phi}(n)\right]^{-1}\right\} E\{\boldsymbol{\Phi}(n)\boldsymbol{\Phi}^T(n)\} E\{\mathbf{v}(n)\} \\
 &- E\left\{\left[\boldsymbol{\Phi}^T(n)\boldsymbol{\Phi}(n)\right]^{-1}\right\} \mathbf{R}_{\phi\hat{\mathbf{u}}}(n)\mathbf{v}(0)
 \end{aligned} \tag{4.6}$$

sendo a matriz $\mathbf{R}_{\phi\hat{\mathbf{u}}}(n)$ de dimensão $N \times N$ definida por

$$E\{\boldsymbol{\Phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\} \triangleq \mathbf{R}_{\phi\hat{\mathbf{u}}}(n). \tag{4.7}$$

Nota-se que esta matriz pode ser estimada através das estatísticas do sinal de excitação, assumindo estacionaridade no sentido amplo (WSS). Além disso, $\mathbf{R}_{\phi\hat{\mathbf{u}}}(n)$ varia deterministicamente conforme as iterações se sucedem.

No Apêndice B é derivada uma equação determinística, em função dos parâmetros do sinal de entrada AR, para os elementos da l -ésima linha e c -ésima coluna de $\mathbf{R}_{\phi\hat{\mathbf{u}}}(n)$, resultando em

$$\begin{aligned}
 r_{\phi\hat{\mathbf{u}}(l,c)}(n) &\cong \sum_{j=1}^P r_{uu(l,j+c)} a_j u_{-1}(-n+j) \\
 &- \sum_{i=1}^P \sum_{j=1}^P r_{uu(i+l,j+c)} a_i a_j u_{-1}(-n+j)
 \end{aligned} \tag{4.8}$$

em que $u_{\cdot 1}(k)$ é uma função degrau unitário (1 se $k \geq 0$ e 0 para outros valores de k) e $r_{uu(a,b)} = E\{u(n-a)u(n-b)\}$.

Os demais valores esperados em (4.6) são avaliados através de (3.1) e da equação (3.33) de Almeida (2004). Assim,

$$E\left\{\left[\boldsymbol{\Phi}^T(n)\boldsymbol{\Phi}(n)\right]^{-1}\right\} \cong \frac{N}{G(G-2)\sigma_z^2} \quad (4.9)$$

em que

$$G = N - P \quad (4.10)$$

são os *graus de liberdade da distribuição chi-quadrada de* $[\boldsymbol{\Phi}^T(n)\boldsymbol{\Phi}(n)]^{-1}$. Considerando a validade da hipótese **H1**, $N \gg P$, a distribuição chi-quadrada tende a se aproximar de uma Gaussiana.

Substituindo-se (3.1) e (4.9) em (4.6), resulta em

$$E\{\mathbf{v}(n+1)\} \cong \beta_1 E\{\mathbf{v}(n)\} + \beta_2 \mathbf{R}_{\phi_{ii}}(n) \mathbf{v}(0) \quad (4.11)$$

em que

$$\beta_1 = \left(1 - \frac{1}{G-2}\right) \quad (4.12)$$

e

$$\beta_2 = -\left(\frac{N}{G(G-2)\sigma_z^2}\right). \quad (4.13)$$

Substituindo-se $\mathbf{v}(n) = \mathbf{w}(n) - \mathbf{w}^o$ em (4.11), obtém-se o modelo do comportamento médio dos coeficientes para o PAP:

$$E\{\mathbf{w}(n+1)\} \cong \beta_1 E\{\mathbf{w}(n)\} + \beta_2 \mathbf{R}_{\phi_{ii}}(n) \mathbf{v}(0) + \frac{1}{G-2} \mathbf{w}^o. \quad (4.14)$$

4.2 Forma Fechada para o Vetor de Erro dos Coeficientes

Uma expressão fechada para descrever o comportamento médio do vetor de erro dos coeficientes pode ser obtida através de substituição iterativa. Reescrevendo a equação (4.11) como sendo

$$E\{\mathbf{v}(n+1)\} \cong \beta_1 E\{\mathbf{v}(n)\} + \beta_2 \mathbf{c}(n) \quad (4.15)$$

em que

$$\mathbf{c}(n) = \mathbf{R}_{\phi_{ii}}(n) \mathbf{v}(0) \quad (4.16)$$

tem-se que

$$\left\{ \begin{array}{l} n=0 \rightarrow E\{\mathbf{v}(1)\} \cong \beta_1 \mathbf{v}(0) + \beta_2 \mathbf{c}(0) \\ n=1 \rightarrow E\{\mathbf{v}(2)\} \cong \beta_1 E\{\mathbf{v}(1)\} + \beta_2 \mathbf{c}(1) = \beta_1 [\beta_1 \mathbf{v}(0) + \beta_2 \mathbf{c}(0)] + \beta_2 \mathbf{c}(1) \\ \qquad \qquad \qquad = \beta_1^2 \mathbf{v}(0) + \beta_1 \beta_2 \mathbf{c}(0) + \beta_2 \mathbf{c}(1) \\ n=2 \rightarrow E\{\mathbf{v}(3)\} \cong \beta_1 E\{\mathbf{v}(2)\} + \beta_2 \mathbf{c}(2) \\ \qquad \qquad \qquad \cong \beta_1 [\beta_1^2 \mathbf{v}(0) + \beta_1 \beta_2 \mathbf{c}(0) + \beta_2 \mathbf{c}(1)] + \beta_2 \mathbf{c}(2) \\ \qquad \qquad \qquad = \beta_1^3 \mathbf{v}(0) + \beta_1^2 \beta_2 \mathbf{c}(0) + \beta_1 \beta_2 \mathbf{c}(1) + \beta_2 \mathbf{c}(2) \\ \qquad \qquad \qquad = \beta_1^3 \mathbf{v}(0) + \sum_{i=0}^2 \beta_1^{2-i} \beta_2 \mathbf{c}(i). \end{array} \right.$$

Esse procedimento é repetido consecutivamente. Efetuando-se uma troca de variável $l = n+1$, a equação fechada reduz-se à

$$\begin{aligned} E\{\mathbf{v}(l)\} &\cong \beta_1 E\{\mathbf{v}(l-1)\} + \beta_2 \mathbf{c}(l-1) \\ &\cong \beta_1^l \mathbf{v}(0) + \sum_{i=0}^{l-1} (\beta_1^{l-i-1} \beta_2 \mathbf{c}(i)) \\ &= \beta_1^l \mathbf{v}(0) + \sum_{i=0}^{l-1} (\beta_1^{l-i-1} \beta_2 \mathbf{R}_{\phi_{ii}}(i) \mathbf{v}(0)). \end{aligned} \quad (4.17)$$

Entretanto, em $n \geq P$ a matriz de correlação $\mathbf{R}_{\phi_{ii}}(n)$ torna-se nula e, portanto, a contribuição de $\mathbf{c}(n)$ também será nula. Dessa forma, em razão da matriz de seleção, a recursão da equação (4.15) em $n = P$, resulta em

$$\begin{aligned} E\{\mathbf{v}(P+1)\} &\cong \beta_1 E\{\mathbf{v}(P)\} + \beta_2 \mathbf{c}(P) \\ &\cong \beta_1 \left[\beta_1^P \mathbf{v}(0) + \beta_1^{P-1} \beta_2 \mathbf{c}(0) + \dots + \beta_2 \mathbf{c}(P-1) \right] + \beta_2 \mathbf{c}(P) \\ &= \beta_1^{P+1} \mathbf{v}(0) + \beta_1 \left[\beta_1^{P-1} \beta_2 \mathbf{c}(0) + \dots + \beta_2 \mathbf{c}(P-1) \right] + \underbrace{\beta_2 \mathbf{c}(P)}_{\mathbf{0}_{N \times 1}} \quad (4.18) \\ &= \beta_1^{P+1} \mathbf{v}(0) + \beta_1 \sum_{i=0}^{P-1} \beta_1^{P-i-1} \beta_2 \mathbf{c}(i). \end{aligned}$$

e assim consecutivamente.

E assim, para $l \geq P+1$, tem-se que

$$\begin{aligned} E\{\mathbf{v}(l)\} &\cong \beta_1 E\{\mathbf{v}(l-1)\} + \beta_2 \mathbf{c}(l-1) \\ &\cong \beta_1^l \mathbf{v}(0) + \beta_1^{l-P} \sum_{i=0}^{P-1} (\beta_1^{P-i-1} \beta_2 \mathbf{c}(i)) \end{aligned} \quad (4.19)$$

Realizando-se uma nova substituição de variável, em que $n = l$ para $n > 0$, obtém-se a equação para o comportamento médio do vetor de erro dos coeficientes no instante n , que resulta em

$$E\{\mathbf{v}(n)\} = \begin{cases} = \mathbf{v}(0) & p/ n \leq 0 \\ \cong \left[\beta_1^n \mathbf{I}_N + \sum_{i=0}^{n-1} \left(\beta_1^{n-i-1} \beta_2 \mathbf{R}_{\phi_{\tilde{u}}}(i) \right) \right] \mathbf{v}(0) & p/ 0 < n \leq P \\ \cong \beta_1^n \mathbf{I}_N \mathbf{v}(0) + \beta_1^{n-P} \mathbf{g} & p/ n > P \end{cases} \quad (4.20)$$

sendo que \mathbf{I}_N é uma matriz identidade de dimensões $N \times N$ e o vetor constante \mathbf{g} é definido como

$$\mathbf{g} \triangleq \sum_{i=0}^{P-1} \left(\beta_1^{P-i-1} \beta_2 \mathbf{R}_{\phi_{\tilde{u}}}(i) \right) \mathbf{v}(0). \quad (4.21)$$

O vetor \mathbf{g} expressa a memória intrínseca associada ao Efeito de Inicialização adquirida nas primeiras P iterações do algoritmo.

Nota-se que a equação (4.20) possui três comportamentos distintos:

- até o início da atualização dos coeficientes, a resposta é o próprio vetor inicial de erro dos coeficientes, $\mathbf{v}(0)$;
- entre $0 < n \leq P$, o Efeito de Inicialização é modelado pela recursão do algoritmo, compreendendo as parcelas do somatório;
- para $n > P$, as novas parcelas do termo gerador que são dependentes de $\mathbf{Q}(-n+1)$ são descartadas, mas devido à característica recursiva da equação, resquícios do Efeito de Inicialização são levados até o infinito. A memória recursiva gerada pelo termo extra até a P -ésima iteração resulta no vetor constante \mathbf{g} , sendo reduzida exponencialmente devido a β_1^{n-P-1} , conforme as iterações se sucedem.

Finalmente, através da terceira igualdade em (4.20), verifica-se que, em regime permanente, o PAP resulta em uma solução não polarizada, visto que $\lim_{n \rightarrow \infty} E\{\mathbf{v}(n)\} \rightarrow \mathbf{0}$, e a série converge para valores de $\beta_1 < 1$, e $G > 2$, conforme (4.14).

4.3 Conclusão

Neste capítulo foi apresentado um novo modelo analítico para o comportamento médio dos coeficientes do pseudo algoritmo de projeções afins no regime transiente. As equações (4.11) e (4.14) expandem o resultado obtido por Almeida (2004). Obteve-se ainda, uma expressão fechada em (4.20) que permite determinar uma condição necessária para a estabilidade do algoritmo.

5 ERRO QUADRÁTICO MÉDIO

Neste capítulo é derivado um novo modelo para a descrição do erro quadrático médio (EQM) para o pseudo algoritmo de projeções afins, considerando o Efeito de Inicialização. Esse procedimento tem o objetivo de corrigir a modelagem no período inicial de transiente em relação aos modelos apresentados em Bershad; Linebarger e McLaughlin (2001), e Almeida (2005). Neste capítulo também são derivadas expressões para o traço dos momentos de segunda ordem do vetor de erro dos coeficientes e para o comportamento quadrático médio do modelo em regime permanente.

5.1 Erro Quadrático Médio em Regime Transiente

O erro escalar instantâneo $e(n)$ do filtro adaptativo para o pseudo algoritmo de projeções afins pode ser escrito em função do vetor de erro dos coeficientes através das equações (2.17), (2.18), (2.19) e (3.2), resultando em

$$e(n) = r(n) - \mathbf{u}^T(n)\mathbf{v}(n). \quad (5.1)$$

Aplicando-se $\mathbf{u}^T(n) = \boldsymbol{\phi}^T(n) + \hat{\mathbf{a}}^T(n)\mathbf{U}^T(n)$, através de (2.50), em (5.1), tem-se que

$$e(n) = r(n) - \boldsymbol{\phi}^T(n)\mathbf{v}(n) - \hat{\mathbf{a}}^T(n)\mathbf{U}^T(n)\mathbf{v}(n) \quad (5.2)$$

e por conseguinte, substituindo-se (3.30) em (5.2), resulta em

$$\begin{aligned} e(n) = & r_a(n) - \boldsymbol{\phi}^T(n)\mathbf{v}(n) + \hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{r}(n-1) \\ & - \hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\mathbf{v}(0) \end{aligned} \quad (5.3)$$

em que

$$r_a(n) = r(n) - \hat{\mathbf{a}}^T(n)\mathbf{r}(n-1). \quad (5.4)$$

Elevando-se ambos os lados de (5.3) ao quadrado, e considerando que $\text{tr}\{x\} = x$, resulta na seguinte expressão para o erro quadrático instantâneo,

$$\begin{aligned}
e^2(n) &= r_a^2(n) - 2r_a(n)\boldsymbol{\phi}^T(n)\mathbf{v}(n) + \mathbf{v}^T(n)\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\mathbf{v}(n) \\
&\quad - 2r_a(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\mathbf{v}(0) \\
&\quad + 2r_a(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{r}(n-1) \\
&\quad + 2\mathbf{v}^T(n)\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\mathbf{v}(0) \\
&\quad - 2\mathbf{v}^T(n)\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{r}(n-1) \\
&\quad - 2\mathbf{v}^T(0)\mathbf{U}(n)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{r}(n-1) \\
&\quad + \mathbf{v}^T(0)\mathbf{U}(n)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\mathbf{v}(0) \\
&\quad + \mathbf{r}^T(n-1)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{r}(n-1).
\end{aligned} \tag{5.5}$$

O erro quadrático médio, condicionado no vetor de erro dos coeficientes, pode ser obtido de (5.5):

$$\begin{aligned}
E\{e^2(n) \mid \mathbf{v}(n)\} &= E\{r_a^2(n) \mid \mathbf{v}(n)\} - 2E\{r_a(n)\boldsymbol{\phi}^T(n) \mid \mathbf{v}(n)\} \mathbf{v}(n) \\
&\quad + \mathbf{v}^T(n)E\{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n) \mid \mathbf{v}(n)\} \mathbf{v}(n) \\
&\quad - 2E\{r_a(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n) \mid \mathbf{v}(n)\} \mathbf{v}(0) \\
&\quad + 2E\{r_a(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{r}(n-1) \mid \mathbf{v}(n)\} \\
&\quad + 2\mathbf{v}^T(n)E\{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n) \mid \mathbf{v}(n)\} \mathbf{v}(0) \\
&\quad - 2\mathbf{v}^T(n)E\{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{r}(n-1) \mid \mathbf{v}(n)\} \\
&\quad - 2\mathbf{v}^T(0)E\{\mathbf{U}(n)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n) \\
&\quad \quad \cdot \hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{r}(n-1) \mid \mathbf{v}(n)\} \\
&\quad + \mathbf{v}^T(0)E\{[\mathbf{U}(n)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)]^2 \mid \mathbf{v}(n)\} \mathbf{v}(0) \\
&\quad + E\{[\mathbf{r}^T(n-1)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)]^2 \mid \mathbf{v}(n)\}.
\end{aligned} \tag{5.6}$$

Apesar do condicionamento no vetor $\mathbf{v}(n)$, o *primeiro termo* do lado direito da equação (5.6) pode ser avaliado de forma similar à utilizada em Almeida (2004). Dessa forma, elevando (5.4) ao quadrado e tomando o seu valor esperado condicionado em $\mathbf{v}(n)$, tem-se que

$$\begin{aligned}
E \left\{ r_a^2(n) \mid \mathbf{v}(n) \right\} &\cong E \left\{ r^2(n) \mid \mathbf{v}(n) \right\} \\
&\quad - 2 \underbrace{E \left\{ r(n) \mathbf{r}^T(n-1) \hat{\mathbf{a}}(n) \mid \mathbf{v}(n) \right\}}_0 \\
&\quad + E \left\{ \hat{\mathbf{a}}^T(n) \mathbf{r}(n-1) \mathbf{r}^T(n-1) \hat{\mathbf{a}}(n) \mid \mathbf{v}(n) \right\}.
\end{aligned} \tag{5.7}$$

O segundo termo do lado direito da equação (5.7) pode ser desprezado assumindo-se que o ruído é estatisticamente independente de suas amostras passadas e de $\hat{\mathbf{a}}(n)$, já que o vetor $\hat{\mathbf{a}}(n)$ é função de $u(n)$. Utilizando-se a propriedade de que $\text{tr}\{x\} = x$ e $\text{tr}\{\mathbf{AB}\} = \text{tr}\{\mathbf{BA}\}$, a equação (5.7) pode então ser expressa por

$$\begin{aligned}
E \left\{ r_a^2(n) \mid \mathbf{v}(n) \right\} &\cong E \left\{ r^2(n) \mid \mathbf{v}(n) \right\} \\
&\quad + \text{tr} \left\{ E \left\{ \mathbf{r}(n-1) \mathbf{r}^T(n-1) \mid \mathbf{v}(n) \right\} E \left\{ \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T(n) \mid \mathbf{v}(n) \right\} \right\}.
\end{aligned} \tag{5.8}$$

Assumindo-se ainda que o ruído de medição é independente e identicamente distribuído, e que $E\{r^2(n) \mid \mathbf{v}(n)\} = \sigma_r^2$, tem-se que

$$E \left\{ \mathbf{r}(n-1) \mathbf{r}^T(n-1) \mid \mathbf{v}(n) \right\} = \sigma_r^2 \mathbf{I}_P \tag{5.9}$$

sendo que \mathbf{I}_P é uma matriz identidade de dimensões $P \times P$. Dessa forma, $E\{r_a^2(n) \mid \mathbf{v}(n)\}$ resulta na expressão

$$E \left\{ r_a^2(n) \mid \mathbf{v}(n) \right\} \cong \sigma_r^2 \left[1 + \text{tr} \left\{ E \left\{ \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T(n) \mid \mathbf{v}(n) \right\} \right\} \right] \tag{5.10}$$

Retomando-se a equação (5.6), o *segundo termo* do lado direito pode ser desprezado substituindo-se (5.4) em (5.6), o que leva à

$$\begin{aligned}
&E \left\{ r_a(n) \boldsymbol{\phi}^T(n) \mid \mathbf{v}(n) \right\} \\
&= E \left\{ \left[r(n) - \hat{\mathbf{a}}^T(n) \mathbf{r}(n-1) \right] \boldsymbol{\phi}^T(n) \mid \mathbf{v}(n) \right\} \\
&= E \left\{ r(n) \boldsymbol{\phi}^T(n) \mid \mathbf{v}(n) \right\} - E \left\{ \hat{\mathbf{a}}^T(n) \mathbf{r}(n-1) \boldsymbol{\phi}^T(n) \mid \mathbf{v}(n) \right\}.
\end{aligned} \tag{5.11}$$

Supondo que $\mathbf{v}(n)$ e $\boldsymbol{\phi}(n)$ sejam estatisticamente independentes (hipótese **H3**), e assumindo-se que o ruído aditivo é i.i.d., de média zero e estatisticamente independente do vetor de entrada $\mathbf{u}(n)$, e que $\boldsymbol{\phi}(n)$ e $\hat{\mathbf{a}}(n)$ são funções do sinal de entrada, tem-se que

$$\begin{aligned}
& E \left\{ \boldsymbol{\phi}^T(n) r_a(n) \mid \mathbf{v}(n) \right\} \\
& \cong E \left\{ \boldsymbol{\phi}^T(n) \mid \mathbf{v}(n) \right\} \underbrace{E \left\{ r(n) \mid \mathbf{v}(n) \right\}}_0 \\
& - E \left\{ \boldsymbol{\phi}^T(n) \hat{\mathbf{a}}^T(n) \mid \mathbf{v}(n) \right\} \underbrace{E \left\{ \mathbf{r}(n-1) \mid \mathbf{v}(n) \right\}}_{\mathbf{0}_{p \times 1}} \\
& = \mathbf{0}_{N \times 1}^T
\end{aligned} \tag{5.12}$$

em que $\mathbf{0}_{N \times 1}$ é um vetor coluna de zeros.

Aplicando-se a propriedade $\text{tr}\{\mathbf{A}^T \mathbf{B}\} = \text{tr}\{\mathbf{B} \mathbf{A}^T\}$ ao *terceiro termo* escalar do lado direito da equação (5.6), obtém-se que

$$\begin{aligned}
& \mathbf{v}^T(n) E \left\{ \boldsymbol{\phi}(n) \boldsymbol{\phi}^T(n) \mid \mathbf{v}(n) \right\} \mathbf{v}(n) \\
& = \text{tr} \left\{ E \left\{ \boldsymbol{\phi}(n) \boldsymbol{\phi}^T(n) \mid \mathbf{v}(n) \right\} \mathbf{v}(n) \mathbf{v}^T(n) \right\}.
\end{aligned} \tag{5.13}$$

O valor esperado no *quarto termo* em (5.6), por sua vez, pode ser avaliado utilizando-se as mesmas considerações utilizadas em (5.12). Substituindo-se (5.4), obtém-se que

$$\begin{aligned}
& E \left\{ \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{U}^T(n) r_a(n) \mid \mathbf{v}(n) \right\} \\
& = E \left\{ \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{U}^T(n) \left[r(n) - \hat{\mathbf{a}}^T(n) \mathbf{r}(n-1) \right] \mid \mathbf{v}(n) \right\} \\
& = E \left\{ \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{U}^T(n) r(n) \mid \mathbf{v}(n) \right\} \\
& - E \left\{ \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{U}^T(n) \hat{\mathbf{a}}^T(n) \mathbf{r}(n-1) \mid \mathbf{v}(n) \right\} \\
& \cong E \left\{ \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{U}^T(n) \mid \mathbf{v}(n) \right\} \underbrace{E \left\{ r(n) \mid \mathbf{v}(n) \right\}}_0 \\
& - E \left\{ \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{U}^T(n) \hat{\mathbf{a}}^T(n) \mid \mathbf{v}(n) \right\} \underbrace{E \left\{ \mathbf{r}(n-1) \mid \mathbf{v}(n) \right\}}_{\mathbf{0}_{p \times 1}} \\
& = \mathbf{0}_{N \times 1}^T.
\end{aligned} \tag{5.14}$$

O *quinto termo* do lado direito de (5.6) pode ser simplificado através de (5.4) e considerando-se que o ruído é estatisticamente independente e de média zero,

$$\begin{aligned}
& E \left\{ r_a(n) \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{r}(n-1) \mid \mathbf{v}(n) \right\} \\
&= E \left\{ \left[r(n) \right] \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{r}(n-1) \mid \mathbf{v}(n) \right\} \\
&\quad - E \left\{ \left[\hat{\mathbf{a}}^T(n) \mathbf{r}(n-1) \right] \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{r}(n-1) \mid \mathbf{v}(n) \right\} \quad (5.15) \\
&\cong \underbrace{E \left\{ r(n) \mid \mathbf{v}(n) \right\}}_0 E \left\{ \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{r}(n-1) \mid \mathbf{v}(n) \right\} \\
&\quad - E \left\{ \left[\hat{\mathbf{a}}^T(n) \mathbf{r}(n-1) \right] \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{r}(n-1) \mid \mathbf{v}(n) \right\}.
\end{aligned}$$

Todavia, o segundo termo do lado direito não é nulo devido aos momentos de segunda ordem do ruído de medida. Aplicando-se o operador traço, e assumindo novamente a independência estatística entre o ruído e o vetor de entrada, tem-se que

$$\begin{aligned}
& E \left\{ \hat{\mathbf{a}}^T(n) \mathbf{r}(n-1) \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{r}(n-1) \mid \mathbf{v}(n) \right\} \\
&= E \left\{ \hat{\mathbf{a}}^T(n) \mathbf{r}(n-1) \mathbf{r}^T(n-1) \mathbf{Q}^T(-n+1) \hat{\mathbf{a}}(n) \mid \mathbf{v}(n) \right\} \\
&= \text{tr} \left\{ E \left\{ \mathbf{r}(n-1) \mathbf{r}^T(n-1) \mathbf{Q}^T(-n+1) \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T(n) \mid \mathbf{v}(n) \right\} \right\} \\
&\cong \text{tr} \left\{ E \left\{ \mathbf{r}(n-1) \mathbf{r}^T(n-1) \mid \mathbf{v}(n) \right\} \mathbf{Q}^T(-n+1) E \left\{ \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T(n) \mid \mathbf{v}(n) \right\} \right\} \quad (5.16)
\end{aligned}$$

em que a potência da matriz de correlação de ruído é dada pela equação (5.9), resultando na seguinte aproximação para o termo,

$$\begin{aligned}
& E \left\{ r_a(n) \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{r}(n-1) \mid \mathbf{v}(n) \right\} \\
&\cong \sigma_r^2 \text{tr} \left\{ \mathbf{Q}^T(-n+1) E \left\{ \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T(n) \mid \mathbf{v}(n) \right\} \right\}. \quad (5.17)
\end{aligned}$$

Ainda sobre o lado direito da equação (5.6), o *sétimo valor esperado* pode ser desprezado, considerando que ambos os vetores $\boldsymbol{\phi}(n)$ e $\hat{\mathbf{a}}(n)$ não possuem correlação com o ruído, e $\mathbf{Q}(-n+1)$ é determinístico, ou seja,

$$\begin{aligned}
& E \left\{ \boldsymbol{\phi}(n) \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{r}(n-1) \mid \mathbf{v}(n) \right\} \\
&\cong E \left\{ \boldsymbol{\phi}(n) \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mid \mathbf{v}(n) \right\} \underbrace{E \left\{ \mathbf{r}(n-1) \mid \mathbf{v}(n) \right\}}_{\mathbf{0}_{p \times 1}} \quad (5.18) \\
&= \mathbf{0}_{N \times 1}.
\end{aligned}$$

O oitavo termo é avaliado utilizando-se as mesmas considerações do sétimo termo, resultando em

$$\begin{aligned}
 & E \left\{ \mathbf{U}(n) \mathbf{Q}^T (-n+1) \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T (n) \mathbf{Q} (-n+1) \mathbf{r}(n-1) \middle| \mathbf{v}(n) \right\} \\
 & \cong E \left\{ \mathbf{U}(n) \mathbf{Q}^T (-n+1) \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T (n) \mathbf{Q} (-n+1) \middle| \mathbf{v}(n) \right\} \\
 & \quad \times E \left\{ \underbrace{\mathbf{r}(n-1)}_{\mathbf{0}_{p \times 1}} \middle| \mathbf{v}(n) \right\} \\
 & = \mathbf{0}_{N \times 1}.
 \end{aligned} \tag{5.19}$$

Finalmente, o último termo escalar do lado direito de (5.6), é avaliado de forma similar ao quinto termo, de forma que

$$\begin{aligned}
 & E \left\{ \left[\hat{\mathbf{a}}^T (n) \mathbf{Q} (-n+1) \mathbf{r}(n-1) \right]^2 \middle| \mathbf{v}(n) \right\} \\
 & = E \left\{ \hat{\mathbf{a}}^T (n) \mathbf{Q} (-n+1) \mathbf{r}(n-1) \mathbf{r}^T (n-1) \right. \\
 & \quad \left. \times \mathbf{Q}^T (-n+1) \hat{\mathbf{a}}(n) \middle| \mathbf{v}(n) \right\}.
 \end{aligned} \tag{5.20}$$

Aplicando-se a propriedade algébrica $\text{tr}\{\mathbf{A}^T \mathbf{B}\} = \text{tr}\{\mathbf{B} \mathbf{A}^T\}$, resulta em

$$\begin{aligned}
 & E \left\{ \left[\hat{\mathbf{a}}^T (n) \mathbf{Q} (-n+1) \mathbf{r}(n-1) \right]^2 \middle| \mathbf{v}(n) \right\} \\
 & = \text{tr} \left\{ \mathbf{Q} (-n+1) E \left\{ \mathbf{r}(n-1) \mathbf{r}^T (n-1) \middle| \mathbf{v}(n) \right\} \mathbf{Q}^T (-n+1) \right. \\
 & \quad \left. \times E \left\{ \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T (n) \middle| \mathbf{v}(n) \right\} \right\}
 \end{aligned} \tag{5.21}$$

e substituindo-se (5.9) em (5.22), obtém-se

$$\begin{aligned}
 & E \left\{ \left[\hat{\mathbf{a}}^T (n) \mathbf{Q} (-n+1) \mathbf{r}(n-1) \right]^2 \middle| \mathbf{v}(n) \right\} \\
 & \cong \text{tr} \left\{ \mathbf{Q} (-n+1) \sigma_r^2 \mathbf{I}_p \mathbf{Q}^T (-n+1) E \left\{ \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T (n) \middle| \mathbf{v}(n) \right\} \right\} \\
 & = \sigma_r^2 \text{tr} \left\{ \mathbf{Q} (-n+1) \mathbf{Q}^T (-n+1) E \left\{ \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T (n) \middle| \mathbf{v}(n) \right\} \right\} \\
 & = \sigma_r^2 \text{tr} \left\{ \mathbf{Q} (-n+1) E \left\{ \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T (n) \middle| \mathbf{v}(n) \right\} \right\}.
 \end{aligned} \tag{5.23}$$

Embora a função densidade de probabilidade conjunta de $\mathbf{v}(n)$ não seja conhecida, uma aproximação para o erro quadrático médio pode ser obtida empregando-se as hipóteses **H6** e **H7** e removendo-se o condicionamento em $\mathbf{v}(n)$. Desta forma, substituindo (5.10), (5.13), (5.17) e (5.23) em (5.6), e tomando-se o valor esperado, resulta em

$$\begin{aligned}
E\{e^2(n)\} &\cong \sigma_r^2 \left[1 + \text{tr} \left\{ E \left\{ \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T(n) \right\} \right\} \right] \\
&+ \text{tr} \left\{ E \left\{ \boldsymbol{\phi}(n) \boldsymbol{\phi}^T(n) \right\} \mathbf{K}(n) \right\} \\
&- 2\sigma_r^2 \text{tr} \left\{ \mathbf{Q}(-n+1) E \left\{ \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T(n) \right\} \right\} \\
&+ 2E \left\{ \mathbf{v}^T(n) \right\} E \left\{ \boldsymbol{\phi}(n) \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{U}^T(n) \right\} \mathbf{v}(0) \\
&+ \mathbf{v}^T(0) E \left\{ \left[\mathbf{U}(n) \mathbf{Q}^T(-n+1) \hat{\mathbf{a}}(n) \right]^2 \right\} \mathbf{v}(0) \\
&+ \sigma_r^2 \text{tr} \left\{ \mathbf{Q}(-n+1) E \left\{ \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T(n) \right\} \right\}
\end{aligned} \tag{5.24}$$

em que

$$\mathbf{K}(n) = E \left\{ \mathbf{v}(n) \mathbf{v}^T(n) \right\} \tag{5.25}$$

são os momentos de segunda ordem do vetor de erro dos coeficientes.

Os demais valores esperados da equação (5.24) são avaliados a seguir. O traço dos momentos de segunda ordem da estimativa dos mínimos quadrados dos coeficientes do processo AR em (5.24), $E\{\hat{\mathbf{a}}(n)\hat{\mathbf{a}}^T(n)\}$, é avaliado em Almeida (2004, p. 74), como

$$\begin{aligned}
\text{tr} \left\{ E \left\{ \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T(n) \right\} \right\} &= \text{tr} \left\{ E \left\{ \hat{\mathbf{a}}^T(n) \hat{\mathbf{a}}(n) \right\} \right\} \\
&\cong \mathbf{a}^T \mathbf{a} + \sigma_z^2 \text{tr} \left\{ E \left\{ \left[\mathbf{U}^T(n) \mathbf{U}(n) \right]^{-1} \right\} \right\}.
\end{aligned} \tag{5.26}$$

Porém, extensivas simulações realizadas neste trabalho, assim como em Bershad, Linebarger e McLaughlin (2001) e Almeida et al. (2005), tem demonstrado que a segunda parcela do lado direito de (5.26) pode ser considerada desprezível em relação a $\mathbf{a}^T \mathbf{a}$ (corroborando com a hipótese **H7**), portanto

$$\text{tr} \left\{ E \left\{ \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T(n) \right\} \right\} \approx \mathbf{a}^T \mathbf{a}. \tag{5.27}$$

No *terceiro e no sexto termos* do lado direito de (5.24), $E\{\hat{\mathbf{a}}(n)\hat{\mathbf{a}}^T(n)\}$, modificado pela matriz $\mathbf{Q}(-n+1)$, pode ser avaliado substituindo (2.52) nesses termos, de forma que

$$\begin{aligned}
& \text{tr} \left\{ \mathbf{Q}(-n+1) E \left\{ \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T(n) \right\} \right\} \\
&= \text{tr} \left\{ \mathbf{Q}(-n+1) \mathbf{a} \mathbf{a}^T \right\} \\
&+ \text{tr} \left\{ \mathbf{Q}(-n+1) E \left\{ \mathbf{z}^T(n) \mathbf{U}(n) \left[\mathbf{U}^T(n) \mathbf{U}(n) \right]^{-1} \right\} \right\} \\
&+ \text{tr} \left\{ \mathbf{Q}(-n+1) E \left\{ \left[\mathbf{U}^T(n) \mathbf{U}(n) \right]^{-1} \mathbf{U}^T(n) \mathbf{z}(n) \mathbf{a}^T \right\} \right\} \\
&+ \text{tr} \left\{ \mathbf{Q}(-n+1) E \left\{ \left[\mathbf{U}^T(n) \mathbf{U}(n) \right]^{-1} \mathbf{U}^T(n) \mathbf{z}(n) \right. \right. \\
&\quad \left. \left. \times \mathbf{z}^T(n) \mathbf{U}(n) \left[\mathbf{U}^T(n) \mathbf{U}(n) \right]^{-1} \right\} \right\}.
\end{aligned} \tag{5.28}$$

O segundo e o terceiro termos do lado direito de (5.28) são desprezados devido à hipótese **H2**, que assume a independência estatística entre $\mathbf{U}(n)$ e $\mathbf{z}(n)$, e sabendo-se que $E\{\mathbf{z}(n)\} = \mathbf{0}_{N \times 1}$. Condicionando-se a equação em $\mathbf{U}(n)$, tem-se que

$$\begin{aligned}
& \text{tr} \left\{ \mathbf{Q}(-n+1) E \left\{ \left[\mathbf{U}^T(n) \mathbf{U}(n) \right]^{-1} \mathbf{U}^T(n) \mathbf{z}(n) \mathbf{z}^T(n) \mathbf{U}(n) \left[\mathbf{U}^T(n) \mathbf{U}(n) \right]^{-1} \mathbf{U}(n) \right\} \right\} \\
&= \text{tr} \left\{ \mathbf{Q}(-n+1) E \left\{ \left[\mathbf{U}^T(n) \mathbf{U}(n) \right]^{-1} \mathbf{U}^T(n) E \left\{ \mathbf{z}(n) \mathbf{z}^T(n) \mid \mathbf{U}(n) \right\} \mathbf{U}(n) \right. \right. \\
&\quad \left. \left. \times \left[\mathbf{U}^T(n) \mathbf{U}(n) \right]^{-1} \right\} \right\}.
\end{aligned} \tag{5.29}$$

Considerando-se ainda que $\mathbf{z}(n)$ é i.i.d., portanto $E\{\mathbf{z}(n) \mathbf{z}^T(n)\} = \sigma_z^2 \mathbf{I}_N$, e aplicando o valor esperado visando retirar o condicionamento em $\mathbf{U}(n)$, chega-se a

$$\begin{aligned}
& \text{tr} \left\{ \mathbf{Q}(-n+1) E \left\{ \left[\mathbf{U}^T(n) \mathbf{U}(n) \right]^{-1} \mathbf{U}^T(n) \mathbf{z}(n) \mathbf{z}^T(n) \mathbf{U}(n) \left[\mathbf{U}^T(n) \mathbf{U}(n) \right]^{-1} \right\} \right\} \\
&= \text{tr} \left\{ \mathbf{Q}(-n+1) E \left\{ \left[\mathbf{U}^T(n) \mathbf{U}(n) \right]^{-1} \mathbf{U}^T(n) \sigma_z^2 \mathbf{I}_N \mathbf{U}(n) \left[\mathbf{U}^T(n) \mathbf{U}(n) \right]^{-1} \right\} \right\} \\
&= \sigma_z^2 \text{tr} \left\{ \mathbf{Q}(-n+1) E \left\{ \left[\mathbf{U}^T(n) \mathbf{U}(n) \right]^{-1} \right\} \right\}.
\end{aligned} \tag{5.30}$$

Dessa forma, esse termo resulta em

$$\begin{aligned} & \text{tr} \left\{ \mathbf{Q}(-n+1) E \left\{ \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T(n) \right\} \right\} \\ & \cong \mathbf{a}^T \mathbf{Q}(-n+1) \mathbf{a} \\ & + \sigma_z^2 \text{tr} \left\{ \mathbf{Q}(-n+1) E \left\{ \left[\mathbf{U}^T(n) \mathbf{U}(n) \right]^{-1} \right\} \right\} \end{aligned} \quad (5.31)$$

e assim, como em (5.27), pode ser reduzido à

$$\text{tr} \left\{ \mathbf{Q}(-n+1) E \left\{ \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T(n) \right\} \right\} \approx \mathbf{a}^T \mathbf{Q}(-n+1) \mathbf{a}. \quad (5.32)$$

O *segundo valor esperado* de (5.24) foi aproximado em (3.1)

$$\mathbf{R}_{\phi\phi}(n) \triangleq E \left\{ \phi(n) \phi^T(n) \right\} \approx \frac{N-P}{N} \sigma_z^2 \mathbf{I}_N$$

enquanto o *quarto termo* à direita de (5.24) foi definido anteriormente em (4.7), como

$$E \left\{ \phi(n) \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{U}^T(n) \right\} \triangleq \mathbf{R}_{\phi\hat{\mathbf{a}}}(n).$$

O valor esperado no *quinto termo* de (5.24) é definido pela matriz

$$E \left\{ \mathbf{U}(n) \mathbf{Q}^T(-n+1) \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{U}^T(n) \right\} \triangleq \mathbf{R}_{\hat{\mathbf{a}}}(n) \quad (5.33)$$

sendo derivado no Apêndice C por

$$r_{\hat{\mathbf{a}}(l,c)}(n) = \sum_{i=1}^P \sum_{j=1}^P r_{uu(i+l,j+c)} a_i a_j q_{-1}(-n+i) q_{-1}(-n+j) \quad (5.34)$$

em que $u_{\cdot 1}(k)$ é uma função degrau unitário (1 se $k \geq 0$ e 0 para outros valores de k), $r_{uu(a,b)} = E\{u(n-a)u(n-b)\}$ e $r_{\hat{\mathbf{a}}(l,c)}(n)$ é o elemento da l -ésima linha e c -ésima coluna de $\mathbf{R}_{\hat{\mathbf{a}}}(n)$. A matriz $\mathbf{R}_{\hat{\mathbf{a}}}(n)$ é a matriz de correlação de $\hat{\mathbf{u}}(n) = \mathbf{U}(n) \mathbf{Q}(-n+1) \hat{\mathbf{a}}(n)$.

Substituindo-se (3.1), (4.7), (5.27) e (5.32) a (5.33) em (5.24) chega-se ao modelo analítico para o comportamento do erro quadrático médio,

$$\begin{aligned} E \left\{ e^2(n) \right\} & \cong \sigma_r^2 \left(1 + \mathbf{a}^T \mathbf{a} \right) + \frac{G}{N} \sigma_z^2 k(n) \\ & + 2E \left\{ \mathbf{v}^T(n) \right\} \mathbf{R}_{\phi\hat{\mathbf{a}}}(n) \mathbf{v}(0) \\ & + \mathbf{v}^T(0) \mathbf{R}_{\hat{\mathbf{a}}}(n) \mathbf{v}(0) - \sigma_r^2 \mathbf{a}^T \mathbf{Q}(-n+1) \mathbf{a} \end{aligned} \quad (5.35)$$

em que $E\{\mathbf{v}(n)\}$ é descrito em (4.20) e $k(n) = \text{tr}\{\mathbf{K}(n)\}$ é o traço dos momentos de segunda ordem de $\mathbf{v}(n)$. Um modelo analítico para o comportamento de $k(n)$ é derivado na próxima seção.

5.2 Momentos de Segunda Ordem em Regime Transiente

Nesta seção são derivadas equações recursivas para os momentos de segunda ordem do vetor de erro dos coeficientes e para o traço dessa matriz, requerida na equação do erro quadrático médio. O desenvolvimento realizado nessa seção segue uma metodologia semelhante à apresentada em Almeida (2004), porém com a nova equação para o vetor de erro dos coeficientes, considerando a inclusão do comportamento do Efeito de Inicialização.

5.2.1 Avaliação dos Momentos de Segunda Ordem

A matriz de momentos de segunda ordem do vetor de erro dos coeficientes é definida em (5.25), como

$$\mathbf{K}(n) = E\{\mathbf{v}(n)\mathbf{v}^T(n)\}$$

sendo que o vetor de erro dos coeficientes é obtido substituindo-se (5.4) em (4.1) para $n \geq 0$, resultando em

$$\begin{aligned} \mathbf{v}(n+1) = & \mathbf{v}(n) - \frac{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \mathbf{v}(n) + \frac{\boldsymbol{\phi}(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} r_a(n) \\ & - \frac{\boldsymbol{\phi}(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\mathbf{v}(0) \\ & + \frac{\boldsymbol{\phi}(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{r}(n-1). \end{aligned} \quad (5.36)$$

Dessa forma, pré-multiplicando (5.36) por sua transposta, e tomando-se o valor esperado do resultado condicionado em $\mathbf{v}(n)$, tem-se que

$$\begin{aligned}
\mathbf{K}_v(n+1) &= \mathbf{v}(n)\mathbf{v}^T(n) - \mathbf{v}(n)\mathbf{v}^T(n)E \left\{ \frac{\phi(n)\phi^T(n)}{\phi^T(n)\phi(n)} \middle| \mathbf{v}(n) \right\} \\
&+ \mathbf{v}(n)E \left\{ \frac{\phi^T(n)r_a(n)}{\phi^T(n)\phi(n)} \middle| \mathbf{v}(n) \right\} \\
&- \mathbf{v}(n)\mathbf{v}^T(0)E \left\{ \frac{\mathbf{U}(n)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\phi^T(n)}{\phi^T(n)\phi(n)} \middle| \mathbf{v}(n) \right\} \\
&+ \mathbf{v}(n)E \left\{ \frac{\mathbf{r}^T(n-1)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\phi^T(n)}{\phi^T(n)\phi(n)} \middle| \mathbf{v}(n) \right\} \\
&- E \left\{ \frac{\phi(n)\phi^T(n)}{\phi^T(n)\phi(n)} \middle| \mathbf{v}(n) \right\} \mathbf{v}(n)\mathbf{v}^T(n) \\
&+ E \left\{ \frac{\phi(n)\phi^T(n)\mathbf{v}(n)\mathbf{v}^T(n)\phi(n)\phi^T(n)}{[\phi^T(n)\phi(n)]^2} \middle| \mathbf{v}(n) \right\} \\
&- E \left\{ r_a(n) \frac{\phi(n)\phi^T(n)\phi^T(n)\mathbf{v}(n)}{[\phi^T(n)\phi(n)]^2} \middle| \mathbf{v}(n) \right\} \\
&+ E \left\{ \frac{\phi(n)\phi^T(n)\mathbf{v}(n)\mathbf{v}^T(0)\mathbf{U}(n)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\phi^T(n)}{[\phi^T(n)\phi(n)]^2} \middle| \mathbf{v}(n) \right\} \\
&- E \left\{ \frac{\phi(n)\mathbf{r}^T(n-1)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\phi^T(n)\phi^T(n)}{[\phi^T(n)\phi(n)]^2} \middle| \mathbf{v}(n) \right\} \mathbf{v}(n) \\
&+ E \left\{ r_a(n) \frac{\phi(n)}{\phi^T(n)\phi(n)} \middle| \mathbf{v}(n) \right\} \mathbf{v}^T(n) + E \left\{ r_a^2(n) \frac{\phi(n)\phi^T(n)}{[\phi^T(n)\phi(n)]^2} \middle| \mathbf{v}(n) \right\} \\
&- E \left\{ r_a(n) \frac{\mathbf{v}^T(n)\phi(n)\phi(n)\phi^T(n)}{[\phi^T(n)\phi(n)]^2} \middle| \mathbf{v}(n) \right\} \\
&- E \left\{ r_a(n) \frac{\phi(n)\mathbf{v}^T(0)\mathbf{U}(n)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\phi^T(n)}{[\phi^T(n)\phi(n)]^2} \middle| \mathbf{v}(n) \right\} \\
&+ E \left\{ r_a(n) \frac{\phi(n)\mathbf{r}^T(n-1)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\phi^T(n)}{[\phi^T(n)\phi(n)]^2} \middle| \mathbf{v}(n) \right\} \\
&- E \left\{ \frac{\phi(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)}{\phi^T(n)\phi(n)} \middle| \mathbf{v}(n) \right\} \mathbf{v}(0)\mathbf{v}^T(n)
\end{aligned}$$

$$\begin{aligned}
& +E \left\{ \frac{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\mathbf{v}(0)\mathbf{v}^T(n)\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^2} \middle| \mathbf{v}(n) \right\} \\
& -E \left\{ r_a(n) \frac{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\mathbf{v}(0)\boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^2} \middle| \mathbf{v}(n) \right\} \\
& +E \left\{ \frac{[\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\mathbf{v}(0)]^2 \boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^2} \middle| \mathbf{v}(n) \right\} \\
& -E \left\{ \frac{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\mathbf{v}(0)\mathbf{r}^T(n-1)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^2} \middle| \mathbf{v}(n) \right\} \\
& +E \left\{ \frac{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{r}(n-1)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \middle| \mathbf{v}(n) \right\} \mathbf{v}^T(n) \\
& -\mathbf{v}^T(n)E \left\{ \frac{\boldsymbol{\phi}(n)\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{r}(n-1)\boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^2} \middle| \mathbf{v}(n) \right\} \\
& +E \left\{ r_a(n) \frac{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{r}(n-1)\boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^2} \middle| \mathbf{v}(n) \right\} \\
& -E \left\{ \frac{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{r}(n-1)\mathbf{v}^T(0)\mathbf{U}(n)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^2} \middle| \mathbf{v}(n) \right\} \\
& +E \left\{ \frac{[\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{r}(n-1)]^2 \boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^2} \middle| \mathbf{v}(n) \right\}
\end{aligned} \tag{5.37}$$

em que $\mathbf{K}_{\mathbf{v}}(n+1) = E\{\mathbf{v}(n+1)\mathbf{v}^T(n+1)|\mathbf{v}(n)\}$.

Para a avaliação do *segundo* e do *sexto termo* do lado direito de (5.37) assume-se a hipótese **H5**, resultando em

$$\begin{aligned}
& \mathbf{v}(n)\mathbf{v}^T(n)E \left\{ \frac{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \middle| \mathbf{v}(n) \right\} \\
& \cong \mathbf{v}(n)\mathbf{v}^T(n)E \left\{ [\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^{-1} \middle| \mathbf{v}(n) \right\} E \left\{ \boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n) \middle| \mathbf{v}(n) \right\}
\end{aligned} \tag{5.38}$$

e

$$\begin{aligned}
& E \left\{ \frac{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \middle| \mathbf{v}(n) \right\} \mathbf{v}(n)\mathbf{v}^T(n) \\
& \cong E \left\{ [\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^{-1} \middle| \mathbf{v}(n) \right\} E \left\{ \boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n) \middle| \mathbf{v}(n) \right\} \mathbf{v}(n)\mathbf{v}^T(n)
\end{aligned} \tag{5.39}$$

respectivamente. Através de **H5** também é possível a avaliação do *quarto* e do *décimo sexto termo* do lado direito de (5.37), levando à

$$\begin{aligned} & \mathbf{v}(n)\mathbf{v}^T(0)E\left\{\frac{\mathbf{U}(n)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\boldsymbol{\phi}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)}\middle|\mathbf{v}(n)\right\} \\ & \cong \mathbf{v}(n)\mathbf{v}^T(0)E\left\{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^{-1}\middle|\mathbf{v}(n)\right\} \\ & \quad \times E\left\{\mathbf{U}(n)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\boldsymbol{\phi}^T(n)\middle|\mathbf{v}(n)\right\} \end{aligned} \quad (5.40)$$

e

$$\begin{aligned} & E\left\{\frac{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)}\middle|\mathbf{v}(n)\right\}\mathbf{v}(0)\mathbf{v}^T(n) \\ & \cong E\left\{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^{-1}\middle|\mathbf{v}(n)\right\} \\ & \quad \times E\left\{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\middle|\mathbf{v}(n)\right\}\mathbf{v}(0)\mathbf{v}^T(n). \end{aligned} \quad (5.41)$$

Os valores esperados do *terceiro* e do *décimo primeiro termos* do lado direito da equação (5.37) podem ser desprezados considerando a independência estatística entre o processo aleatório de ruído $r(n)$ em relação ao processo aleatório de entrada $u(n)$,

$$\begin{aligned} & E\left\{\frac{\boldsymbol{\phi}^T(n)r_a(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)}\middle|\mathbf{v}(n)\right\} = E\left\{\frac{\boldsymbol{\phi}^T(n)r(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)}\middle|\mathbf{v}(n)\right\} - E\left\{\frac{\boldsymbol{\phi}^T(n)\hat{\mathbf{a}}^T(n)\mathbf{r}(n-1)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)}\middle|\mathbf{v}(n)\right\} \\ & \cong E\left\{\frac{\boldsymbol{\phi}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)}\middle|\mathbf{v}(n)\right\}\underbrace{E\{r(n)|\mathbf{v}(n)\}}_0 - E\left\{\frac{\boldsymbol{\phi}^T(n)\hat{\mathbf{a}}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)}\middle|\mathbf{v}(n)\right\}\underbrace{E\{\mathbf{r}(n-1)|\mathbf{v}(n)\}}_{\mathbf{0}_{p \times 1}} \\ & = \mathbf{0}_{N \times 1}^T. \end{aligned} \quad (5.42)$$

Assume-se, nesta análise, que $r(n)$ possui média zero, sendo independente e identicamente distribuído (i.i.d.). O *oitavo* e o *décimo terceiro* termos da equação (5.37) levam ao mesmo resultado,

$$\begin{aligned} & E\left\{r_a(n)\frac{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\boldsymbol{\phi}^T(n)\mathbf{v}(n)}{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^2}\middle|\mathbf{v}(n)\right\} \\ & = \underbrace{E\{r_a(n)|\mathbf{v}(n)\}}_0 E\left\{\frac{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\boldsymbol{\phi}^T(n)\mathbf{v}(n)}{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^2}\middle|\mathbf{v}(n)\right\} \\ & = \mathbf{0}_{N \times N}. \end{aligned} \quad (5.43)$$

Similarmente, o valor esperado no *quinto* e o *vigésimo primeiro termo* à direita de (5.37) podem ser desprezados, em virtude de

$$\begin{aligned}
& E \left\{ \frac{\mathbf{r}^T(n-1)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\boldsymbol{\phi}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \middle| \mathbf{v}(n) \right\} \\
& \cong E \left\{ \underbrace{\mathbf{r}^T(n-1)}_{\mathbf{0}_{1 \times P}} \middle| \mathbf{v}(n) \right\} E \left\{ \frac{\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\boldsymbol{\phi}^T(n)}{\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)} \middle| \mathbf{v}(n) \right\} \quad (5.44) \\
& = \mathbf{0}_{1 \times N}.
\end{aligned}$$

Considera-se, nessa avaliação, que $\mathbf{Q}(-n+1)$ é uma matriz determinística, embora variante no tempo.

Uma simplificação para o *sétimo termo* de (5.37) pode ser obtida através da hipótese **H5**, de forma que

$$\begin{aligned}
& E \left\{ \frac{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\mathbf{v}(n)\mathbf{v}^T(n)\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^2} \middle| \mathbf{v}(n) \right\} \\
& \cong E \left\{ \boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\mathbf{v}(n)\mathbf{v}^T(n)\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n) \middle| \mathbf{v}(n) \right\} E \left\{ [\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^{-2} \middle| \mathbf{v}(n) \right\} \quad (5.45)
\end{aligned}$$

em virtude do denominador ser elevado ao quadrado, tornando tal hipótese mais acurada.

De forma a avaliar o *nono* e o *décimo sétimo termo*, à direita de (5.37), inicialmente aplica-se a hipótese **H5**, o que resulta em

$$\begin{aligned}
& E \left\{ \frac{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\mathbf{v}(n)\mathbf{v}^T(0)\mathbf{U}(n)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^2} \middle| \mathbf{v}(n) \right\} \\
& \cong E \left\{ [\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^{-2} \middle| \mathbf{v}(n) \right\} E \left\{ \boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n) \right. \quad (5.46) \\
& \quad \left. \times \mathbf{v}(0)\mathbf{v}^T(n)\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n) \middle| \mathbf{v}(n) \right\} \mathbf{v}(0)\mathbf{v}^T(n).
\end{aligned}$$

O segundo valor esperado a direita de (5.46) é composto de momentos de sexta ordem, sendo dessa forma fatorado através do *teorema da fatoração dos momentos Gaussianos*, assumindo-se que as variáveis aleatórias envolvidas são conjuntamente gaussianas. A decomposição em momentos de segunda ordem para o conjunto de variáveis $E\{\mathbf{y}_1\mathbf{y}_2^T\mathbf{w}_1\mathbf{w}_2^T\mathbf{y}_3\mathbf{y}_4^T\}$ é apresentado no Anexo C como

$$\begin{aligned}
& E\{\mathbf{y}_1\mathbf{y}_2^T\mathbf{w}_1\mathbf{w}_2^T\mathbf{y}_3\mathbf{y}_4^T\} = E\{\mathbf{y}_1\mathbf{y}_2^T\}\mathbf{w}_1\mathbf{w}_2^T E\{\mathbf{y}_3\mathbf{y}_4^T\} \\
& \quad + E\{\mathbf{y}_1\mathbf{y}_3^T\}\mathbf{w}_2\mathbf{w}_1^T E\{\mathbf{y}_2\mathbf{y}_4^T\} + E\{\mathbf{y}_2\mathbf{y}_3^T\}\mathbf{w}_1^T E\{\mathbf{y}_1\mathbf{y}_4^T\}\mathbf{w}_2. \quad (5.47)
\end{aligned}$$

Salienta-se que o teorema da fatoração não é uma aproximação em se tratando de variáveis Gaussianas (HAYKIN, 1996). Definindo-se as

variáveis $\mathbf{y}_1 = \boldsymbol{\phi}(n)$, $\mathbf{y}_2^T = \hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)$, $\mathbf{w}_1 = \mathbf{v}(0)$, $\mathbf{w}_2^T = \mathbf{v}^T(n)$ e $\mathbf{y}_3^T = \mathbf{y}_4^T = \boldsymbol{\phi}^T(n)$ em (5.47) (a gaussianidade da variável \mathbf{y}_2^T é suportada pela hipótese **H7**), resulta em

$$\begin{aligned}
& E\left\{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\mathbf{v}(0)\mathbf{v}^T(n)\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\mid\mathbf{v}(n)\right\} \\
&= E\left\{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\mid\mathbf{v}(n)\right\}\mathbf{v}(0)\mathbf{v}^T(n)E\left\{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\mid\mathbf{v}(n)\right\} \\
&+ E\left\{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\mid\mathbf{v}(n)\right\}\mathbf{v}(n)\mathbf{v}^T(0)E\left\{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\mid\mathbf{v}(n)\right\} \\
&+ E\left\{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\mid\mathbf{v}(n)\right\}\text{tr}\left[E\left\{\mathbf{U}(n)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\boldsymbol{\phi}^T(n)\mid\mathbf{v}(n)\right\}\mathbf{v}(n)\mathbf{v}^T(0)\right].
\end{aligned} \tag{5.48}$$

Substituindo-se (5.48) em (5.46), chega-se à expressão para o termo:

$$\begin{aligned}
& E\left\{\frac{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\mathbf{v}(n)\mathbf{v}^T(0)\mathbf{U}(n)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^2}\mid\mathbf{v}(n)\right\} \\
&\cong E\left\{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^{-2}\mid\mathbf{v}(n)\right\}\times\left\{E\left\{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\mid\mathbf{v}(n)\right\}\right. \\
&\quad \times\mathbf{v}(0)\mathbf{v}^T(n)E\left\{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\mid\mathbf{v}(n)\right\} \\
&\quad +E\left\{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\mid\mathbf{v}(n)\right\}\mathbf{v}(n)\mathbf{v}^T(0)E\left\{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\mid\mathbf{v}(n)\right\} \\
&\quad \left.+E\left\{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\mid\mathbf{v}(n)\right\}\text{tr}\left[E\left\{\mathbf{U}(n)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\boldsymbol{\phi}^T(n)\mid\mathbf{v}(n)\right\}\right.\right. \\
&\quad \left.\left.\times\mathbf{v}(n)\mathbf{v}^T(0)\right]\right\}.
\end{aligned} \tag{5.49}$$

Os valores esperados dos *termos dez e vinte e dois* à direita de (5.37) são obtidos de forma similar à (5.44), ou seja,

$$\begin{aligned}
& E\left\{\frac{\boldsymbol{\phi}(n)\mathbf{r}^T(n-1)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\boldsymbol{\phi}^T(n)\boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^2}\mid\mathbf{v}(n)\right\} \\
&= E\left\{\frac{\mathbf{r}^T(n-1)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^2}\mid\mathbf{v}(n)\right\} \\
&\cong \underbrace{E\left\{\mathbf{r}^T(n-1)\mid\mathbf{v}(n)\right\}}_{\mathbf{0}_{1 \times P}} E\left\{\frac{\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^2}\mid\mathbf{v}(n)\right\} \\
&= \mathbf{0}_N
\end{aligned} \tag{5.50}$$

sendo considerado que o ruído de medida é i.i.d e média zero, estatisticamente independente do sinal de excitação $u(n)$ e que a matriz $\mathbf{Q}_{(-n+1)}$ é determinística.

Aplicando-se a hipótese **H5** no *décimo segundo termo* à direita da igualdade em (5.37), tem-se que ele resulta em

$$\begin{aligned} E \left\{ r_a^2(n) \frac{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^2} \mid \mathbf{v}(n) \right\} \\ \cong E \left\{ [\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^{-2} \mid \mathbf{v}(n) \right\} E \left\{ r_a^2(n)\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n) \mid \mathbf{v}(n) \right\}. \end{aligned} \quad (5.51)$$

Esses momentos de quarta ordem podem ser fatorados através do *teorema da fatoração dos momentos Gaussianos*. A decomposição em momentos de segunda ordem para o conjunto de variáveis $E\{y_1\mathbf{y}_2\mathbf{y}_3^T y_4\}$ é obtido no Anexo B,

$$\begin{aligned} E\{y_1\mathbf{y}_2\mathbf{y}_3^T y_4\} = E\{y_1\mathbf{y}_2\} E\{\mathbf{y}_3^T y_4\} + E\{\mathbf{y}_2 y_4\} E\{y_1\mathbf{y}_3^T\} \\ + E\{y_1 y_4\} E\{\mathbf{y}_2\mathbf{y}_3^T\}. \end{aligned} \quad (5.52)$$

Substituindo-se as variáveis $y_1 = y_4 = r_a(n)$, $\mathbf{y}_2 = \mathbf{y}_3 = \boldsymbol{\phi}(n)$, em que as variáveis y_1 e y_4 são consideradas gaussianas devido à hipótese **H7**, tem-se que

$$\begin{aligned} E \left\{ r_a^2(n)\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n) \mid \mathbf{v}(n) \right\} \\ = \underbrace{E \left\{ r_a(n)\boldsymbol{\phi}(n) \mid \mathbf{v}(n) \right\} E \left\{ \boldsymbol{\phi}^T(n)r_a(n) \mid \mathbf{v}(n) \right\}}_{\mathbf{0}_{N \times N}} \\ + \underbrace{E \left\{ \boldsymbol{\phi}(n)r_a(n) \mid \mathbf{v}(n) \right\} E \left\{ \boldsymbol{\phi}^T(n)r_a(n) \mid \mathbf{v}(n) \right\}}_{\mathbf{0}_{N \times N}} \\ + E \left\{ r_a^2(n) \mid \mathbf{v}(n) \right\} E \left\{ \boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n) \mid \mathbf{v}(n) \right\} \end{aligned} \quad (5.53)$$

sendo que os dois primeiros valores esperados de (5.53) são nulos devido à característica de média zero do ruído de medida $r(n)$ e sua independência em relação ao vetor $\boldsymbol{\phi}(n)$. Dessa forma, o primeiro valor esperado não nulo à direita de (5.53) pode ser aproximado por

$$E \left\{ r_a^2(n)\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n) \mid \mathbf{v}(n) \right\} \cong E \left\{ r_a^2(n) \mid \mathbf{v}(n) \right\} E \left\{ \boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n) \mid \mathbf{v}(n) \right\} \quad (5.54)$$

sendo que o primeiro valor esperado do lado direito de (5.54) já foi avaliado na equação (5.10) e, portanto, a expressão resulta em

$$\begin{aligned}
& E \left\{ r_a^2(n) \boldsymbol{\phi}(n) \boldsymbol{\phi}^T(n) \middle| \mathbf{v}(n) \right\} \\
& \cong E \left\{ \boldsymbol{\phi}(n) \boldsymbol{\phi}^T(n) \mathbf{v}(n) \right\} \left[1 + \text{tr} \left\{ E \left\{ \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T(n) \middle| \mathbf{v}(n) \right\} \right\} \right] \sigma_r^2
\end{aligned} \tag{5.55}$$

e substituindo (5.55) em (5.51), chega-se à

$$\begin{aligned}
& E \left\{ r_a^2(n) \frac{\boldsymbol{\phi}(n) \boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n) \boldsymbol{\phi}(n)]^2} \middle| \mathbf{v}(n) \right\} \\
& \cong E \left\{ [\boldsymbol{\phi}^T(n) \boldsymbol{\phi}(n)]^{-2} \middle| \mathbf{v}(n) \right\} E \left\{ \boldsymbol{\phi}(n) \boldsymbol{\phi}^T(n) \middle| \mathbf{v}(n) \right\} \\
& \quad \times \left[1 + \text{tr} \left\{ E \left\{ \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T(n) \middle| \mathbf{v}(n) \right\} \right\} \right] \sigma_r^2.
\end{aligned} \tag{5.56}$$

A equação (5.4), $r_a(n) = r(n) \hat{\mathbf{a}}^T(n) \mathbf{r}(n-1)$, pode ser aplicada ao *décimo quarto* e ao *décimo oitavo termo* de (5.37), levando à

$$\begin{aligned}
& E \left\{ r_a(n) \frac{\boldsymbol{\phi}(n) \mathbf{v}^T(0) \mathbf{U}(n) \mathbf{Q}^T(-n+1) \hat{\mathbf{a}}(n) \boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n) \boldsymbol{\phi}(n)]^2} \middle| \mathbf{v}(n) \right\} \\
& = E \left\{ r(n) \frac{\boldsymbol{\phi}(n) \mathbf{v}^T(0) \mathbf{U}(n) \mathbf{Q}^T(-n+1) \hat{\mathbf{a}}(n) \boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n) \boldsymbol{\phi}(n)]^2} \middle| \mathbf{v}(n) \right\} \\
& \quad - E \left\{ \mathbf{r}^T(n-1) \hat{\mathbf{a}}(n) \frac{\boldsymbol{\phi}(n) \mathbf{v}^T(0) \mathbf{U}(n) \mathbf{Q}^T(-n+1) \hat{\mathbf{a}}(n) \boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n) \boldsymbol{\phi}(n)]^2} \middle| \mathbf{v}(n) \right\} \\
& \cong \underbrace{E \{ r(n) \}}_0 E \left\{ \frac{\boldsymbol{\phi}(n) \mathbf{v}^T(0) \mathbf{U}(n) \mathbf{Q}^T(-n+1) \hat{\mathbf{a}}(n) \boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n) \boldsymbol{\phi}(n)]^2} \middle| \mathbf{v}(n) \right\} \\
& \quad - \underbrace{E \{ \mathbf{r}^T(n-1) \}}_{\mathbf{0}_{N \times P}} E \left\{ \hat{\mathbf{a}}(n) \frac{\boldsymbol{\phi}(n) \mathbf{v}^T(0) \mathbf{U}(n) \mathbf{Q}^T(-n+1) \hat{\mathbf{a}}(n) \boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n) \boldsymbol{\phi}(n)]^2} \middle| \mathbf{v}(n) \right\} \\
& = \mathbf{0}_{N \times P}.
\end{aligned} \tag{5.57}$$

O *décimo quinto* e o *vigésimo terceiro termo* à direita da igualdade de (5.37) são avaliados assumindo-se uma hipótese similar à **H5** entre o numerador e o denominador, e utilizando-se (5.4), obtém-se

$$\begin{aligned}
& E \left\{ r_a(n) \frac{\boldsymbol{\phi}(n) \mathbf{r}^T(n-1) \mathbf{Q}^T(-n+1) \hat{\mathbf{a}}(n) \boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n) \boldsymbol{\phi}(n)]^2} \middle| \mathbf{v}(n) \right\} \\
& \cong E \left\{ [\boldsymbol{\phi}^T(n) \boldsymbol{\phi}(n)]^{-2} \middle| \mathbf{v}(n) \right\} \underbrace{E \{ r(n) \}}_0 E \left\{ \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{r}(n-1) \right. \\
& \quad \left. \times \boldsymbol{\phi}(n) \boldsymbol{\phi}^T(n) \middle| \mathbf{v}(n) \right\} \\
& - E \left\{ [\boldsymbol{\phi}^T(n) \boldsymbol{\phi}(n)]^{-2} \middle| \mathbf{v}(n) \right\} E \left\{ \hat{\mathbf{a}}^T(n) \mathbf{r}(n-1) \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{r}(n-1) \right. \\
& \quad \left. \times \boldsymbol{\phi}(n) \boldsymbol{\phi}^T(n) \middle| \mathbf{v}(n) \right\}.
\end{aligned} \tag{5.58}$$

O primeiro valor esperado à direita de (5.58) é nulo em virtude do ruído de medida apresentar média zero, enquanto que o segundo termo é avaliado através do teorema da fatoração dos momentos Gaussianos de (5.52), no Anexo B. Dessa forma, realizando a seguinte equivalência de variáveis, tendo a sua gaussianidade suportada pela hipótese **H7**, $y_1 = \hat{\mathbf{a}}^T(n) \mathbf{r}(n-1)$, $y_4 = \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{r}(n-1)$, $\mathbf{y}_2 = \boldsymbol{\phi}(n)$, e $\mathbf{y}_3^T = \boldsymbol{\phi}^T(n)$, obtém-se que

$$\begin{aligned}
& E \left\{ \left[\hat{\mathbf{a}}^T(n) \mathbf{r}(n-1) \right] \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{r}(n-1) \boldsymbol{\phi}(n) \boldsymbol{\phi}^T(n) \middle| \mathbf{v}(n) \right\} \\
& = \underbrace{E \left\{ \hat{\mathbf{a}}^T(n) \mathbf{r}(n-1) \boldsymbol{\phi}(n) \middle| \mathbf{v}(n) \right\} E \left\{ \boldsymbol{\phi}^T(n) \left[\hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{r}(n-1) \right] \middle| \mathbf{v}(n) \right\}}_{\mathbf{0}_N} \\
& \quad + \underbrace{E \left\{ \boldsymbol{\phi}(n) \hat{\mathbf{a}}^T(n) \mathbf{r}(n-1) \middle| \mathbf{v}(n) \right\} E \left\{ \left[\hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{r}(n-1) \right] \boldsymbol{\phi}^T(n) \middle| \mathbf{v}(n) \right\}}_{\mathbf{0}_N} \\
& \quad + E \left\{ \left[\hat{\mathbf{a}}^T(n) \mathbf{r}(n-1) \right] \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{r}(n-1) \middle| \mathbf{v}(n) \right\} E \left\{ \boldsymbol{\phi}(n) \boldsymbol{\phi}^T(n) \middle| \mathbf{v}(n) \right\}
\end{aligned} \tag{5.59}$$

sendo que os dois primeiros valores esperados do lado direito de (5.59) são desprezados devido às características do processo aleatório $r(n)$. O primeiro valor esperado do último termo do lado direito de (5.59) já foi aproximado em (5.17) como

$$\begin{aligned}
& E \left\{ \hat{\mathbf{a}}^T(n) \mathbf{r}(n-1) \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{r}(n-1) \boldsymbol{\phi}(n) \boldsymbol{\phi}^T(n) \middle| \mathbf{v}(n) \right\} \\
& \cong \sigma_r^2 \text{tr} \left\{ \mathbf{Q}^T(-n+1) E \left\{ \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T(n) \middle| \mathbf{v}(n) \right\} \right\} E \left\{ \boldsymbol{\phi}(n) \boldsymbol{\phi}^T(n) \middle| \mathbf{v}(n) \right\}.
\end{aligned} \tag{5.60}$$

Substituindo-se (5.60) em (5.58), obtém-se

$$\begin{aligned}
 & E \left\{ r_a(n) \frac{\boldsymbol{\phi}(n) \mathbf{r}^T(n-1) \mathbf{Q}^T(-n+1) \hat{\mathbf{a}}(n) \boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n) \boldsymbol{\phi}(n)]^2} \middle| \mathbf{v}(n) \right\} \\
 & \cong -E \left\{ [\boldsymbol{\phi}^T(n) \boldsymbol{\phi}(n)]^{-2} \middle| \mathbf{v}(n) \right\} E \left\{ \boldsymbol{\phi}(n) \boldsymbol{\phi}^T(n) \middle| \mathbf{v}(n) \right\} \\
 & \quad \times \sigma_r^2 \text{tr} \left\{ \mathbf{Q}^T(-n+1) E \left\{ \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T(n) \middle| \mathbf{v}(n) \right\} \right\}.
 \end{aligned} \tag{5.61}$$

O *décimo nono termo* do lado direito de (5.37) é derivado através de hipótese similar à **H5**, considerando que o termo no denominador é elevado ao quadrado, ou seja,

$$\begin{aligned}
 & E \left\{ \frac{[\hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{U}^T(n) \mathbf{v}(0)]^2 \boldsymbol{\phi}(n) \boldsymbol{\phi}^T(n)}{[\boldsymbol{\phi}^T(n) \boldsymbol{\phi}(n)]^2} \middle| \mathbf{v}(n) \right\} \\
 & \cong E \left\{ [\boldsymbol{\phi}^T(n) \boldsymbol{\phi}(n)]^{-2} \middle| \mathbf{v}(n) \right\} E \left\{ [\hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{U}^T(n) \right. \\
 & \quad \left. \times \mathbf{v}(0)]^2 \boldsymbol{\phi}(n) \boldsymbol{\phi}^T(n) \middle| \mathbf{v}(n) \right\}.
 \end{aligned} \tag{5.62}$$

O segundo valor esperado à direita de (5.62) é avaliado através do teorema da fatoração dos momentos Gaussianos de (5.52), assumindo-se as variáveis $\mathbf{y}_2 = \boldsymbol{\phi}(n)$, $\mathbf{y}_3^T = \boldsymbol{\phi}^T(n)$ e $y_1 = y_4 = \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{U}^T(n) \mathbf{v}(0)$, e que $\hat{\mathbf{a}}(n) \approx \mathbf{a}$, tendo-se a hipótese **H7**. Assim, é possível decompô-las em correlações de segunda ordem entre os pares de variáveis,

$$\begin{aligned}
 & E \left\{ [\hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{U}^T(n) \mathbf{v}(0)]^2 \boldsymbol{\phi}(n) \boldsymbol{\phi}^T(n) \middle| \mathbf{v}(n) \right\} \\
 & = 2E \left\{ [\hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{U}^T(n) \mathbf{v}(0)] \boldsymbol{\phi}(n) \middle| \mathbf{v}(n) \right\} \\
 & \quad \times E \left\{ \boldsymbol{\phi}^T(n) [\hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{U}^T(n) \mathbf{v}(0)] \middle| \mathbf{v}(n) \right\} \\
 & + E \left\{ [\hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{U}^T(n) \mathbf{v}(0)]^2 \middle| \mathbf{v}(n) \right\} E \left\{ \boldsymbol{\phi}(n) \boldsymbol{\phi}^T(n) \middle| \mathbf{v}(n) \right\}.
 \end{aligned} \tag{5.63}$$

Dessa forma, substituindo-se (5.63) em (5.62), tem-se que

$$\begin{aligned}
& E \left\{ \frac{[\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\mathbf{v}(0)]^2 \phi(n)\phi^T(n)}{[\phi^T(n)\phi(n)]^2} \middle| \mathbf{v}(n) \right\} \\
& \cong E \left\{ [\phi^T(n)\phi(n)]^{-2} \middle| \mathbf{v}(n) \right\} \\
& \quad \times \left[2E \left\{ \phi(n) [\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\mathbf{v}(0)] \middle| \mathbf{v}(n) \right\} \right. \\
& \quad \quad \times E \left\{ [\mathbf{v}^T(0)\mathbf{U}(n)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)] \phi^T(n) \middle| \mathbf{v}(n) \right\} \\
& \quad \quad \left. + E \left\{ [\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\mathbf{v}(0)]^2 \middle| \mathbf{v}(n) \right\} \right. \\
& \quad \quad \left. \times E \left\{ \phi(n)\phi^T(n) \middle| \mathbf{v}(n) \right\} \right]. \tag{5.64}
\end{aligned}$$

Os valores esperados do *vigésimo e vigésimo quarto termos* de (5.37) podem ser eliminados levando em consideração as características do ruído de medida e a independência estatística com o vetor de excitação $\mathbf{u}(n)$,

$$\begin{aligned}
& E \left\{ \frac{\phi(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\mathbf{v}(0)\mathbf{r}^T(n-1)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\phi^T(n)}{[\phi^T(n)\phi(n)]^2} \middle| \mathbf{v}(n) \right\} \\
& = E \left\{ \frac{\mathbf{r}^T(n-1)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\mathbf{v}(0)\phi(n)\phi^T(n)}{[\phi^T(n)\phi(n)]^2} \middle| \mathbf{v}(n) \right\} \\
& \cong E \left\{ \underbrace{\mathbf{r}^T(n-1)}_{\mathbf{0}_{1 \times P}} \middle| \mathbf{v}(n) \right\} \\
& \quad \times E \left\{ \frac{\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\mathbf{v}(0)\phi(n)\phi^T(n)}{[\phi^T(n)\phi(n)]^2} \middle| \mathbf{v}(n) \right\} \\
& = \mathbf{0}_N. \tag{5.65}
\end{aligned}$$

E finalmente, aplicando-se **H5** no *vigésimo quinto termo* à direita de (5.37), tem-se que

$$\begin{aligned}
& E \left\{ \frac{[\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{r}(n-1)]^2 \phi(n)\phi^T(n)}{[\phi^T(n)\phi(n)]^2} \middle| \mathbf{v}(n) \right\} \\
& \cong E \left\{ [\phi^T(n)\phi(n)]^{-2} \middle| \mathbf{v}(n) \right\} E \left\{ [\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{r}(n-1)]^2 \phi(n)\phi^T(n) \middle| \mathbf{v}(n) \right\}. \tag{5.66}
\end{aligned}$$

em que segundo valor esperado à direita de (5.66) considera os momentos de quarta ordem, podendo ser fatorado em pares de momentos de segunda ordem através do procedimento descrito no Anexo B, e considerando as variáveis Gaussianas (hipótese **H7**) $y_1 = y_4 = \hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{r}(n-1)$, $\mathbf{y}_2 = \boldsymbol{\phi}(n)$ e $\mathbf{y}_3^T = \boldsymbol{\phi}^T(n)$, o que resulta em

$$\begin{aligned}
& E \left\{ \left[\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{r}(n-1) \right]^2 \boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n) \middle| \mathbf{v}(n) \right\} \\
&= E \left\{ \underbrace{\mathbf{r}^T(n-1)}_{\mathbf{0}_{1 \times P}} \middle| \mathbf{v}(n) \right\} E \left\{ \mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\boldsymbol{\phi}(n) \middle| \mathbf{v}(n) \right\} E \left\{ \boldsymbol{\phi}^T(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{r}(n-1) \middle| \mathbf{v}(n) \right\} \\
&\quad + E \left\{ \underbrace{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{r}(n-1)}_{\mathbf{0}_N} \middle| \mathbf{v}(n) \right\} E \left\{ \hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{r}(n-1)\boldsymbol{\phi}^T(n) \middle| \mathbf{v}(n) \right\} \\
&\quad + E \left\{ \left[\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{r}(n-1) \right]^2 \middle| \mathbf{v}(n) \right\} E \left\{ \boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n) \middle| \mathbf{v}(n) \right\}
\end{aligned} \tag{5.67}$$

sendo que o primeiro valor esperado do último termo do lado direito de (5.67) já foi avaliado em (5.23), e desta forma, (5.66) resulta em

$$\begin{aligned}
& E \left\{ \frac{\left[\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{r}(n-1) \right]^2 \boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)}{\left[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n) \right]^2} \middle| \mathbf{v}(n) \right\} \\
&\cong E \left\{ \left[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n) \right]^{-2} \middle| \mathbf{v}(n) \right\} E \left\{ \boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n) \middle| \mathbf{v}(n) \right\} \\
&\quad \times \sigma_r^2 \text{tr} \left\{ \mathbf{Q}(-n+1) E \left\{ \hat{\mathbf{a}}(n)\hat{\mathbf{a}}^T(n) \middle| \mathbf{v}(n) \right\} \right\}.
\end{aligned} \tag{5.68}$$

Embora a função densidade de probabilidade conjunta de $\mathbf{v}(n)$ não é conhecida, uma aproximação para $\mathbf{K}(n+1)$ pode ser obtida considerando-se que $\mathbf{v}(n)$ e $\boldsymbol{\phi}(n)$ são estatisticamente independentes (hipótese **H3**), e empregando as hipóteses **H6** e **H7**. Substituindo-se os resultados obtidos em (5.38) à (5.45), (5.49), (5.50), (5.56), (5.57), (5.61), (5.64), (5.65) e (5.68) em (5.37), e tomando-se o valor esperado visando retirar o condicionamento em $\mathbf{v}(n)$. Esse procedimento resulta na expressão

$$\mathbf{K}(n+1) \cong \mathbf{K}(n)$$

$$\begin{aligned}
& - E\{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^{-1}\}\mathbf{K}(n)E\{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\} \\
& - E\{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^{-1}\}E\{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\}\mathbf{K}(n) \\
& + E\{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^{-2}\}E\{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\mathbf{v}(n)\mathbf{v}^T(n)\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\} \\
& + E\{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^{-2}\}\sigma_r^2\text{tr}\{1 + E\{\hat{\mathbf{a}}(n)\hat{\mathbf{a}}^T(n)\}\}E\{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\} \\
& - E\{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^{-1}\}E\{\mathbf{v}(n)\}\mathbf{v}^T(0)E\{\mathbf{U}(n)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\boldsymbol{\phi}^T(n)\} \\
& - E\{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^{-1}\}E\{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\}\mathbf{v}(0)E\{\mathbf{v}^T(n)\} \\
& + 2E\{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^{-2}\}E\{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\} \\
& \quad \times \mathbf{v}(0)E\{\mathbf{v}^T(n)\}E\{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\} \\
& + 2E\{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^{-2}\}E\{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\}E\{\mathbf{v}(n)\}\mathbf{v}^T(0) \\
& \quad \times E\{\mathbf{U}(n)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\boldsymbol{\phi}^T(n)\} \\
& + 2E\{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^{-2}\}E\{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\} \\
& \quad \times \text{tr}\{E\{\mathbf{U}(n)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\boldsymbol{\phi}^T(n)\}E\{\mathbf{v}(n)\}\mathbf{v}^T(0)\} \\
& + E\{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^{-2}\}E\{[\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\mathbf{v}(0)]^2\}E\{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\} \\
& + 2E\{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^{-2}\}E\{\boldsymbol{\phi}(n)\hat{\mathbf{a}}^T(n)\mathbf{Q}(-n+1)\mathbf{U}^T(n)\}\mathbf{v}(0)\mathbf{v}^T(0) \\
& \quad \times E\{\mathbf{U}(n)\mathbf{Q}^T(-n+1)\hat{\mathbf{a}}(n)\boldsymbol{\phi}^T(n)\} \\
& - 2E\{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^{-2}\}\sigma_r^2\text{tr}\{\mathbf{Q}^T(-n+1)E\{\hat{\mathbf{a}}(n)\hat{\mathbf{a}}^T(n)\}\}E\{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\} \\
& + E\{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^{-2}\}\sigma_r^2\text{tr}\{\mathbf{Q}^T(-n+1)E\{\hat{\mathbf{a}}(n)\hat{\mathbf{a}}^T(n)\}\}E\{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\}
\end{aligned} \tag{5.69}$$

sendo $\mathbf{K}(n+i) = E\{\mathbf{v}(n+i)\mathbf{v}^T(n+i)\}$ para $i = \{0,1\}$.

A avaliação do *segundo valor esperado* do *quarto termo* do lado direito de (5.69) é obtida através da equação (3.85) de Almeida (2004, p. 76),

$$\begin{aligned}
& E\{\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\mathbf{v}(n)\mathbf{v}^T(n)\boldsymbol{\phi}(n)\boldsymbol{\phi}^T(n)\} \\
& \cong \frac{G}{N}\sigma_z^2 \left[\frac{G}{N}\text{tr}\{\mathbf{K}(n)\} + \left(1 - \frac{G}{N}\right)E\{\mathbf{v}^T(n)\}E\{\mathbf{v}(n)\} \right] R_{\phi\phi}. \tag{5.70}
\end{aligned}$$

Detalhes sobre o procedimento empregado nessa aproximação podem ser obtidos diretamente na referência.

O primeiro valor esperado deste mesmo termo, assim como nos demais termos da equação (5.69), é expresso através da equação (3.73) de Almeida (2004, p. 77),

$$E\{[\boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)]^2\}^{-1} \cong \left[\left(\frac{G}{N} \right)^2 \sigma_z^4 (G^2 + 2G) \right]^{-1}. \quad (5.71)$$

Dessa forma, substituindo-se as aproximações obtidas em (3.1), (4.7), (4.9), (5.27), (5.32), (5.33), (5.70) e (5.71) em (5.69), resulta na equação recursiva:

$$\begin{aligned} \mathbf{K}(n+1) \cong & \mathbf{K}(n) - \gamma_1 \mathbf{K}(n) + \gamma_2 \text{tr}[\mathbf{K}(n)] \mathbf{I}_N \\ & + \gamma_3 E\{\mathbf{v}^T(n)\} E\{\mathbf{v}(n)\} \mathbf{I}_N + \gamma_4 \mathbf{I}_N \\ & + \gamma_5 E\{\mathbf{v}(n)\} \mathbf{v}^T(0) \mathbf{R}_{\phi\bar{u}}^T(n) \\ & + \gamma_5 \mathbf{R}_{\phi\bar{u}}(n) \mathbf{v}(0) E\{\mathbf{v}^T(n)\} \\ & + 2\gamma_6 \mathbf{v}^T(0) \mathbf{R}_{\phi\bar{u}}^T(n) E\{\mathbf{v}(n)\} \mathbf{I}_N \\ & + \gamma_6 \mathbf{v}^T(0) \mathbf{R}_{\bar{u}} \mathbf{v}(0) \mathbf{I}_N \\ & + \gamma_7 \mathbf{R}_{\phi\bar{u}}(n) \mathbf{v}(0) \mathbf{v}^T(0) \mathbf{R}_{\phi\bar{u}}^T(n) \\ & - \gamma_6 \left[\sigma_r^2 \mathbf{a}^T \mathbf{Q}(-n+1) \mathbf{a} \right] \mathbf{I}_N \end{aligned} \quad (5.72)$$

em que

$$\left\{ \begin{aligned} \gamma_1 &= \frac{2}{(G-2)}; \gamma_2 = \frac{G}{N(G^2 + 2G)}; \gamma_3 = \frac{P}{N(G^2 + 2G)} \\ \gamma_4 &= \frac{\sigma_r^2 N(1 + \mathbf{a}^T \mathbf{a})}{\sigma_z^2 G(G^2 + 2G)}; \gamma_5 = \frac{2N}{\sigma_z^2 G(G^2 + 2G)} - \frac{N}{\sigma_z^2 G(G-2)} \\ \gamma_6 &= \frac{N}{\sigma_z^2 G(G^2 + 2G)}; \gamma_7 = \frac{2N^2}{\sigma_z^4 G^2 (G^2 + 2G)}. \end{aligned} \right. \quad (5.73)$$

Como citado anteriormente, equações determinísticas para $\mathbf{R}_{\phi\bar{u}}(n)$ e $\mathbf{R}_{\bar{u}}(n)$ são derivadas no Apêndice B e Apêndice C, respectivamente.

5.2.2 Traço da Equação Recursiva dos Momentos de Segunda Ordem

A equação para o erro quadrático médio em (5.35) requer a expressão do traço de $\mathbf{K}(n)$. Dessa forma, aplicando o operador traço em ambos os lados de (5.72), obtém-se

$$\begin{aligned}
 k(n+1) &\cong k(n) - \gamma_1 k(n) + \gamma_2 \operatorname{tr} \{ k(n) \mathbf{I}_N \} \\
 &+ \gamma_3 \operatorname{tr} \left\{ E \{ \mathbf{v}^T(n) \} E \{ \mathbf{v}(n) \} \mathbf{I}_N \right\} \\
 &+ \gamma_4 \operatorname{tr} \{ \mathbf{I}_N \} + \gamma_5 \operatorname{tr} \left\{ E \{ \mathbf{v}(n) \} \mathbf{v}^T(0) \mathbf{R}_{\phi_{\bar{u}}}^T(n) \right\} \\
 &+ \gamma_5 \operatorname{tr} \left\{ \mathbf{R}_{\phi_{\bar{u}}}(n) \mathbf{v}(0) E \{ \mathbf{v}^T(n) \} \right\} \\
 &+ 2\gamma_6 \operatorname{tr} \left\{ \mathbf{v}^T(0) \mathbf{R}_{\phi_{\bar{u}}}^T(n) E \{ \mathbf{v}(n) \} \mathbf{I}_N \right\} \\
 &+ \gamma_6 \operatorname{tr} \left\{ \mathbf{v}^T(0) \mathbf{R}_{\bar{u}} \mathbf{v}(0) \mathbf{I}_N \right\} \\
 &+ \gamma_7 \operatorname{tr} \left\{ \mathbf{R}_{\phi_{\bar{u}}}(n) \mathbf{v}(0) \mathbf{v}^T(0) \mathbf{R}_{\phi_{\bar{u}}}^T(n) \right\} \\
 &- \gamma_6 \operatorname{tr} \left\{ \left[\sigma_r^2 \mathbf{a}^T \mathbf{Q}(-n+1) \mathbf{a} \right] \mathbf{I}_N \right\}
 \end{aligned} \tag{5.74}$$

em que $k(n+i) = \operatorname{tr} \{ \mathbf{K}(n+i) \}$, para $i = \{0,1\}$. A equação (5.74) pode ser avaliada utilizando-se as propriedades algébricas do operador traço.

Dessa forma, o *terceiro termo* do lado direito da equação (5.74) resulta em

$$\operatorname{tr} \{ \operatorname{tr} \{ \mathbf{K}(n) \} \mathbf{I}_N \} = k(n)N \tag{5.75}$$

enquanto, o *quarto termo* de (5.74) é expresso por

$$\operatorname{tr} \left\{ E \{ \mathbf{v}^T(n) \} E \{ \mathbf{v}(n) \} \mathbf{I}_N \right\} = E \{ \mathbf{v}^T(n) \} E \{ \mathbf{v}(n) \} N. \tag{5.76}$$

O *sexto e sétimo termos* de (5.74) são definidos por

$$\operatorname{tr} \left\{ E \{ \mathbf{v}(n) \} \mathbf{v}^T(0) \mathbf{R}_{\phi_{\bar{u}}}^T(n) \right\} = \mathbf{v}^T(0) \mathbf{R}_{\phi_{\bar{u}}}^T(n) E \{ \mathbf{v}(n) \} \tag{5.77}$$

e de forma similar, o *oitavo termo* de (5.74) resulta em

$$\operatorname{tr} \left\{ \mathbf{v}^T(0) \mathbf{R}_{\phi_{\bar{u}}}^T(n) E \{ \mathbf{v}(n) \} \mathbf{I}_N \right\} = \mathbf{v}^T(0) \mathbf{R}_{\phi_{\bar{u}}}^T(n) E \{ \mathbf{v}(n) \} N. \tag{5.78}$$

Considerando-se que $\mathbf{R}_{\phi_{\bar{u}}}(n) \mathbf{v}(0)$ é um vetor de dimensões $N \times 1$, o *décimo termo* de (5.74) pode ser calculado como

$$\begin{aligned}
 &\operatorname{tr} \left\{ \mathbf{R}_{\phi_{\bar{u}}}(n) \mathbf{v}(0) \mathbf{v}^T(0) \mathbf{R}_{\phi_{\bar{u}}}^T(n) \right\} \\
 &= \mathbf{v}^T(0) \mathbf{R}_{\phi_{\bar{u}}}^T(n) \mathbf{R}_{\phi_{\bar{u}}}(n) \mathbf{v}(0).
 \end{aligned} \tag{5.79}$$

Finalmente, substituindo-se (5.75) à (5.79) em (5.74) chega-se a equação recursiva para $k(n)$:

$$\begin{aligned}
k(n+1) &\cong (1 - \gamma_1 + \gamma_2 N) k(n) \\
&\quad + \gamma_3 N E \left\{ \mathbf{v}^T(n) \right\} E \left\{ \mathbf{v}(n) \right\} + \gamma_4 N \\
&\quad + 2(\gamma_5 + \gamma_6 N) \mathbf{v}^T(0) \mathbf{R}_{\phi_{\bar{u}}}^T(n) E \left\{ \mathbf{v}(n) \right\} \\
&\quad + \gamma_6 N \mathbf{v}^T(0) \mathbf{R}_{\bar{u}}(n) \mathbf{v}(0) \\
&\quad + \gamma_7 \mathbf{v}^T(0) \mathbf{R}_{\phi_{\bar{u}}}^T(n) \mathbf{R}_{\phi_{\bar{u}}}(n) \mathbf{v}(0) \\
&\quad - \gamma_6 N \sigma_r^2 [\mathbf{a}^T \mathbf{Q}(-n+1) \mathbf{a}].
\end{aligned} \tag{5.80}$$

5.2.3 Forma Fechada para $k(n)$

A equação (5.80) pode ser obtida na forma fechada através de substituição iterativa. O procedimento a seguir visa obter tal expressão, sendo apresentado o método nas duas primeiras iterações e generalizando posteriormente.

$n = 0$: No instante inicial, tem-se que

$$\begin{aligned}
k(1) &\cong (1 - \gamma_1 + \gamma_2 N) k(0) + \gamma_3 N E \left\{ \mathbf{v}^T(0) \right\} E \left\{ \mathbf{v}(0) \right\} \\
&\quad + \gamma_4 N + 2(\gamma_5 + \gamma_6 N) \mathbf{v}^T(0) \mathbf{R}_{\phi_{\bar{u}}}^T(0) E \left\{ \mathbf{v}(0) \right\} \\
&\quad + \gamma_6 N \mathbf{v}^T(0) \mathbf{R}_{\bar{u}}(0) \mathbf{v}(0) + \gamma_7 \mathbf{v}^T(0) \mathbf{R}_{\phi_{\bar{u}}}^T(0) \mathbf{R}_{\phi_{\bar{u}}}(0) \mathbf{v}(0) \\
&\quad - \gamma_6 N \sigma_r^2 [\mathbf{a}^T \mathbf{Q}(1) \mathbf{a}].
\end{aligned} \tag{5.81}$$

$n = 1$: Na segunda iteração, (5.80) resulta em

$$\begin{aligned}
k(2) &\cong (1 - \gamma_1 + \gamma_2 N) k(1) + \gamma_3 N E \left\{ \mathbf{v}^T(1) \right\} E \left\{ \mathbf{v}(1) \right\} \\
&\quad + \gamma_4 N + 2(\gamma_5 + \gamma_6 N) \mathbf{v}^T(0) \mathbf{R}_{\phi_{\bar{u}}}^T(1) E \left\{ \mathbf{v}(1) \right\} \\
&\quad + \gamma_6 N \mathbf{v}^T(0) \mathbf{R}_{\bar{u}}(1) \mathbf{v}(0) - \gamma_6 N \sigma_r^2 [\mathbf{a}^T \mathbf{Q}(0) \mathbf{a}] \\
&\quad + \gamma_7 \mathbf{v}^T(0) \mathbf{R}_{\phi_{\bar{u}}}^T(1) \mathbf{R}_{\phi_{\bar{u}}}(1) \mathbf{v}(0)
\end{aligned} \tag{5.82}$$

e aplicando $k(1)$ em $k(2)$,

$$\begin{aligned}
k(2) \equiv & (1 - \gamma_1 + \gamma_2 N) \left[\begin{aligned} & (1 - \gamma_1 + \gamma_2 N) k(0) + \gamma_3 NE \{ \mathbf{v}^T(0) \} E \{ \mathbf{v}(0) \} \\ & + \gamma_4 N + 2(\gamma_5 + \gamma_6 N) \mathbf{v}^T(0) \mathbf{R}_{\phi\bar{u}}^T(0) E \{ \mathbf{v}(0) \} \\ & + \gamma_6 N \mathbf{v}^T(0) \mathbf{R}_{\bar{u}}(0) \mathbf{v}(0) - \gamma_6 N \sigma_r^2 [\mathbf{a}^T \mathbf{Q}(1) \mathbf{a}] \\ & + \gamma_7 \mathbf{v}^T(0) \mathbf{R}_{\phi\bar{u}}^T(0) \mathbf{R}_{\phi\bar{u}}(0) \mathbf{v}(0) \end{aligned} \right] \\
& + \gamma_3 NE \{ \mathbf{v}^T(1) \} E \{ \mathbf{v}(1) \} + 2(\gamma_5 + \gamma_6 N) \mathbf{v}^T(0) \mathbf{R}_{\phi\bar{u}}^T(1) E \{ \mathbf{v}(1) \} \\
& + \gamma_4 N + \gamma_6 N \mathbf{v}^T(0) \mathbf{R}_{\bar{u}}(1) \mathbf{v}(0) + \gamma_7 \mathbf{v}^T(0) \mathbf{R}_{\phi\bar{u}}^T(1) \mathbf{R}_{\phi\bar{u}}(1) \mathbf{v}(0) \\
& - \gamma_6 N \sigma_r^2 [\mathbf{a}^T \mathbf{Q}(0) \mathbf{a}].
\end{aligned} \tag{5.83}$$

Rearranjando os termos de (5.83), resulta em

$$\begin{aligned}
k(2) \equiv & (1 - \gamma_1 + \gamma_2 N)^2 k(0) + \gamma_3 (1 - \gamma_1 + \gamma_2 N) NE \{ \mathbf{v}^T(0) \} E \{ \mathbf{v}(0) \} \\
& + \gamma_3 NE \{ \mathbf{v}^T(1) \} E \{ \mathbf{v}(1) \} + \gamma_4 (1 - \gamma_1 + \gamma_2 N) N + \gamma_4 N \\
& + 2(1 - \gamma_1 + \gamma_2 N) (\gamma_5 + \gamma_6 N) \mathbf{v}^T(0) \mathbf{R}_{\phi\bar{u}}^T(0) E \{ \mathbf{v}(0) \} \\
& + 2(\gamma_5 + \gamma_6 N) \mathbf{v}^T(0) \mathbf{R}_{\phi\bar{u}}^T(1) E \{ \mathbf{v}(1) \} \\
& + \gamma_6 (1 - \gamma_1 + \gamma_2 N) N \mathbf{v}^T(0) \mathbf{R}_{\bar{u}}(0) \mathbf{v}(0) + \gamma_6 N \mathbf{v}^T(0) \mathbf{R}_{\bar{u}}(1) \mathbf{v}(0) \\
& + \gamma_7 (1 - \gamma_1 + \gamma_2 N) \mathbf{v}^T(0) \mathbf{R}_{\phi\bar{u}}^T(0) \mathbf{R}_{\phi\bar{u}}(0) \mathbf{v}(0) \\
& + \gamma_7 \mathbf{v}^T(0) \mathbf{R}_{\phi\bar{u}}^T(1) \mathbf{R}_{\phi\bar{u}}(1) \mathbf{v}(0) \\
& - \gamma_6 (1 - \gamma_1 + \gamma_2 N) N \sigma_r^2 [\mathbf{a}^T \mathbf{Q}(1) \mathbf{a}] - \gamma_6 N \sigma_r^2 [\mathbf{a}^T \mathbf{Q}(0) \mathbf{a}].
\end{aligned} \tag{5.84}$$

Este procedimento é repetido sucessivamente, resultando na forma fechada para a equação fechada para o traço dos momentos de segunda ordem.

Na iteração $n = P$, devido à natureza da matriz $\mathbf{Q}(-n+1)|_{n \geq P} = \mathbf{0}_P$, tem-se que $\mathbf{R}_{\phi\bar{u}}(n)|_{n \geq P} = \mathbf{0}_N$, $\mathbf{R}_{\bar{u}}(n)|_{n \geq P} = \mathbf{0}_N$, e $\mathbf{a}^T \mathbf{Q}(-n+1) \mathbf{a}|_{n \geq P} = 0$. Dessa forma:

$$k(P+1) \equiv (1 - \gamma_1 + \gamma_2 N) k(P) + \gamma_3 NE \{ \mathbf{v}^T(P) \} E \{ \mathbf{v}(P) \} + \gamma_4 N. \tag{5.85}$$

Entretanto, a influência dos elementos dependentes da matriz $\mathbf{Q}(-n+1)$ irá permanecer ao longo do tempo pelo efeito de memória

recursiva, sendo multiplicado por um fator de decaimento exponencial $(1-\gamma_1+\gamma_2N)^{n-P}$. Dessa forma, resolvendo (5.85) para $k(n)$, sendo $n > P$, tem-se que

$$\begin{aligned} k(n) &\cong (1-\gamma_1+\gamma_2N)^n k(0) \\ &+ \gamma_3 N \sum_{i=0}^{n-1} (1-\gamma_1+\gamma_2N)^{n-i-1} E\{\mathbf{v}^T(i)\} E\{\mathbf{v}(i)\} \\ &+ \gamma_4 N \sum_{i=0}^{n-1} (1-\gamma_1+\gamma_2N)^{n-i-1} + (1-\gamma_1+\gamma_2N)^{n-P} \tau \end{aligned} \quad (5.86)$$

em que a constante τ é determinada por

$$\begin{aligned} \tau &\triangleq 2 \sum_{i=0}^{P-1} (1-\gamma_1+\gamma_2N)^{P-i-1} (\gamma_5+\gamma_6N) \mathbf{v}^T(0) \mathbf{R}_{\phi_{\hat{u}}}^T(i) E\{\mathbf{v}(i)\} \\ &+ \gamma_6 N \sum_{i=0}^{P-1} (1-\gamma_1+\gamma_2N)^{P-i-1} \mathbf{v}^T(0) \mathbf{R}_{\hat{u}}^T(i) \mathbf{v}(0) \\ &- \gamma_6 N \sum_{i=0}^{P-1} (1-\gamma_1+\gamma_2N)^{P-i-1} \sigma_r^2 [\mathbf{a}^T \mathbf{Q}(-i+1) \mathbf{a}] \\ &+ \gamma_7 \sum_{i=0}^{P-1} (1-\gamma_1+\gamma_2N)^{P-i-1} \mathbf{v}^T(0) \mathbf{R}_{\phi_{\hat{u}}}^T(i) \mathbf{R}_{\phi_{\hat{u}}}(i) \mathbf{v}(0). \end{aligned} \quad (5.87)$$

A equação (5.86) não contempla os instantes $0 < n \leq P$, devido à complexidade computacional, mas apresenta um custo computacional reduzido em relação à versão recursiva para $n > P$. O valor esperado de $\mathbf{v}(n)$ é apresentado na equação (4.20).

5.3 Comportamento em Regime Permanente

Um filtro adaptativo não polarizado é dito operando em regime permanente quando

$$\begin{cases} \lim_{n \rightarrow \infty} E\{\mathbf{v}(n)\} = \mathbf{0}_{N \times 1} \\ \lim_{n \rightarrow \infty} E\{\mathbf{K}(n+1)\} = \lim_{n \rightarrow \infty} E\{\mathbf{K}(n)\} = \mathbf{K}_{\infty}. \end{cases} \quad (5.88)$$

A primeira condição é verificável através da equação (4.20), para β_1 na região de estabilidade (valor menor que a unidade). Nesta seção são obtidas equações determinísticas para o erro quadrático médio e o desvio quadrático médio em regime permanente e ambiente estacionário, respectivamente.

5.3.1 Desvio Quadrático Médio

O desvio quadrático médio (DQM) descreve a potência média das flutuações dos coeficientes em regime permanente. Assumindo-se a convergência em regime permanente para (5.72) podemos escrever

$$\mathbf{K}_\infty \approx \mathbf{K}_\infty - \gamma_1 \mathbf{K}_\infty + \gamma_2 k_\infty \mathbf{I}_N + \gamma_4 \mathbf{I}_N \quad (5.89)$$

sendo que \mathbf{K}_∞ é definido em (5.88) e $k_\infty = \text{tr}\{\mathbf{K}_\infty\}$. Dessa forma, isolando-se \mathbf{K}_∞ , obtém-se que

$$\mathbf{K}_\infty \approx \frac{1}{\gamma_1} [\gamma_2 k_\infty + \gamma_4] \mathbf{I}_N \quad (5.90)$$

e aplicando-se o operador traço em (5.90), tem-se que

$$k_\infty \approx \text{tr} \left\{ \frac{1}{\gamma_1} [\gamma_2 k_\infty + \gamma_4] \mathbf{I}_N \right\} = \frac{N}{\gamma_1} [\gamma_2 k_\infty + \gamma_4]. \quad (5.91)$$

Isolando k_∞ , chega-se em

$$k_\infty \approx \eta N \gamma_4 \quad (5.92)$$

onde η é definido como

$$\eta = \frac{1}{\gamma_1 - N \gamma_2} = \frac{(G-2)(G+2)}{(G+6)}. \quad (5.93)$$

Substituindo (5.93) em (5.92), resulta na expressão para o desvio quadrático médio:

$$\lim_{n \rightarrow \infty} k(n) \approx \left[\frac{(G-2)(G+2)}{(G+6)} \right] \gamma_4 N. \quad (5.94)$$

5.3.2 Erro Quadrático Médio em Regime Permanente

O erro quadrático médio em regime permanente é definido por:

$$\xi \triangleq \lim_{n \rightarrow \infty} E \{ e^2(n) \}. \quad (5.95)$$

Dessa forma, fazendo (5.35) tender ao infinito, tem-se que

$$\xi \approx \sigma_\phi^2 k_\infty + (1 + \mathbf{a}^T \mathbf{a}) \sigma_r^2. \quad (5.96)$$

Aplicando-se a equação (5.94) em (5.96)

$$\xi \approx \sigma_\phi^2 \left[\frac{(G-2)(G+2)}{(G+6)} \right] N\gamma_4 + (1 + \mathbf{a}^T \mathbf{a}) \sigma_r^2. \quad (5.97)$$

Substituindo-se γ_4 de (5.73) em (5.97), obtém-se a expressão determinística para o erro quadrático médio em regime permanente,

$$\xi \approx (1 + \mathbf{a}^T \mathbf{a}) \left[1 + \frac{N(G-2)(G+2)}{(G+6)(G^2+2G)} \right] \sigma_r^2. \quad (5.98)$$

Os resultados obtidos em (5.94) e (5.98) corroboram os apresentados em (ALMEIDA, 2004, pág. 83).

Através da expressão matemática para o regime permanente em (5.98) observa-se que o mecanismo de memória recursiva aplicado no transiente inicial, e que inclui o Efeito de Inicialização, não afeta o regime permanente. Dessa forma, os resultados obtidos nesta seção são idênticos aos obtidos em (ALMEIDA, 2004, p. 83) e as mesmas considerações são válidas para este trabalho. Assim, aumentando os graus de liberdade G da distribuição chi-quadrada, e consequentemente a relação $N \gg P$, leva à redução de $1 + [N(G-2)(G+2)] / [(G+6)(G^2+2G)]$, que por sua vez conduz à redução do EQM em regime permanente.

5.3.3 Erro Quadrático Médio em Excesso

O erro quadrático médio em excesso (EQM_{ex}) no regime permanente é obtido subtraindo de (5.96) o mínimo erro quadrático médio que seria produzido pela solução ótima do filtro de Wiener, ou seja, a potência do ruído, de forma que

$$\text{EQM}_{\text{ex}} \equiv \left[\frac{N(G-2)(G+2)}{(G+6)(G^2+2G)} \right] (1 + \mathbf{a}^T \mathbf{a}) \sigma_r^2 + (\mathbf{a}^T \mathbf{a}) \sigma_r^2. \quad (5.99)$$

5.4 Conclusão

Neste capítulo foram derivadas equações para o erro quadrático médio em regime transitório e permanente, além de expressões para os momentos de segunda ordem do vetor de erro dos coeficientes. A equação derivada para o EQM (5.35) é equivalente à equação (3.63) de (ALMEIDA, 2004),

$$E\{e^2(n)\} = \left(1 + \mathbf{a}^T \mathbf{a} + \text{tr} \left\{ E \left\{ \left[\mathbf{U}^T(n) \mathbf{U}(n) \right]^{-1} \right\} \right\} \right) \sigma_r^2 + \text{tr} \left[\mathbf{R}_{\phi\phi} \mathbf{K}(n) \right] \quad (5.100)$$

exceto pela inclusão dos três últimos termos do lado direito, que são contribuições deste trabalho para modelar o Efeito de Inicialização.

De forma associada, é apresentada uma expressão recursiva para o traço dos momentos de segunda ordem em regime transiente e sua derivação na forma fechada. A nova equação recursiva de $\mathbf{K}(n)$ é semelhante à obtida anteriormente por Almeida (2004), exceto pelos seis últimos termos. Os quatro primeiros termos do lado direito de (5.69) determinam a constante de tempo do decaimento no modelo de $\mathbf{K}(n)$. A equação na forma fechada para $k(n)$ foi obtida para $n \geq P$, instante em que o escalar τ , que representa a memória intrínseca dos momentos de segunda ordem devido ao Efeito de Inicialização, se torna um valor determinístico. Essa valor se esvanece exponencialmente conforme as iterações se sucedem.

6 RESULTADOS E DISCUSSÕES

Neste capítulo são apresentados alguns resultados representativos de um conjunto maior de simulações abrangendo variadas combinações de parâmetros. Essas simulações visam validar o modelo proposto, além de efetuar uma comparação com o modelo de Almeida (2004). A referência empregada é a simulação Monte Carlo do pseudo algoritmo de projeções afins.

As simulações são realizadas considerando o problema de identificação de sistema da Figura 6, sendo conduzidas em um ambiente controlado com sinais sintéticos. A análise proposta avalia qualitativamente, e quantitativamente, o comportamento do regime transitório inicial e do regime permanente. As figuras de mérito empregadas são o comportamento médio dos coeficientes e o erro quadrático médio.

6.1 Condições de Teste

Os parâmetros de simulação empregados na maioria dos exemplos visam enfatizar o Efeito de Inicialização, visto que os modelos de Almeida (2004) descrevem adequadamente o comportamento do PAP quando esse efeito não é significativo. Nessa condição, como será mostrado, os modelos propostos nesse trabalho produzem previsões do comportamento do PAP mais precisas que as obtidas pelos modelos de Almeida (2004). Quando não particularmente especificados, os parâmetros de simulação utilizados são:

- o sinal de excitação de entrada $u(n)$ é um processo aleatório AR, sendo que as 10.000 primeiras amostras são descartadas de forma a diminuir o efeito da não-estacionaridade decorrente do transitório de inicialização;
- o ruído de medida é produzido por um processo branco Gaussiano de média zero e variância $\sigma_r^2 = 10^{-6}$;
- a variância da inovação $z(n)$ é ajustada de forma que a relação sinal-ruído $SNR = 10 \cdot \log_{10}(\mathbf{w}^{oT} \mathbf{R}_u \mathbf{w}^o / \sigma_r^2)$ seja de 60 dB;
- a inicialização do filtro adaptativo é realizada na origem, ou seja, o vetor inicial de coeficientes é $\mathbf{w}(0) = [0 \ 0 \ \dots \ 0]^T$;

- a matriz $\mathbf{U}(n)$ é completamente preenchida antes do início da atualização dos coeficientes;
- as matrizes $\mathbf{R}_{\phi_i}(n)$ e $\mathbf{R}_i(n)$ são obtidas através das equações fornecidas no Apêndice B e Apêndice C, respectivamente.

Além disso, duas plantas \mathbf{w}^o diferentes, com 128 amostras cada, são utilizadas:

- resposta ao impulso normalizada de um canal de eco de um simulador híbrido de rede da norma ITU-T G.168 (RECOMMENDATION ITU-T G.168, 02/2002, p. 88) (ANEXO D.2, TABELA D.5), conforme Figura 8;

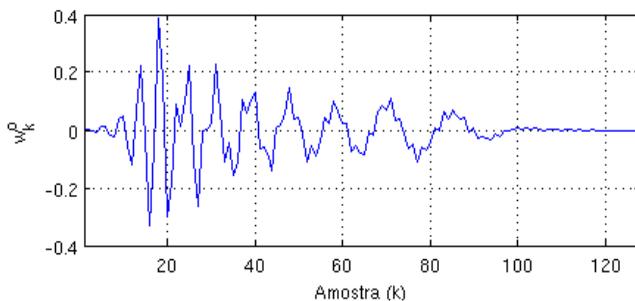


Figura 8 – Resposta impulsiva de um canal de eco (ITU-T G.168).

- resposta ao impulso de um ambiente acústico de longa duração utilizada nos exemplos de (ALMEIDA, 2004).

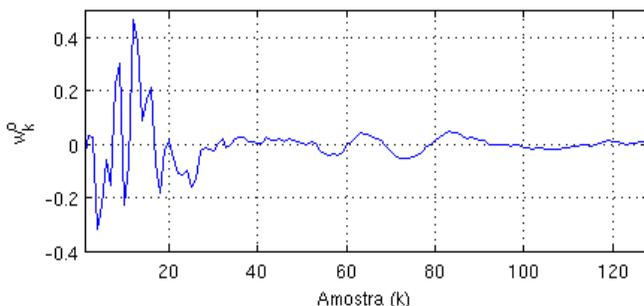


Figura 9 - Resposta impulsiva utilizada em Almeida (2004).

6.2 O Efeito de Inicialização

Essa seção apresenta um exemplo, dentre diversos resultados obtidos, que demonstra a diferença de desempenho entre os algoritmos APA e PAP, considerando o uso de passo unitário. Conforme discussão na Seção 3, essa diferença ocorre em virtude do Efeito de Inicialização. Comparações o resultado da equação (3.48), com o novo modelo derivado e o anteriormente obtido por Almeida (2004) são fornecidos.

A Figura 10 apresenta a curva de aprendizado do erro quadrático médio quando os dois algoritmos são sujeitos exatamente às mesmas condições. Foram utilizados como parâmetros: um sinal de entrada autoregressivo, conforme equação (2.31), com vetor de coeficientes $\mathbf{a} = [1,96 \ -0,9893]$ e $z(n)$ um processo aleatório branco Gaussiano com variância $\sigma_z^2 \approx 0,0575$; o ruído aditivo é um processo Gaussiano, de média zero e variância $\sigma_r^2 \approx 10^{-6}$; $N = 128$; $P = 2$, $\mu = 1$ e $\mathbf{w}(0) = [0 \ 0 \ \dots \ 0]^T$. A planta utilizada é a resposta ao impulso de um canal de eco da norma ITU-T G.168, conforme Figura 8. A simulação Monte Carlo foi obtida através de uma média de 500 promediações.

O resultado presente na Figura 10 demonstra uma pequena diferença entre os dois algoritmos para um processo AR(2), através de simulação Monte Carlo, contrariando as afirmações de Bouteille et al. (1999). Salienta-se que a curva azul de (3.48) esta sobreposta à curva espiculada em cor preta, que representa (2.38), demonstrando a equivalência da equação para o algoritmo APA. A curva em vermelho representa o processo de atualização dos coeficientes do algoritmo PAP de (2.47). O novo modelo, representado pela curva em magenta, descreve o comportamento do PAP, enquanto o modelo anteriormente derivado em Almeida (2004), em ciano, produz uma boa descrição da curva do APA após um determinado número de iterações. Um erro de 12 dB ocorre na primeira iteração, conforme detalhe da figura.

Porém salienta-se que os parâmetros deste exemplo foram escolhidos para demonstrar uma situação extrema considerando uma baixa ordem do processo autoregressivo, e que geralmente, para esse caso, essa diferença não ocorre, ou é insignificante. Por exemplo, a Figura 11 foi gerada com um processo AR(1) - $u(n) = -0,9u(n-1) + z(n)$; $\sigma_z^2 \approx 1$; $P = 1$; e demais parâmetros utilizados na simulação anterior. Ela demonstra a similaridade na curva de aprendizado entre os dois algoritmos e os modelos de Almeida (2004) e proposto e anteriormente obtido. Nessa situação, $\mathbf{\kappa}(0)$ é desprezível.

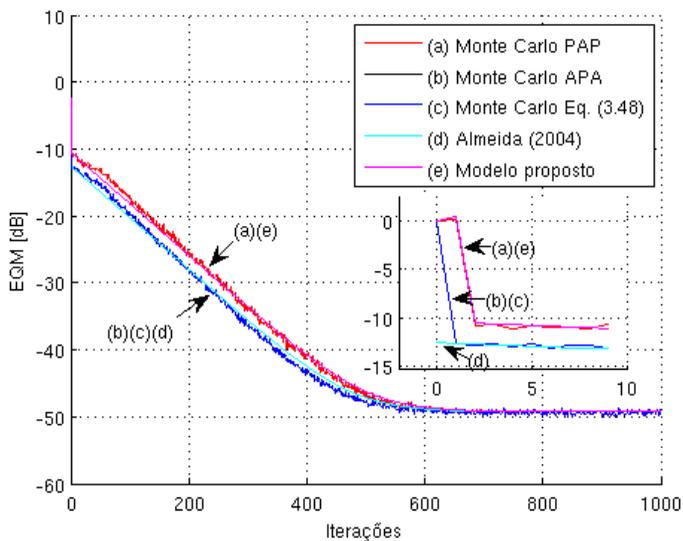


Figura 10 – Comportamento do EQM para um processo AR(2); $N = 128$ e $P = 2$. (a) PAP em vermelho; (b) APA em preto; (c) equação (3.48) em azul; (d) modelo teórico de Almeida (2005) em ciano e (e) modelo proposto em magenta.

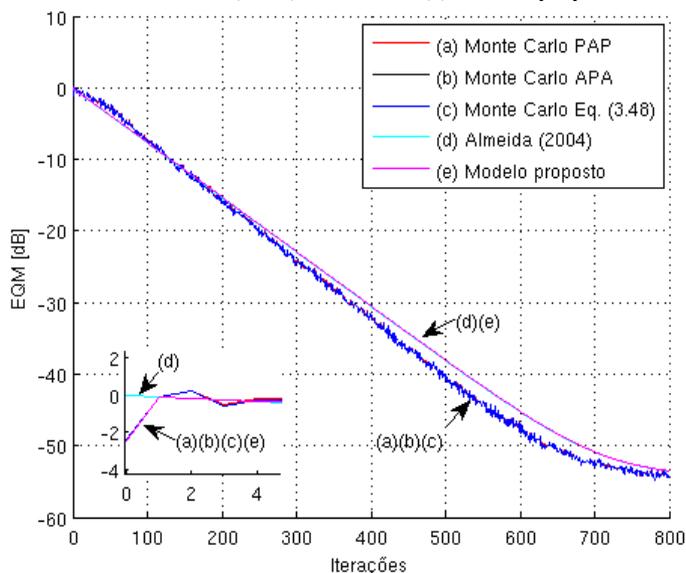


Figura 11 – Comportamento do EQM para um processo AR(1); $N = 64$ e $P = 1$. (a) PAP em vermelho; (b) APA em preto; (c) equação (3.48) em azul; (d) modelo teórico de Almeida (2005) em ciano e (e) modelo proposto em magenta.

6.3 Comportamento Médio dos Coeficientes

Esta seção apresenta os resultados para o comportamento médio dos coeficientes e para o vetor de erro dos coeficientes utilizando os modelos obtidos no Capítulo 4. São fornecidas comparações entre o modelo proposto na equação (4.14) e o modelo de Almeida (2004). Como referência, foram utilizadas 5000 realizações de Monte Carlo. As figuras apresentadas em escala logarítmica visam facilitar a visualização da diferença gerada no período inicial do regime transiente para o vetor de erro dos coeficientes, em virtude da inclusão de (3.30).

• Exemplo 6.3.1

Os seguintes parâmetros foram empregados na simulação: o sinal de excitação $u(n)$ é obtido através de um modelo AR(1), que é definido pela equação $u(n) = -0,9u(n-1) + z(n)$. A variância da inovação é $\sigma_z^2 \approx 1,03$; o número de coeficientes do filtro é $N = 64$, e o número de vetores passados é $P = 1$.

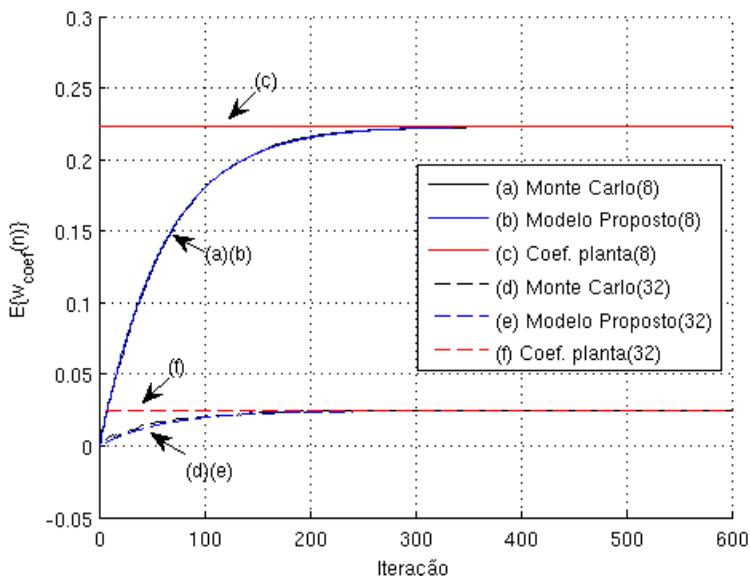


Figura 12 – Comportamento médio do 8º e do 32º coeficiente para o Exemplo 6.3.1. AR(1); $N = 64$ e $P = 1$. (a)(d) simulação Monte Carlo (preto); (b)(e) modelo proposto em (4.14) (azul); e (c)(f) coeficiente da planta (vermelho).

Os coeficientes da planta empregados são os primeiros 64 coeficientes da resposta ao impulso de um ambiente acústico utilizado em Almeida (2004), conforme Figura 9. Essa configuração de parâmetros é similar à utilizada no Exemplo 3.7.1 de Almeida (2004, p. 67).

A Figura 12 apresenta o comportamento médio do oitavo e trigésimo segundo coeficientes e a curva gerada pelo modelo analítico da equação (4.14). Ambas as predições realizadas pelo modelo proposto são similares às obtidas pelo modelo de Almeida (2004, p. 67) e encontram-se praticamente sobrepostas. Essa semelhança pode ser verificada detalhadamente em escala logarítmica na Figura 13, para as primeiras 80 iterações da evolução do décimo coeficiente. Ou seja, para essa configuração de parâmetros, a influência da ausência de $\kappa(0)$ é praticamente desprezível, visto que P é pequeno e a inicialização do vetor de coeficientes está próxima à solução ótima. Dessa forma, o modelo proposto e o modelo de Almeida (2004) apresentam excelente concordância com a simulação Monte Carlo.

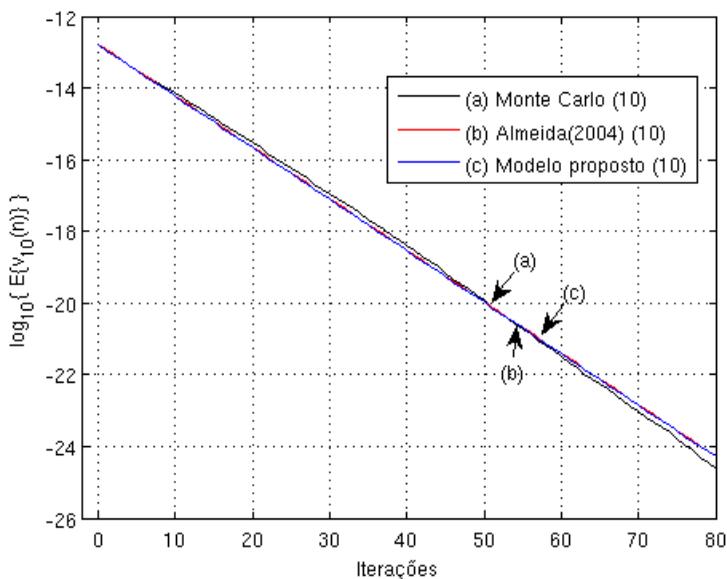


Figura 13 – Comportamento médio do 10º coeficiente de $v(n)$ para o Exemplo 6.3.1 em escala logarítmica. AR(1); $N = 64$ e $P = 1$. (a) simulação Monte Carlo (preto); (b) modelo em Almeida (2004) (vermelho) e (c) modelo proposto em (4.11) (azul).

• **Exemplo 6.3.2**

Para este exemplo é utilizada a seguinte configuração de parâmetros: o sinal de excitação é gerado por um processo AR(2) $u(n) = 1,96u(n-1) - 0,9893u(n-2) + z(n)$; $\sigma_z^2 = 0,0575$; $N = 128$; $P = 2$. Os coeficientes da planta são a resposta ao impulso de um canal de eco de linha fornecido na norma ITU-T G.168. Esse exemplo visa demonstrar um caso no qual P é pequeno, $\mathbf{w}(0)$ é inicializado em zero, ou seja, próximo à solução ótima, e o sinal de entrada é fortemente correlacionado.

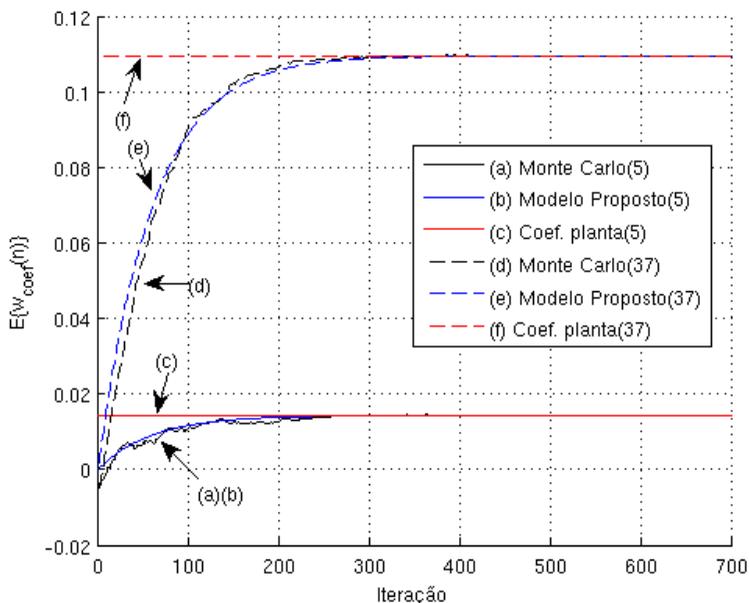


Figura 14 – Comportamento médio do 5º e 37º coeficiente para o Exemplo 6.3.2. AR(2); $N = 128$ e $P = 2$. (a)(d) simulação Monte Carlo (preto); (b)(e) modelo proposto em (4.14) (azul); e (c)(f) coeficiente da planta (vermelho).

Através da Figura 14 observa-se uma adequada acurácia do modelo proposto em (4.14) em relação à simulação Monte Carlo. Além disso, nota-se uma perfeita modelagem dos coeficientes da planta em regime permanente.

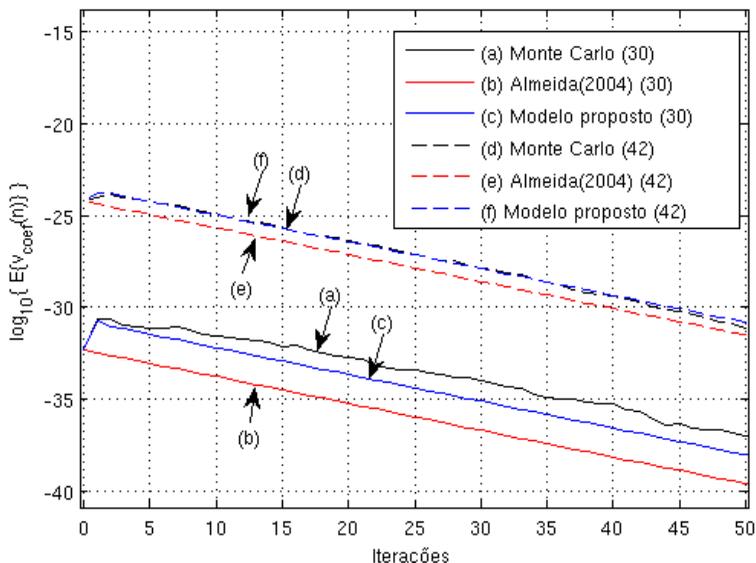


Figura 15 – Comportamento médio do 30º e 42º coeficiente de $\mathbf{v}(n)$ para o Exemplo 6.3.2 em escala logarítmica. AR(2); $N = 128$ e $P = 2$. (a)(d) simulação Monte Carlo (preto); (b)(e) modelo em Almeida (2004) (vermelho) e (c)(f) modelo proposto em (4.11) (azul).

A Figura 15, por sua vez, apresenta a evolução do 30º e 42º coeficiente do vetor de erro dos coeficientes $\mathbf{v}(n)$ para as primeiras 50 iterações do modelo obtido em (4.11) e o desenvolvido em Almeida (2004). O resultado é apresentado em escala logarítmica. Verifica-se que o modelo proposto acompanha a simulação Monte Carlo com maior acurácia que o modelo de Almeida (≈ 2 dB para o 30º coeficiente e ≈ 1 dB para o 42º).

Esse exemplo tem como principal objetivo demonstrar que o efeito gerado pela inicialização pode acarretar em um comportamento não desprezível, mesmo para um baixo valor de ordem do algoritmo. Isso ocorre devido à elevada correlação de $\mathbf{u}(n)$, que conseqüentemente influencia a matriz $\mathbf{R}_{\phi_i}(n)$, em virtude do Efeito de Inicialização. Devido ao efeito de memória da equação recursiva, descrito na Subseção 4.2, essa diferença se propaga até o regime permanente, sendo reduzido exponencialmente, quando então se esvanece.

• **Exemplo 6.3.3**

Seja um processo de filtragem adaptativa composto pelos seguintes parâmetros: sinal de excitação gerado por um processo AR(30) com coeficientes $a_i = (-0,99)^i$ para $i = 1, \dots, 30$; variância da inovação $\sigma_z^2 \approx 1$; $N = 64$ e $P = 30$. Os coeficientes da planta são as 64 primeiras amostras da resposta ao impulso de um canal de eco da norma ITU-T G.168, conforme Figura 8.

Os coeficientes foram inicializados em $\mathbf{w}(0) = [100 \ 100 \ \dots \ 100]^T$. Neste exemplo a matriz $\mathbf{R}_{\phi_i}(n)$ foi estimada através de realizações do sinal de excitação.

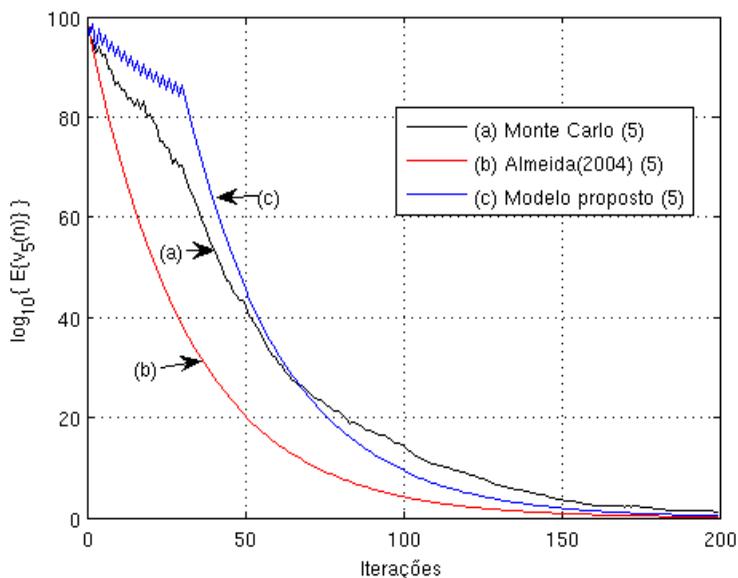


Figura 16 – Comportamento médio do 5º coeficiente de $\mathbf{v}(n)$ para o filtro do Exemplo 6.3.3 em escala linear; $\mathbf{R}_{\phi_i}(n)$ estimado numericamente; AR(30); $N = 64$ e $P = 30$. (a) simulação Monte Carlo (preto); (b) modelo em Almeida (2004) (vermelho) e (c) modelo proposto em (4.14) (azul).

A Figura 16 apresenta a comparação entre o modelo proposto em (4.11) para o quinto coeficiente de $\mathbf{v}(n)$ em relação ao modelo de Almeida (2004). Pode-se observar uma maior precisão no modelo proposto. Salienta-se que esse resultado foi obtido com $\mathbf{R}_{\phi_{ii}}(n)$ estimado numericamente a partir de realizações do sinal de excitação.

No caso de avaliação de $\mathbf{R}_{\phi_{ii}}(n)$ através do equacionamento do Apêndice B o modelo proposto apresenta resultado semelhante aos obtidos pelo modelo de Almeida (2004) o que indica necessidade de aprimoramento futuro na estimação dessa matriz.

Essa diferença provavelmente é ocasionada pela aproximação de $\hat{\mathbf{a}}(n)$ por \mathbf{a} , considerada pela hipótese **H7** e pelo fato de N não ser muito maior que P , não satisfazendo a hipótese **H1**. Salienta-se que esse exemplo é um caso extremo visando demonstrar os limites de validade do modelo.

6.4 Erro Quadrático Médio

Esta seção apresenta resultados que demonstram o incremento na acurácia de predição do erro quadrático médio pelo modelo proposto de (5.35) em relação ao modelo de Almeida (2004), utilizando como referência a simulação Monte Carlo. Seis exemplos são apresentados, ressaltando as características do modelo proposto. Nas simulações, também são apresentados os resultados da equação para o regime permanente (5.98) e do mínimo erro quadrático médio produzido pelo filtro de Wiener na condição ótima, ou seja, a potência do ruído $r(n)$.

Para determinar o número mínimo de promediações R para o procedimento de teste, o método descrito no Anexo E é utilizado, assumindo-se que a variável aleatória $e(n)$ é Gaussiana. Considerando os parâmetros $\alpha = 0,05$, potência $(1-\beta)$ de 0,9 e $\lambda = 1,1$, resulta na necessidade de serem realizadas 573 promediações para que seja obtido um intervalo de confiança de 0,5%.

• Exemplo 6.4.1

Os parâmetros utilizados neste exemplo são similares ao primeiro exemplo de Almeida (2004, p. 85): o sinal de excitação $u(n)$ é gerado através de um processo AR(1) $u(n) = -0,9u(n-1) + z(n)$; $\sigma_z^2 \approx 1,03$; $N = 64$; $P = 1$. Os coeficientes da planta empregados são os 64 primeiros coeficientes da resposta ao impulso de um ambiente acústico utilizado em Almeida (2004), conforme Figura 9.

Neste primeiro exemplo, verifica-se através da Figura 17 que ambos os modelos possuem resultados similares e se aproximam da simulação Monte Carlo. Entretanto, no detalhe superior do gráfico da Figura 17 nota-se que para o modelo de Almeida (2004) apresenta uma diferença de aproximadamente 3,5 dB na primeira iteração de atualização dos coeficientes, enquanto o modelo proposto se ajusta muito bem à curva da simulação. Este resultado se deve a característica atenuada do Efeito de Inicialização (nota-se que $P = 1$) que resulta nos termos extras em (5.35) em relação à (5.98), assim como em 6.3.1. Essa similaridade é válida para demais exemplos de Almeida (2004).

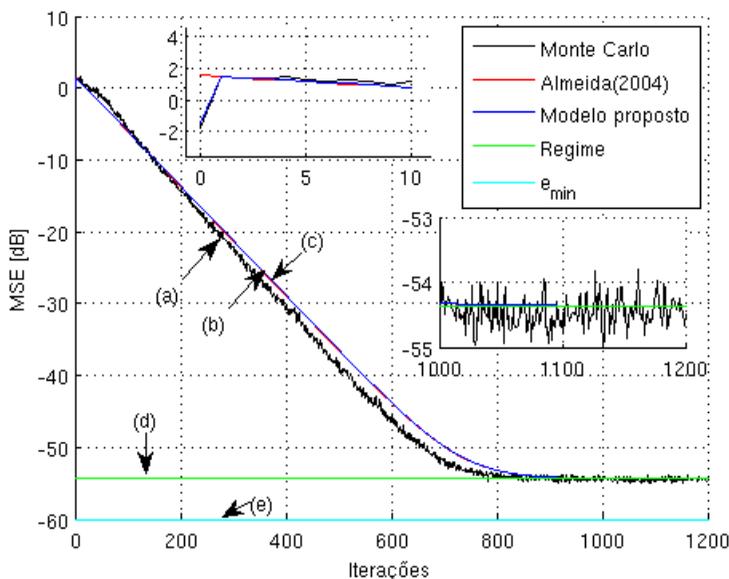


Figura 17 – Erro quadrático médio para o Exemplo 6.4.1. AR(1); $N = 64$; $P = 1$. (a) Simulação Monte Carlo (preto); (b) modelo em Almeida (2004) (vermelho); (c) modelo proposto em (5.35) (azul), (d) regime permanente segundo (5.98); e (e) erro mínimo.

• Exemplo 6.4.2

Este exemplo utiliza os mesmos parâmetros do Exemplo 6.3.2: o processo AR(2) utilizado é gerado através da equação $u(n) = 1,96u(n-1) - 0,9893u(n-2) + z(n)$; $\sigma_z^2 = 0,0575$; $N = 64$; $P = 2$. Os coeficientes da planta são fornecidos pela resposta ao impulso de um canal de eco conforme modelo 4 do Anexo D.2 da norma ITU-T G.168.

Em geral, o Efeito de Inicialização é pouco notado em uma configuração de parâmetros com ordem pequena. Entretanto, um sinal de excitação altamente correlacionado pode acarretar em uma discrepância nas primeiras P iterações, resultando em uma grande diferença entre $\mathbf{U}(n)\mathbf{v}(0) = r(n-1)$ em $n = 0$. Nota-se na Figura 18 que, para o exemplo em questão, o modelo de Almeida (2004) apresenta um erro de predição de aproximadamente 13 dB nas primeiras iterações, diminuindo à medida que a adaptação evolui. As predições produzidas pelo modelo proposto correspondem adequadamente à curva da simulação Monte Carlo.

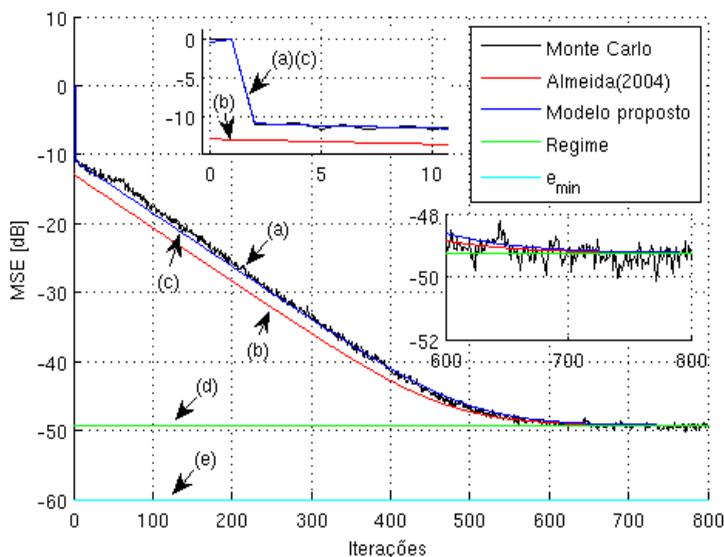


Figura 18 – Erro quadrático médio para o Exemplo 6.4.2. AR(2); $N = 64$; $P = 2$. (a) Simulação Monte Carlo (preto); (b) modelo em Almeida (2004) (vermelho), (c) modelo proposto em (5.35) (azul); (d) regime permanente segundo (5.98); e (e) erro mínimo.

- **Exemplo 6.4.3**

Neste exemplo são utilizados os mesmos parâmetros empregados no Exemplo 6.4.2, para os valores de $P = \{2, 8\}$. A Figura 19 apresenta comparações entre o modelo proposto (5.35) e a simulação Monte Carlo para esses valores de P , considerando $N = 128$.

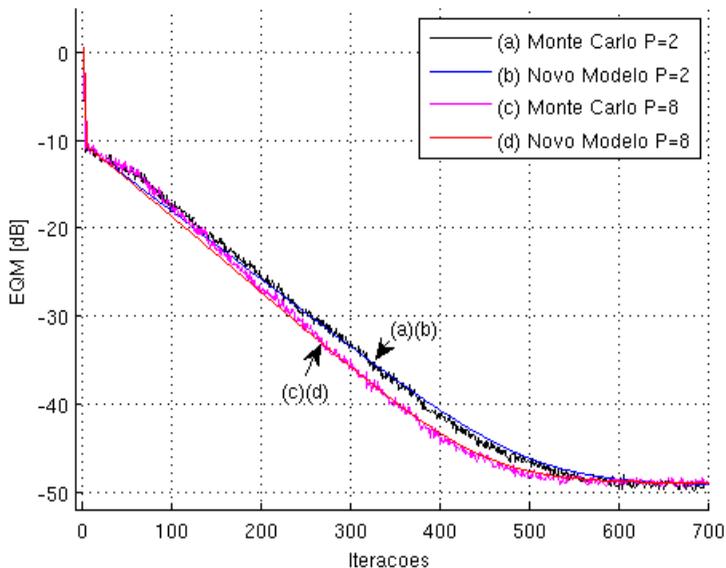


Figura 19 – Erro quadrático médio para o Exemplo 6.4.3. AR(2); $N = 64$. simulação Monte Carlo: (a) $P = 2$ (preto) e (c) $P = 8$ (magenta); modelo proposto em (5.35): (b) $P = 2$ (azul) e (d) $P = 8$ (vermelho).

Observa-se que a curva do modelo proposto se ajusta precisamente à simulação do algoritmo para ambos os casos.

- **Exemplo 6.4.4**

Este exemplo utiliza como sinal de excitação um processo AR(15) com coeficientes $a_i = \{3,0623 -3,4051 1,0137 0,9671 -0,4409 -0,8332 0,9136 -0,1986 -0,1313 -0,0624 0,1744 0,0049 -0,1331 0,0997 -0,0323\}$. Estes coeficientes foram obtidos a partir da estimativa dos parâmetros de um sinal de *eletroencefalografia* (EEG) utilizando o método de Burg. Os demais parâmetros são: $\sigma_z^2 = 0,025$; $N = 64$; $P = 15$.

Para a planta são utilizados os primeiros 64 coeficientes da resposta ao impulso de um canal de eco de linha, conforme o quarto modelo do Anexo D.2 da norma ITU-T G.168. Esse exemplo utiliza parâmetros que acentuam a influência de $\kappa(0)$, e portanto, do Efeito de Inicialização (P e dispersão de auto-valores do sinal de excitação elevados) e, portanto desfavorece o modelo de Almeida (2004).

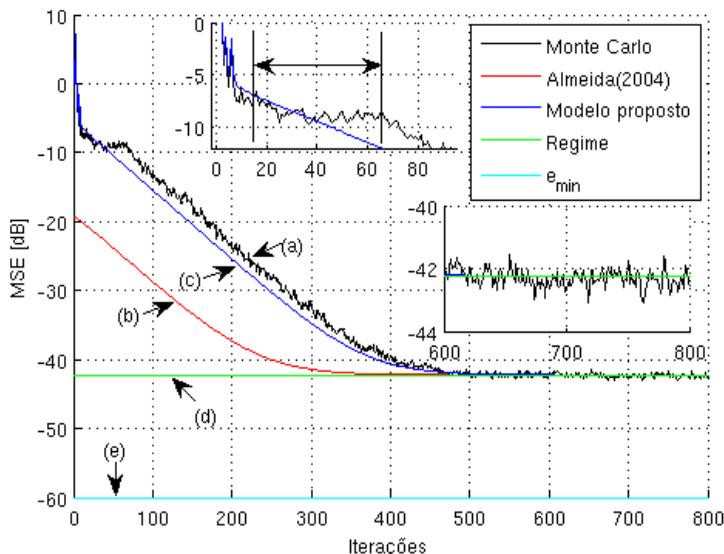


Figura 20 – Erro quadrático médio para o Exemplo 6.4.4. AR(15); $N = 64$; $P = 15$. (a) Simulação Monte Carlo (preto); (b) modelo em Almeida (2004) (vermelho); (c) modelo proposto em (5.35) (azul); (d) regime permanente segundo (5.98); e (e) erro mínimo.

De acordo com a Figura 20, o modelo proposto produz excelente descrição do regime transitório inicial e em regime permanente. Verifica-se também um período intermediário, entre os instantes P e N , no qual o modelo proposto se aproxima, mas não consegue descrever completamente a dinâmica do algoritmo (veja detalhe superior da Figura 20). Este comportamento foi observado em todos os exemplos realizados para o erro quadrático médio. Em regime permanente, ambos os modelos predizem com eficácia o comportamento médio do algoritmo, tendo suas curvas sobrepostas com o resultado da equação obtida para o regime permanente.

• **Exemplo 6.4.5**

Este exemplo apresenta resultados considerando uma inicialização idêntica à resposta ao impulso da planta, $\mathbf{w}(0) = \mathbf{w}^0$. O processo AR(15) utilizado é o mesmo do Exemplo 6.4.4; $\sigma_z^2 = 0,0154$; $P = 15$. Foram utilizados dois comprimentos para a planta: $N = 64$ e $N = 128$.

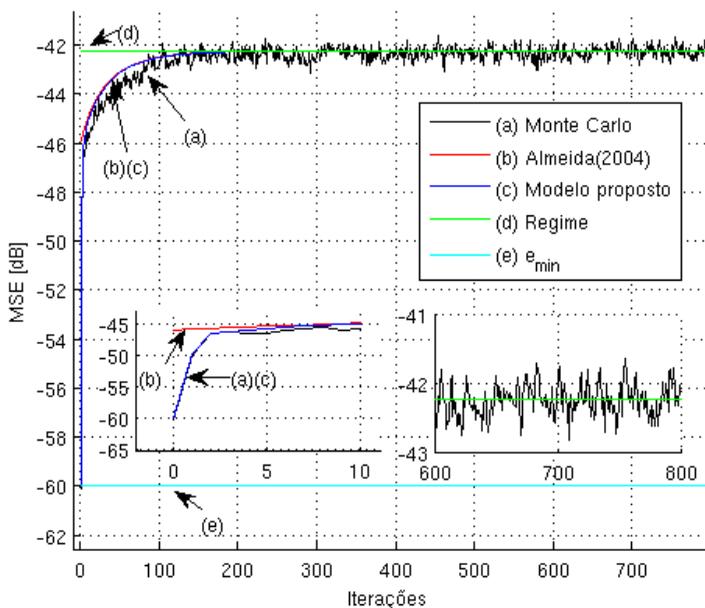


Figura 21 – Erro quadrático médio para o Exemplo 6.4.5. AR(15); $\mathbf{w}(0) = \mathbf{w}^0$; $N = 64$, $P = 15$. (a) Simulação Monte Carlo (preto); (b) modelo de (ALMEIDA, 2004) (vermelho) e (c) modelo proposto (azul).

Observa-se através das Figuras 21 ($N = 64$) e 22 ($N = 128$), que para o caso de uma inicialização nos coeficientes ótimos, o novo modelo apresenta uma ótima concordância com o comportamento médio do PAP através da simulação Monte Carlo.

O modelo de Almeida (2004), entretanto, apresenta um erro inicial de aproximadamente 14 dB. Essa diferença é visualizada no detalhe do canto inferior esquerdo. Nota-se, que embora a diferença em decibéis seja grande, o valor absoluto é muito pequeno.

Em regime permanente, ambos os modelos se ajustam perfeitamente à simulação Monte Carlo. Este exemplo demonstra que mesmo em uma situação em que $\mathbf{v}(0) = \mathbf{0}_{N \times 1}$, $\boldsymbol{\kappa}(0)$ não é nulo, e portanto, há a ocorrência do Efeito de Inicialização, devido à relação entre $r(n-1)$ e $\mathbf{U}^T(n)\mathbf{v}(0)$, em $n = 0$.

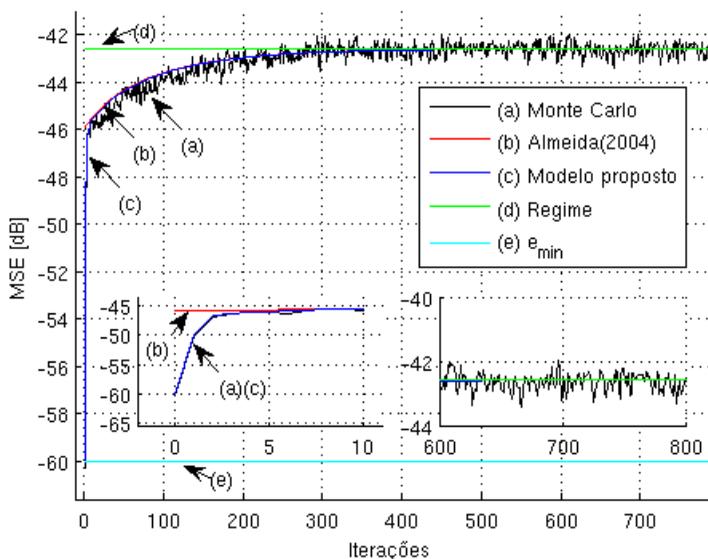


Figura 22 – Erro quadrático médio para o Exemplo 6.4.5. AR(15); $\mathbf{w}(0) = \mathbf{w}^o$; $N = 128$, $P = 15$. (a) Simulação Monte Carlo (preto); (b) modelo de (ALMEIDA, 2004) (vermelho); (c) modelo proposto (azul); (d) regime permanente segundo (5.98); e (e) erro mínimo.

• Exemplo 6.4.6

Os parâmetros utilizados nesta simulação são: sinal de excitação gerado por um processo AR(30) com coeficientes $a_i = (-0,99)^i$ para $i = 1, \dots, 30$; variância da inovação $\sigma_z^2 \approx 0,01$; $N = \{40, 64, 128\}$; e $P = 30$. Os coeficientes da planta são as N primeiras amostras da resposta ao impulso do canal de eco de linha da Figura 8. Os coeficientes foram inicializados em $\mathbf{w}(0) = [0 \ 0 \ \dots \ 0]^T$. As Figuras 22 à 24 demonstram o resultado obtido para $N = 128, 64$ e 40 , respectivamente. Este exemplo acentua o período em que o Efeito de Inicialização ocorre (P elevado) demonstrando como o modelo proposto se ajusta ao comportamento médio do algoritmo obtido por simulação Monte Carlo. Observa-se através da Figura 23 ($N = 128$) que o modelo proposto em (5.35) apresenta uma excelente concordância com a simulação Monte Carlo. Nessa situação, conforme visualizado no quadro superior da Figura 23, o modelo de Almeida (2004) apresenta um erro inicial de 22 dB, com progressiva diminuição ao longo do processo de adaptação, até que em regime permanente ambos modelos produzem a mesma diferença de aproximadamente 0,7 dB em relação à simulação Monte Carlo.

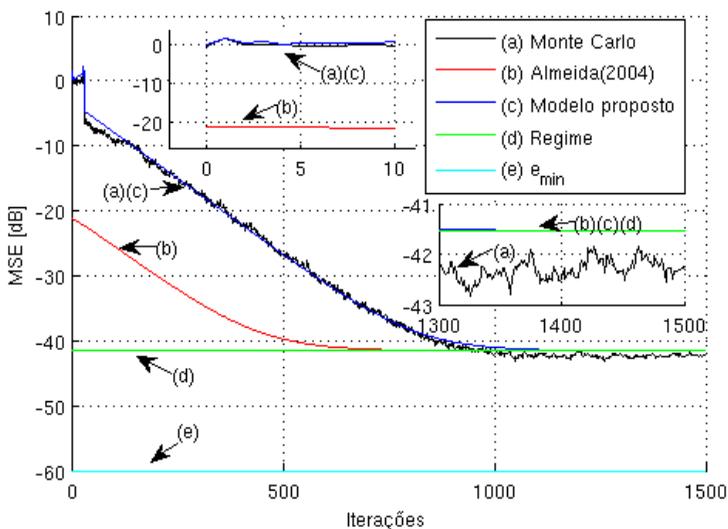


Figura 23 – Erro quadrático médio para o Exemplo 6.4.6. AR(30); $N = 128$; $P = 30$. (a) Simulação Monte Carlo (preto); (b) modelo em Almeida (2004) (vermelho); (c) modelo proposto em (5.35) (azul); (d) regime permanente segundo (5.98); e (e) erro mínimo.

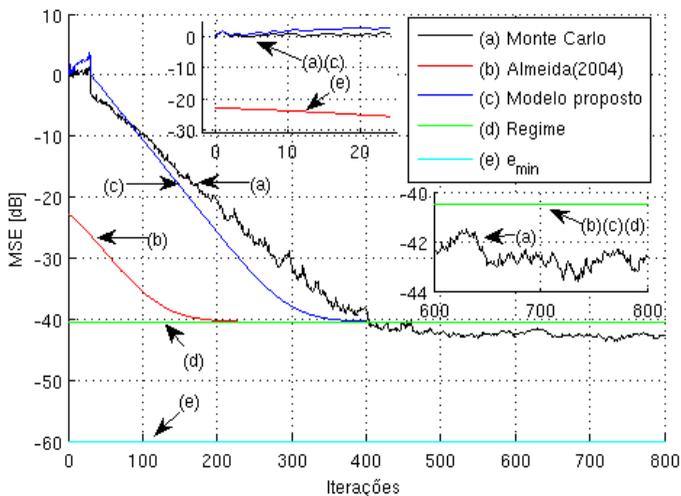


Figura 24 – Erro quadrático médio para o Exemplo 6.4.6. AR(30); $N = 64$; $P = 30$. (a) Simulação Monte Carlo (preto); (b) modelo em Almeida (2004) (vermelho); (c) modelo proposto em (5.35) (azul); (d) regime permanente segundo (5.98); e (e) erro mínimo.

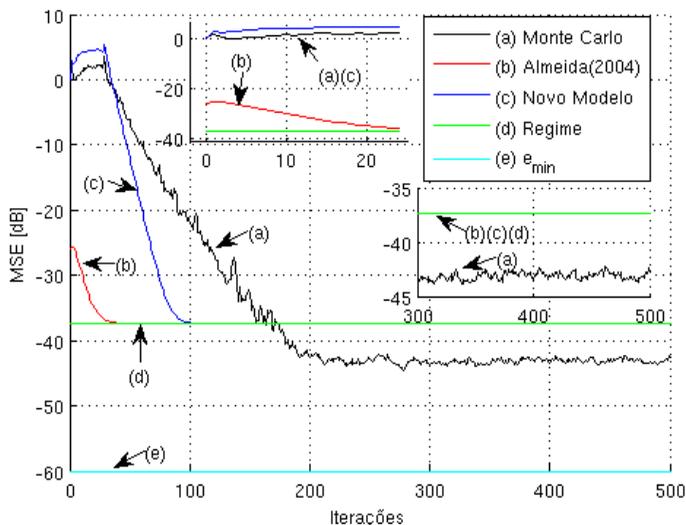


Figura 25 – Erro quadrático médio para o Exemplo 6.4.6. AR(30); $N = 40$; $P = 30$. (a) Simulação Monte Carlo (preto); (b) modelo em Almeida (2004) (vermelho); (c) modelo proposto em (5.35) (azul); (d) regime permanente segundo (5.98); e (e) erro mínimo.

As Figuras 24 e 25, em virtude do relaxamento da hipótese **H1** ($N \gg P$), apresentam uma degradação significativa na acurácia do modelo proposto em relação à simulação Monte Carlo. Esse comportamento é observado tanto no início do transiente, quanto no regime permanente. Entretanto, ainda com previsões mais próximas que as do modelo de Almeida (2004).

Exaustivas simulações indicam que a imprecisão na aproximação da matriz $\mathbf{R}_{\hat{\mu}}(n)$, apresentada no Apêndice B, não acarreta degradação perceptível na modelagem do erro quadrático médio, como verificado no comportamento médio dos coeficientes.

De acordo com os exemplos realizados e à luz dos modelos teóricos verificou-se que dois fatores realçam consideravelmente o comportamento gerado pelo Efeito de Inicialização no pseudo algoritmo de projeções afins: (a) entrada de excitação $u(n)$ altamente correlacionada; (b) inicialização do vetor de coeficientes longe da solução ótima da planta. Entretanto, a ordem do algoritmo, P , influencia o tamanho da matriz $\mathbf{U}(n)$, e conseqüentemente afeta o efeito também.

Finalmente, nota-se através da Figura 20 que há um mecanismo na equação do erro quadrático médio que ainda não é perfeitamente modelado pela equação proposta, compreendido entre as iterações P e N .

6.5 Conclusão

Neste capítulo foram apresentadas simulações que comprovam a disparidade no comportamento dos algoritmos APA e PAP, mesmo sob a restrição de passo unitário, e que essa diferença possui mais visibilidade em determinadas combinações de parâmetros.

Através dos resultados obtidos foi verificado que os modelos propostos apresentam um incremento significativo (chegando a 22 dB) na acurácia das predições em relação aos mesmos modelos de Almeida (2004), quando o Efeito de Inicialização é significativo. Em regime permanente os resultados obtidos são idênticos aos de Almeida (2004).

O *toolbox* Matlab utilizado para gerar as figuras dessa seção, encontra-se disponível em (BARCELOS, 2013). Os *scripts* “PAPie.m” e “PAPie2.m” foram utilizados para gerar as figuras da Seção 6.2, enquanto demais figuras foram gerados por “simul.m” ou versões alteradas deste *script*.

7 CONCLUSÕES E TRABALHO FUTURO

Neste trabalho foi realizado um estudo sobre o comportamento do pseudo algoritmo de projeções afins, aplicado a um problema de identificação de sistema, e restrito ao caso de passo de adaptação unitário e entrada autoregressiva e a estrutura do filtro adaptativo em linha de retardo. A motivação desse estudo foi a existência de um efeito não considerado na dinâmica inicial do regime transitório dos modelos anteriormente descritos, e que diferencia o comportamento entre o APA convencional e o PAP, erroneamente descritos na literatura como sendo equivalentes para passo unitário. Constatou-se que essa condição é necessária, porém não é suficiente para garantir a equivalência entre ambos os algoritmos. O mecanismo gerador dessa diferença foi denominado no presente trabalho de Efeito de Inicialização, sendo caracterizado pela estrutura escalar do erro na atualização dos coeficientes do PAP e uma inicialização arbitrária de $\mathbf{w}(0)$.

A expressão matemática (3.30), resultante desse estudo, permitiu gerar novos modelos analíticos para o PAP. Dessa forma, modelos determinísticos recursivos foram obtidos para o comportamento médio dos coeficientes, o erro quadrático médio e os momentos de segunda ordem do vetor de erro dos coeficientes. Também foram obtidas equações fechadas, que permitiram uma análise mais refinada do mecanismo intrínseco do efeito e da memória decorrente da recursividade do algoritmo. Constatou-se que o modelo proposto, assim como o de Almeida (2004), é um estimador não polarizado para o vetor de erro dos coeficientes. Além disso, em regime permanente, ambos os modelos possuem comportamento similar, visto que o efeito gerado pelos novos termos e pela memória recursiva se esvanece.

Os modelos analíticos propostos demonstraram uma excelente concordância com a simulação Monte Carlo, em situações que o modelo apresentado em Almeida et al. (2005) não possui uma predição tão acurada. Deve-se isso à ausência de $\kappa(0)$ no PAP, expresso pela equação (3.49) e, portanto, ocorrendo nas situações em que: (a) entrada de excitação é altamente correlacionada e (b) a inicialização do vetor de coeficientes é realizada longe da solução ótima. Os resultados obtidos pelos modelos propostos demonstraram maior precisão no transiente em relação ao modelo de Almeida et al. (2005), e desempenho similar em regime permanente.

7.1 Propostas para Continuação do Trabalho

Alguns pontos são sugeridos para continuação da pesquisa:

- **Passo não-unitário:** quando o PAP trabalha com passo diferente do unitário, a projeção de atualização ocorre de forma não ortogonal. Essa situação, embora ocasione uma diminuição da velocidade convergência, é de grande interesse em aplicações práticas visando garantir a estabilidade do algoritmo;
- **Ambiente não-estacionário:** uma característica fundamental do filtro adaptativo é sua capacidade de acompanhar modificações estatísticas dos sinais de entrada. Neste sentido, uma análise semelhante à realizada por Almeida (2004) faz-se necessária, para avaliar a capacidade de rastreamento do PAP;
- **Ordem e tamanho deficiente:** aplicações práticas geralmente apresentam restrições significativas em relação à capacidade computacional disponível. Informações sobre o desempenho do PAP em situações em que P é menor que a ordem do AR, e N é inferior ao comprimento da resposta impulsiva da planta, são de grande interesse para o projetista;
- **Estabilidade e robustez:** assim como em Rupp (2011), os resultados deste trabalho podem ser utilizados não apenas para melhorar modelos teóricos previamente desenvolvidos, como também para a análise de robustez e estabilidade;
- **Modelo estocástico para o APA:** o equacionamento obtido na Seção 3.4 e os resultados da Seção 6.2 indicam a possibilidade de extensão no modelo de Almeida (2004), de forma a possibilitar a modelagem do algoritmo APA convencional;
- **Algoritmo APA de baixo custo computacional:** assumindo a invariância no tempo da planta, o resultado obtido na Seção 3.4 indica o desenvolvimento de um novo algoritmo de baixo custo computacional, visto que a equação (3.48) apresenta o mesmo comportamento do APA, e o mesmo custo computacional do PAP após a primeira iteração.

REFERÊNCIAS

- ALMEIDA, S. J. M. **Análise Estatística do Comportamento de uma Classe de Algoritmos de Projeções Afins**. 2004. 180p. Tese (Doutorado em Engenharia Elétrica). Programa de Pós-Graduação em Engenharia Elétrica, UFSC, Florianópolis, 2004.
- ALMEIDA, S. J. M. et al. A Statistical Analysis of the Affine Projection Algorithm for Unity Step Size and Autoregressive Inputs. **IEEE Transactions on Circuits and Systems – I: Regular Papers**, v. 52, p. 1394-1405, 2005.
- ANTONIOU, A. **Digital Signal Processing - Signals, Systems, and Filters**. New York: McGraw-Hill, 2005.
- BARCELOS, R. B. Toolbox de simulação do efeito de inicialização e modelo do PAP. Versão 1.1.19. Disponível em: <https://docs.google.com/file/d/0B3iVEfiTjNt4QkI1LXdQb0Jlanc/edit?usp=sharing> (url reduzida: <http://goo.gl/sHhJ0>). Florianópolis: [s.n.]. 2013.
- BERSHAD, N. J.; CELKA, P.; VESIN, J.-M. Stochastic analysis of gradient adaptive identification of nonlinear systems with memory for gaussian data and noisy input and output measurements. **IEEE Transactions on Circuits and Systems**, vol. 47, n. 3, 1999. 675-689.
- BERSHAD, N. J.; LINEBARGER, D.; MCLAUGHLIN, S. A stochastic analysis of the affine projection algorithm for gaussian autoregressive inputs. **Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing**, 2001. 3837-3840.
- BOUTEILLE, F.; SCALART, P.; CORAZZA, M. Pseudo affine projection algorithm: new solution for adaptive identification. In: Proc. Eurospeech. Budapest: [s.n.]. 1999. p. 427-430.
- COSTA, M. H. **Comportamento do algoritmo FXLMS em sistemas com caminho secundário não-linear**: aplicação ao controle ativo de ruído. 2001. Tese (Doutorado em Engenharia Elétrica). Engenharia, Universidade Federal de Santa Catarina (UFSC), Florianópolis, 2001.
- COSTA, M. H. **Avaliação da Função de Correlação de um Sinal Autoregressivo**. 2011. 5p. Relatório interno. Universidade Federal de Santa Catarina, Florianópolis, 2011.
- COSTA, M. H. et al. New Insights into the Weight Behaviour of the Affine Projection Algorithm. In: Proc. European Signal Process. Conf. Bucharest: [s.n.]. 2012a. p. 2610-2614.

COSTA, M. H. et al. Efeito de Inicialização no Modelo do Comportamento dos Coeficientes do Algoritmo de Projeções Afins. In: XXX Simpósio Brasileiro de Telecomunicações (SBrT'12). Brasília: [s.n.]. 2012b. p. xx-xx.

DINIZ, P. S. R. **Adaptive Filtering: Algorithms and Practical Implementation**. 3. ed. New York: Springer, 2008.

FARHANG-BOROUJENY, B. **Adaptive Filters: Theory and Applications**. 1st. ed. Chichester: Wiley, 1999.

GAY, S. L.; TAVATHIA, S. The fast affine projection algorithm. In: Proc IEEE Int. Conf. Acoust., Speech, Signal Process. Detroit: [s.n.]. Maio 1995. p. 3023–3026.

GONZALEZ, A. et al. Affine Projection Algorithms: Evolution to Smart and Fast Algorithms and Applications. In: Proc. European Signal Process. Conf. Bucharest: [s.n.]. 2012. p. 1965-1969.

HAYKIN, S. **Adaptive Filter Theory**. 3rd. ed. Englewood Cliffs: Prentice-Hall, 1996.

HAYKIN, S. **Adaptive Filter Theory**. 4nd. ed. Englewood Cliffs: Prentice-Hall, 2002.

KAY, S. M. **Intuitive Probability and Random Processes Using Matlab**. New York: Springer, 2006.

MANOLAKIS, G.; INGLE, V. K.; KOGON, S. M. **Statistical and adaptive signal processing - spectral estimation, signal modeling, adaptive filtering and array processing**. 1ª. ed. Norwood: Artech House, 2005.

MAZO, J. E. On the independence theory of equalizer convergence. **Bell Syst. Tech. Journal**, vol. 58, Maio-Junho 1979. 963-993.

MONTGOMERY, D. C.; RUNGER, G. C. **Estatística Aplicada e Probabilidade para Engenheiros**. 2. ed. Rio de Janeiro: LTC, 1999.

NICOLAU, V. **Cancelamento de Realimentação Acústica em Aparelhos Auditivos através do Método de Erro de Predição com Comprimento Deficiente**. 2010. Dissertação (Mestrado em Engenharia Elétrica). Universidade Federal de Santa Catarina, Florianópolis, Anexo B, 2010.

OPPENHEIM, A. V.; SCHAEFER, R. W.; BUCK, J. R. **Discrete-Time Signal Processing**. 2nd. ed. New Jersey: Prentice Hall, 1998.

OZEKI, K.; UMEDA, T. An adaptive filtering algorithm using orthogonal projection to an affine subspace and its properties. **Electron. Commun. Jpn.**, v. 67-A, n. 5, p. 19-27, 1984.

PAPOULIS, A.; PILLAI, S. U. **Probability, Random Variables and Stochastic Processes**. 4nd. ed. New York: McGraw-Hill, 2002.

RECOMMENDATION ITU-T G.168 International telephone connections and circuits – Apparatus associated with long-distance telephone circuits. [S.l.]: [s.n.]. 02/2002.

RUPP, M. A Family of Filter Algorithms with Decorrelating Properties. **IEEE Transactions on Signal Processing**, vol. 46, n. 3, 1998. 771-775.

RUPP, M. Pseudo-affine projection algorithms revis-ited: robustness and stability analysis. **IEEE Transactions on Signal Processing**, vol. 59, no. 5, Maio 2011. 2017–2023.

SAMSON, C.; REDDY, V. U. Fixed Point Error Analysis of the Normalized Ladder Algorithms. **IEEE Transactions on Acoustics, Speech and Signal Processing**, v. 31, n. 5, Outubro 1983. 1177-1191.

SAYED, A. H. **Fundamentals of Adaptive Filtering**. 1st. ed. New Jersey: Wiley, 2003.

SHENOI, B. A. **Introduction to Digital Signal Processing and Filter Design**. New Jersey: John Wiley & Sons, 2006.

SHIN, H.-C.; SAYED, A. H. Mean-Square Performance of a Family of Affine Projection Algorithms. **IEEE Transactions on Signal Processing**, 52, n. 1, Janeiro 2004. 90-102.

WIDROW, B.; STEARNS, S. D. **Adaptive Signal Processing**. Englewood Cliffs: Prentice-Hall, 1985.

ZWILLINGER, D. **Standard Mathematical Tables and Formulae**. 31st. ed. Boca Raton: CRC, 2003.

APÊNDICE A – ESTIMAÇÃO DOS PARÂMETROS DE UM PROCESSO AUTOREGRESSIVO ATRAVÉS DE PREDITOR LINEAR PELO MÉTODO DOS MÍNIMOS QUADRADOS

O método dos mínimos quadrados (também conhecido como método de regressão linear) para formulação de um problema de estimação foi introduzido em 1821, pelo matemático alemão Carl Friedrich Gauss (1777-1855), visando determinar a órbita do asteroide Ceres (MANOLAKIS; INGLE e KOGON, 2005, p. 395). A análise obtida nesse apêndice para a estimação dos parâmetros de um processo autoregressivo através de um preditor linear guarda similaridade com a estimação para o problema de filtragem em Manolakis; Ingle e Kogon (2005, p. 395) e Haykin (1996, p. 506). Esta abordagem desempenha um papel fundamental no mecanismo de atualização dos coeficientes do PAP, conforme Subseção 2.3.2.

ESTIMADOR DOS MÍNIMOS QUADRADOS LINEAR

A estimação dos mínimos quadrados (EMQ, em inglês *Least Square Estimate* - LSE) utiliza como critério de desempenho a minimização da soma dos quadrados dos desvios (erro de estimação) das observações em relação ao modelo de regressão, operando em blocos de tamanho P . Considerando-se um vetor $\hat{\mathbf{u}}(n) = [\hat{u}(n) \hat{u}(n-1) \dots \hat{u}(n-N+1)]^T$ e o preditor linear de (9.2.2) em Manolakis; Ingle e Kogon (2005, p. 450),

$$\hat{u}(n) = \sum_{k=1}^P \mathbf{u}(n-k) \hat{a}_k(n) \quad (\text{A.1})$$

tem-se que P amostras de $\hat{u}(n)$, na forma vetorial, resulta em

$$\hat{\mathbf{u}}(n) = \mathbf{U}(n) \hat{\mathbf{a}}(n) \quad (\text{A.2})$$

sendo que $\mathbf{U}(n) = [\mathbf{u}(n-1) \mathbf{u}(n-2) \dots \mathbf{u}(n-P)]$ é uma matriz composta de P regressores passados de dimensões $N \times P$ e o vetor $\hat{\mathbf{a}}(n) = [\hat{a}_1(n) \hat{a}_2(n) \dots \hat{a}_P(n)]^T$, com dimensões $P \times 1$, é a estimação dos mínimos quadrados do vetor de coeficientes \mathbf{a} . O vetor de erro de predição é fornecido pela equação:

$$\boldsymbol{\phi}(n) = \mathbf{u}(n) - \mathbf{U}(n) \hat{\mathbf{a}}(n) \quad (\text{A.3})$$

em que o vetor $\boldsymbol{\phi}(n) = [\phi_0(n) \phi_1(n) \dots \phi_{N-1}(n)]^T$. Farhang-Boroujeny cita que para o problema de filtragem,

[...] uma interpretação perspicaz para o método dos mínimos quadrados é sua propriedade de *ajuste de curva* (*curve-fitting*). Considere uma curva na qual os pontos são as amostras da saída desejada do filtro adaptativo. Dessa forma, amostras da saída do filtro (dada uma sequência de entrada) podem ser consideradas para formar outra curva. O problema na escolha dos parâmetros do filtro, de forma a encontrar o melhor ajuste entre as duas curvas, se resume ao método dos mínimos quadrados, caso seja definido que o melhor ajuste é o que minimiza uma soma ponderada dos valores quadráticos da diferença entre as amostras da curva. (FARHANG-BOROUJENY, 1999, p.413, tradução nossa).

Dessa forma, o autor resume o procedimento do estimador dos mínimos quadrados para o problema de filtragem. De forma análoga, os coeficientes do estimador para um processo AR podem ser obtidos através da minimização do somatório dos quadrados do erro de predição, ou seja, a energia do sinal de erro, dada por

$$\xi(n) = \boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n). \quad (\text{A.4})$$

DERIVAÇÃO GEOMÉTRICA DAS EQUAÇÕES NORMAIS

A minimização da equação (A.4) em função dos parâmetros \mathbf{a} , resulta nas *equações normais*. Essa abordagem é semelhante à empregada na obtenção das equações de Wiener-Hopf pelo critério do estimador de mínimo erro quadrático médio. A solução do problema de estimação dos mínimos quadrados (EMQ) para o problema de filtragem é apresentada na Seção 8.2.1 de Manolakis; Ingle e Kogon (2005).

Todavia, as equações normais podem ser obtidas através de interpretação geométrica, como apresentado a seguir.

Considerando-se um espaço vetorial, como o espaço de Hilbert, formado por vetores com produto interno

$$\langle \mathbf{u}_i(n), \mathbf{u}_j(n) \rangle \triangleq \mathbf{u}_i^T(n)\mathbf{u}_j(n) \quad (\text{A.5})$$

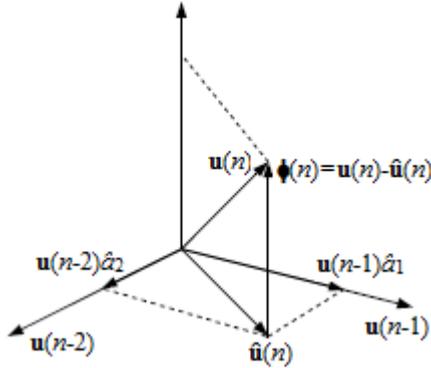


Figura 26 – Interpretação vetorial da estimação pelo EMQ para $N = 3$ e $M = 2$ (Baseada na figura 8.5 de (MANOLAKIS; INGLE e KOGON, 2005)).

sendo i e j , o i -ésimo e o j -ésimo elemento de $\mathbf{U}(n)$, respectivamente, e com tamanho,

$$\|\mathbf{u}_i\|^2 \triangleq \langle \mathbf{u}_i, \mathbf{u}_i \rangle = \mathbf{u}_i^T \mathbf{u}_i \quad (\text{A.6})$$

o preditor linear de (A.2) gera um espaço vetorial N -dimensional e os P vetores $\mathbf{u}(n-i)$ formam um subespaço gerado pelas colunas da matriz de dados $\mathbf{U}(n)$.

Observa-se pela Figura 26 que o erro de estimação entre $\mathbf{u}(n)$ e $\hat{\mathbf{u}}(n)$ é o vetor $\boldsymbol{\phi}(n)$, sendo ortogonal ao subespaço gerado pelas colunas da matriz de dados $\mathbf{U}(n)$, quando $\boldsymbol{\phi}(n)$ for minimizado. Embora, qualquer estimativa $\hat{\mathbf{u}}(n)$ sempre esteja contida nesse subespaço, $\mathbf{u}(n)$ só estará quando $\boldsymbol{\phi}(n) = \mathbf{0}_{N \times 1}$.

O *princípio da ortogonalidade* é empregado para se obter o estimador dos mínimos quadrados linear. Enquanto no caso geral do filtro ótimo de Wiener, o princípio é utilizado com sentido estocástico de mínimo erro quadrático médio, para o EMQ resulta em uma média temporal baseada na covariância (HAYKIN, 1996, p. 487). Dessa forma, ele define que o k -ésimo componente do vetor de gradiente $\nabla_k \xi(n)$ é a derivada da função de custo em relação ao vetor de coeficientes a_k , resultando em

$$\nabla_k \xi(n) = \frac{\partial \xi(n)}{\partial a_k} \quad (\text{A.7})$$

em que $k = 1, \dots, M$. Substituindo-se (A.4) na (A.7), tem-se que

$$\nabla_k \xi(n) = 2 \sum_{l=0}^{N-1} \left(\phi_l(l) \frac{\partial \phi_l(l)}{\partial a_k} \right). \quad (\text{A.8})$$

Diferenciando a equação do erro de estimação com relação à $a_k(n)$, tem-se a expressão:

$$\frac{\partial \phi_l(n)}{\partial a_k} = -u(n-k) \quad (\text{A.9})$$

para $k = 1, \dots, P$. Substituindo-se (A.9) em (A.8), obtém-se

$$\nabla_k \xi(n) = -2 \sum_{l=0}^{N-1} (\phi_l(l) u(l-k)). \quad (\text{A.10})$$

Minimizando a função de custo com relação ao vetor de coeficientes \mathbf{a} , é necessário que a seguinte condição seja satisfeita,

$$\nabla_k \xi(n) = \mathbf{0}_{P \times 1} \quad (\text{A.11})$$

resultando em ortogonalidade vetorial, similar ao princípio da ortogonalidade obtida na solução de Wiener,

$$\mathbf{U}^T(n) \boldsymbol{\phi}(n) = \mathbf{0}_{P \times 1}. \quad (\text{A.12})$$

Assim, tem-se que o erro de estimação é ortogonal ao subespaço gerado pelos vetores coluna de $\mathbf{U}(n)$, quando operando na condição dos mínimos quadrados, de acordo com a Figura 26.

Substituindo-se a equação (A.3) em (A.12), leva à

$$\mathbf{U}^T(n) [\mathbf{u}(n) - \mathbf{U}(n) \hat{\mathbf{a}}(n)] = \mathbf{0}_{P \times 1} \quad (\text{A.13})$$

resultando nas equações normais para o estimador dos mínimos quadrados do vetor de coeficientes \mathbf{a} , conforme expressão

$$\mathbf{U}^T(n) \mathbf{U}(n) \hat{\mathbf{a}}(n) = \mathbf{U}^T(n) \mathbf{u}(n) \quad (\text{A.14})$$

Pré-multiplicando ambos os lados de (A.14) por $[\mathbf{U}^T(n) \mathbf{U}(n)]^{-1}$, tem-se que

$$\hat{\mathbf{a}}(n) = [\mathbf{U}^T(n) \mathbf{U}(n)]^{-1} \mathbf{U}^T(n) \mathbf{u}(n). \quad (\text{A.15})$$

Considerando o caso em que $N \neq P$, tem-se que $[\mathbf{U}^T(n) \mathbf{U}(n)]^{-1} \mathbf{U}^T(n)$ é a Pseudo Inversa de Moore Penrose (OZEKI e UMEDA, 1984). A solução das equações normais existe, e é única, se a matriz $\mathbf{U}^T(n) \mathbf{U}(n)$ for não singular. Finalmente, aplicando o teorema de Pitágoras no triângulo retângulo da Figura 26, tem-se a seguinte relação entre o preditor linear $\hat{\mathbf{u}}(n)$ e o vetor de erro $\boldsymbol{\phi}(n)$:

$$\|\mathbf{u}(n)\|^2 = \|\hat{\mathbf{u}}(n)\|^2 + \|\boldsymbol{\phi}(n)\|^2. \quad (\text{A.16})$$

OPERADOR DE PROJEÇÃO

A estimativa pelos mínimos quadrados $\hat{\mathbf{u}}(n)$ de $\mathbf{u}(n)$ pode ser obtida através do operador de projeção. Segundo Manolakis (MANOLAKIS; INGLE e KOGON, 2005, p. 402), a estimativa dos mínimos quadrados de $\mathbf{u}(n)$ pode ser obtida através das equações (A.2) e (A.15), resultando em

$$\hat{\mathbf{u}}(n) = \mathbf{U}(n)\hat{\mathbf{a}}(n) = \mathbf{P}_U(n)\mathbf{u}(n) \quad (\text{A.17})$$

em que $\mathbf{P}_U(n)$ é uma matriz $N \times N$ de projeção no subespaço gerado pelas colunas de $\mathbf{U}(n)$, e definida pela expressão

$$\mathbf{P}_U(n) = \mathbf{U}(n) \left[\mathbf{U}^T(n)\mathbf{U}(n) \right]^{-1} \mathbf{U}^T(n) \quad (\text{A.18})$$

sendo simétrica e idempotente, ou seja,

$$\mathbf{P}_U(n) = \mathbf{P}_U^T(n) \quad (\text{A.19})$$

$$\mathbf{P}_U^2(n) = \mathbf{P}_U^T(n)\mathbf{P}_U(n) = \mathbf{P}_U(n). \quad (\text{A.20})$$

Dessa forma, o vetor de erro de estimação pelos mínimos quadrados pode ser expresso por

$$\begin{aligned} \boldsymbol{\phi}(n) &= \mathbf{u}(n) - \mathbf{P}_U(n)\mathbf{u}(n) \\ &= (\mathbf{I}_N - \mathbf{P}_U(n))\mathbf{u}(n) \\ &= \mathbf{P}_\perp(n)\mathbf{u}(n) \end{aligned} \quad (\text{A.21})$$

em que $\mathbf{P}_\perp(n) = \mathbf{I}_N - \mathbf{P}_U(n)$ é o operador de projeção complementar ortogonal ao subespaço gerado pelas colunas de $\mathbf{U}(n)$, e \mathbf{I}_N é uma matriz identidade de dimensões $N \times N$.

Esse erro na estimação de $\mathbf{u}(n)$ pode ser geometricamente analisado através da Figura 26, onde constata-se que:

1. O vetor $\hat{\mathbf{u}}(n)$ é obtido como uma combinação linear dos vetores $\mathbf{u}(n-1)$, ..., $\mathbf{u}(n-N+1)$;
2. A diferença $\boldsymbol{\phi}(n) = \mathbf{u}(n) - \hat{\mathbf{u}}(n)$ é um vetor ortogonal ao subespaço gerado por $\mathbf{U}(n)$;
3. O vetor $\hat{\mathbf{u}}(n)$ tem a menor distância Euclidiana em relação ao vetor $\mathbf{u}(n)$.

PROPRIEDADES DESTA ANÁLISE

Quatro propriedades importantes podem ser derivadas de Haykin (1996, p. 502) e Manolakis; Ingle e Kogoon (2005, p. 403):

1. O estimador dos mínimos quadrados $\hat{\mathbf{a}}(n)$ não é polarizado, desde que o vetor de erro de medida $\boldsymbol{\phi}(n)$ tenha média zero, ou seja, $E\{\hat{\mathbf{a}}(n)\} = \mathbf{a}$;
2. Quando $\boldsymbol{\phi}(n)$ é branco e de média zero, a matriz de correlação do vetor de erro $\boldsymbol{\phi}(n)$ tem suas amostras descorrelacionadas, com variância constante na diagonal dada por $\sigma_{\phi}^2 \mathbf{I}_N$;
3. Quando o processo de erro de medida $\boldsymbol{\phi}(n)$ é branco e com média zero, a estimativa dos mínimos quadrados $\hat{\mathbf{a}}(n)$ é o melhor estimador linear não polarizado (BLUE);
4. Quando o processo de erro de medida é branco e Gaussiano, com média zero, o estimador dos mínimos quadrados $\hat{\mathbf{a}}(n)$ atinge o limite mínimo de Cramér-Rao para estimativas não polarizadas;
5. Pela análise de ortogonalidade vetorial similar ao princípio da ortogonalidade, não há informação de $\boldsymbol{\phi}(n)$ contida em $\mathbf{U}(n)$.

As demonstrações são similares às obtidas nas referências citadas para o problema de filtragem.

APÊNDICE B – MATRIZ DE CORRELAÇÃO $\mathbf{R}_{\hat{\phi}\hat{a}}(n)$

Este apêndice trata da determinação analítica da matriz de correlação cruzada $\mathbf{R}_{\hat{\phi}\hat{a}}(n)$, definida em (4.7) como:

$$\mathbf{R}_{\hat{\phi}\hat{a}}(n) \triangleq E \left\{ \boldsymbol{\phi}(n) \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{U}^T(n) \right\}.$$

Definindo-se uma matriz determinística de dimensões $N \times N$,

$$\boldsymbol{\Gamma}(n) \triangleq \boldsymbol{\phi}(n) \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{U}^T(n) \quad (\text{B.1})$$

e substituindo-se $\boldsymbol{\phi}(n) = \mathbf{u}(n) - \mathbf{U}(n) \hat{\mathbf{a}}(n)$, tem-se

$$\begin{aligned} \boldsymbol{\Gamma}(n) &= \mathbf{u}(n) \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{U}^T(n) \\ &\quad - \mathbf{U}(n) \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{U}^T(n). \end{aligned} \quad (\text{B.2})$$

Desprezando-se as flutuações de $\hat{\mathbf{a}}(n)$ em relação ao vetor de coeficientes \mathbf{a} , conforme hipótese **H7**, a equação (B.2) pode ser aproximada por:

$$\begin{aligned} \boldsymbol{\Gamma}(n) &\cong \mathbf{u}(n) \mathbf{a}^T \mathbf{Q}(-n+1) \mathbf{U}^T(n) \\ &\quad - \mathbf{U}(n) \mathbf{a} \mathbf{a}^T \mathbf{Q}(-n+1) \mathbf{U}^T(n). \end{aligned} \quad (\text{B.3})$$

Simulações estatísticas extensivamente realizadas corroboram a validade da aproximação apresentada em (B.3). Uma equação determinística e conformável (ZWILLINGER, 2003) pode ser obtida para a matriz $\mathbf{R}_{\hat{\phi}\hat{a}}(n)$ equacionando-se os elementos da matriz $\boldsymbol{\Gamma}(n)$,

$$\begin{aligned} \eta_{l,c}(n) &\cong u(n-l) \sum_{j=1}^P a_j u_{-1}(-n+j) u(n-j-c+1) \\ &\quad - \sum_{i=1}^P u(n-i-l+1) a_i \sum_{j=1}^P a_j u_{-1}(-n+j) u(n-j-c+1) \end{aligned} \quad (\text{B.4})$$

em que $l = 1, \dots, N$ é o índice de linha e $c = 1, \dots, N$ é o índice de coluna, e $u_{-1}(k)$ é uma função degrau unitário (1 se $k \geq 0$ e 0 para outros valores de k). Tomando-se o valor esperado de (B.4) obtém-se a equação para cada elemento da matriz $\mathbf{R}_{\hat{\phi}\hat{a}}(n)$, dada por:

$$\begin{aligned} r_{\hat{\phi}\hat{a}}(l,c)(n) &\cong E \left\{ u(n-l) \sum_{j=1}^P a_j u_{-1}(-n+j) u(n-j-c+1) \right\} \\ &\quad - E \left\{ \sum_{i=1}^P u(n-i-l+1) a_i \sum_{j=1}^P a_j u_{-1}(-n+j) u(n-j-c+1) \right\}. \end{aligned} \quad (\text{B.5})$$

Considerando-se que os termos de $u_{-1}(-n+j)$ e de \mathbf{a} são determinísticos, chega-se à

$$\begin{aligned}
r_{\phi\bar{u}(l,c)}(n) &\cong \sum_{j=1}^P E \{u(n-l)u(n-j-c)\} a_j u_{-1}(-n+j) \\
&- \sum_{i=1}^P \sum_{j=1}^P E \{u(n-i-l)u(n-j-c)\} a_i a_j u_{-1}(-n+j).
\end{aligned} \tag{B.6}$$

Os valores esperados em (B.6) são obtidos da matriz de correlação de dimensões $(N+P) \times (N+P)$ do sinal de entrada, ou seja,

$$r_{uu(a,b)} \triangleq E \{u(n-a)u(n-b)\}. \tag{B.7}$$

Assim, (B.6) resulta em

$$\begin{aligned}
r_{\phi\bar{u}(l,c)}(n) &\cong \sum_{j=1}^P r_{uu(l,j+c)} a_j u_{-1}(-n+j) \\
&- \sum_{i=1}^P \sum_{j=1}^P r_{uu(i+l,j+c)} a_i a_j u_{-1}(-n+j).
\end{aligned} \tag{B.8}$$

A equação (B.8) apresenta comportamento variante no tempo. A expressão determinística para a função de correlação de um sinal real autoregressivo em função dos coeficientes do modelo e da potência da inovação é apresentada no Anexo D.

APÊNDICE C – MATRIZ DE CORRELAÇÃO $\mathbf{R}_{\hat{u}}(n)$

Este apêndice apresenta a derivação de uma equação determinística para a matriz $\mathbf{R}_{\hat{u}}(n)$, definida em (5.33) como:

$$\mathbf{R}_{\hat{u}}(n) \triangleq E \left\{ \mathbf{U}(n) \mathbf{Q}^T (-n+1) \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{U}^T(n) \right\}.$$

Baseado no mesmo procedimento descrito no Apêndice B, obtém-se a descrição determinística para essa matriz. Inicialmente, define-se uma matriz determinística de dimensões $N \times N$,

$$\mathbf{\Omega}(n) \triangleq \mathbf{U}(n) \mathbf{Q}^T (-n+1) \hat{\mathbf{a}}(n) \hat{\mathbf{a}}^T(n) \mathbf{Q}(-n+1) \mathbf{U}^T(n). \quad (\text{C.1})$$

Desprezando-se as flutuações de $\hat{\mathbf{a}}(n)$ em relação ao vetor de coeficientes \mathbf{a} , conforme hipótese **H7**, a equação (C.1) pode ser aproximada por:

$$\mathbf{\Omega}(n) \cong \mathbf{U}(n) \mathbf{Q}^T (-n+1) \mathbf{a} \mathbf{a}^T \mathbf{Q}(-n+1) \mathbf{U}^T(n). \quad (\text{C.2})$$

Uma equação determinística e conformável (ZWILLINGER, 2003) pode ser obtida para a matriz $\mathbf{R}_{\hat{u}}(n)$ através dos elementos da matriz $\mathbf{\Omega}(n)$,

$$\begin{aligned} \omega_{l,c}(n) &= \sum_{i=1}^P u(n-i-l+1) u_{-1}(-n+i) a_i \sum_{j=1}^P a_j u_{-1}(-n+j) u(n-j-c+1) \\ &= \sum_{i=1}^P \sum_{j=1}^P u(n-i-l) u(n-j-c) a_i a_j u_{-1}(-n+i) u_{-1}(-n+j) \end{aligned} \quad (\text{C.3})$$

sendo que $l = 1, \dots, N$ e $c = 1, \dots, N$ representam os índices de cada linha e coluna da matriz, respectivamente. Tomando-se o valor esperado (C.3), resulta em

$$r_{\hat{u}(l,c)}(n) = \sum_{i=1}^P \sum_{j=1}^P r_{uu(i+l, j+c)} a_i a_j q_{-1}(-n+i) q_{-1}(-n+j) \quad (\text{C.4})$$

em que $r_{uu(a,b)}$ é definido em (B.7),

$$r_{uu(a,b)} \triangleq E \left\{ u(n-a) u(n-b) \right\}.$$

No Anexo D é apresentada uma expressão determinística para a função de correlação de um sinal real autoregressivo.

ANEXO A – INTERPRETAÇÃO GEOMÉTRICA DO APA

Neste anexo é apresentada a interpretação geométrica do APA derivada em (ALMEIDA, 2004, p. 38) com base em (OZEKI e UMEDA, 1984). Com o objetivo de tornar a explanação mais clara, esboça-se primeiro o algoritmo NLMS que, como já foi visto, é um caso particular do algoritmo APA utilizando-se apenas um vetor de dados de entrada, conforme mostra a equação (2.39).

Na Figura 27 apresenta-se uma interpretação geométrica do comportamento de atualização do vetor de coeficientes $\mathbf{w}(n)$ no algoritmo NLMS, mostrando seu traçado até atingir o valor ótimo \mathbf{w}^o .

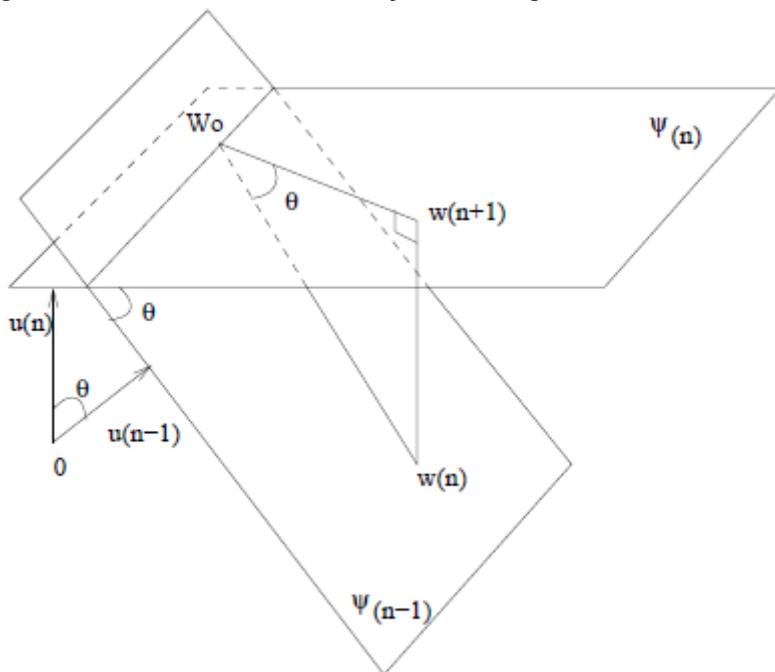


Figura 27 – Interpretação geométrica do algoritmo NLMS (Figura 2.2 de (ALMEIDA, 2004)).

O conjunto definido por $\Psi(n)$ representa o conjunto de todos os vetores de coeficientes que produzem a saída $y(n)$ igual a $d(n)$ na topologia de estimação linear da Figura 4.b para todos os vetores de entrada $\mathbf{u}(n)$, os quais formam um hiperplano no espaço Euclidiano N -dimensional (OZEKI e UMEDA, 1984). Note que o vetor ótimo \mathbf{w}^o

pertence necessariamente a esse hiperplano. Da mesma forma, \mathbf{w}^o pertencerá a todos os hiperplanos $\psi(n-k)$ associados a cada vetor de entrada $\mathbf{u}(n-k)$. Assim, \mathbf{w}^o pertence à intersecção de todos os hiperplanos $\psi(n-k)$, $\forall k$. Formalmente, o hiperplano $\Psi(n)$ é definido por

$$\Psi(n-k) = \left\{ \mathbf{w}; \mathbf{w} \in \mathfrak{R}^n, \mathbf{u}^T(n-k)\mathbf{w} = d(n-k) \right\}. \quad (\text{D.1})$$

A equação de atualização de $\mathbf{w}(n)$ para passo unitário é dada por:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \Delta\mathbf{w}(n) \quad (\text{D.2})$$

em que $\Delta\mathbf{w}(n)$ é função da entrada $\mathbf{u}(n)$, e saídas $d(n)$ e $y(n)$. Assim,

$$\Delta\mathbf{w}(n) = f(u(n), \dots, u(n-M+1), d(n), \dots, d(n-M+1), y(n), \dots, y(n-M+1)) \quad (\text{D.3})$$

Como apresentado mediante ilustração na Figura 27, o algoritmo NLMS ajusta o vetor de coeficientes $\mathbf{w}(n)$ adicionando um fator de correção $\Delta\mathbf{w}(n)$. Como o ajuste de $\mathbf{w}(n)$ é feito de forma a minimizar a norma de $\Delta\mathbf{w}(n)$, a direção de $\Delta\mathbf{w}(n)$ deve ser ortogonal ao subespaço $\Psi(n)$ e $\Psi(n-1)$, o ângulo θ da Figura 27. Isto permite observar que, dado o vetor $\mathbf{w}(n)$, a velocidade de convergência estará diminuindo quando o ângulo θ aproximar-se de 0 ou π radianos, contrário ao caso de quando θ aproximar-se de $(\pi/2)$ ou $(3\pi/2)$ radianos. Através da Figura 27, verifica-se ainda que

$$\mathbf{w}(n+1) - \mathbf{w}(n) = \mathbf{w}(n) - \mathbf{w}^o \mid \cos \theta. \quad (\text{D.4})$$

Assim, uma redução do ângulo θ leva a uma maior distância entre $\mathbf{w}(n+1)$ e \mathbf{w}^o para o mesmo valor de diferença entre $\mathbf{w}(n)$ e \mathbf{w}^o . A situação é mostrada graficamente de perfil na Figura 28. Assim, a velocidade de convergência do algoritmo tende a aumentar quando o ângulo entre $\Psi(n)$ e $\Psi(n-1)$ se aproxima de $(\pi/2)$ ou $(3\pi/2)$ radianos.

Outro ponto importante é a relação entre o ângulo θ e o ângulo entre $\mathbf{u}(n-1)$ e $\mathbf{u}(n)$. Pela definição do hiperplano $\Psi(n)$ em (D.1), $\Psi(n-1)$ é caracterizado pela expressão $\mathbf{u}^T(n-1)\mathbf{w}(n) = d_1(n)$ e $\Psi(n)$ pela expressão $\mathbf{u}^T(n)\mathbf{w}(n+1) = d(n)$. Assim $\Psi(n-1)$ é o subespaço formado pelas extremidades de todos os vetores em \mathfrak{R}^N cuja projeção ortogonal em $\mathbf{u}(n-1)$ é igual ao mesmo valor $d_1(n)$. Assim, $\mathbf{u}(n-1)$ é perpendicular a $\Psi(n-1)$. Da mesma forma, $\mathbf{u}(n)$ é perpendicular a $\Psi(n)$. Logo o ângulo θ entre $\Psi(n)$ e $\Psi(n-1)$ é também o ângulo entre $\mathbf{u}(n)$ e $\mathbf{u}(n-1)$.

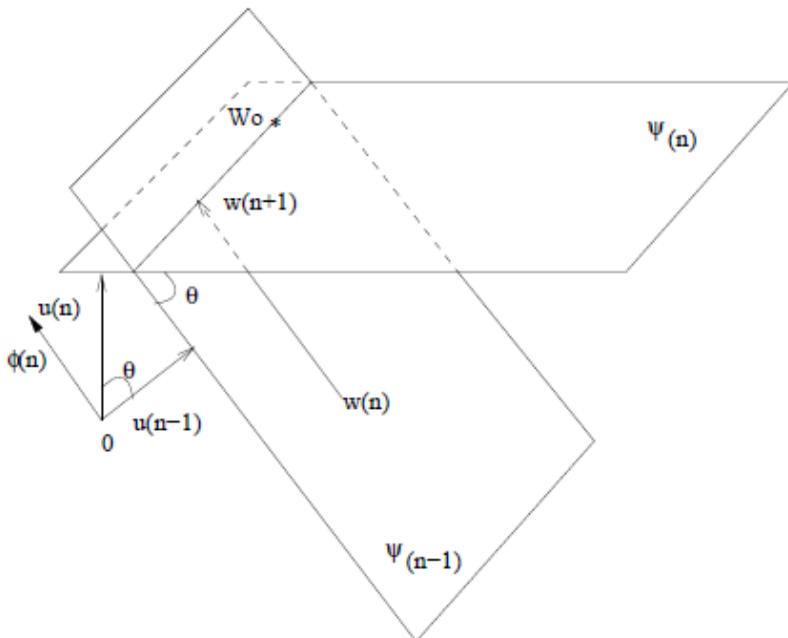


Figura 28 – Interpretação geométrica do novo algoritmo (Figura 2.3 de (ALMEIDA, 2004))

Este fenômeno é resultado da direção de ajuste do vetor de coeficientes ser função da direção imposta pelo vetor de entrada $\mathbf{u}(n)$. Para melhorar a situação, a direção de modificação do vetor de coeficientes precisa ser reconsiderada.

Baseado na Figura 27 fica evidente que para manter a velocidade de convergência independente do ângulo entre $\mathbf{u}(n)$ e $\mathbf{u}(n-1)$, a linha que vai de $\mathbf{w}(n)$ para $\mathbf{w}(n+1)$ deve ser traçada diretamente de $\mathbf{w}(n)$ para $\Psi(n) \cap \Psi(n-1)$ conforme a ilustração da Figura 28 e não perpendicularmente a $\Psi(n)$ como na Figura 27.

ANEXO B – DETERMINAÇÃO PARA $E\{y_1 y_2 y_3^T y_4\}$

A seguinte derivação é baseada no Anexo 4 de Costa (2001) e resulta na fatoração de $E\{y_1 y_2 y_3^T y_4\}$ em termos de valores esperados de no máximo duas variáveis aleatórias. Supõe-se que y_1 e y_4 sejam variáveis aleatórias escalares e \mathbf{y}_2 e \mathbf{y}_3 sejam vetores de variáveis aleatórias Gaussianas de média zero ($\mathbf{y}_i = [y_{i1} y_{i2} \dots y_{iN}]^T$ para $i = 2, 3$).

Arbitrando-se uma matriz \mathbf{B} como

$$\mathbf{B} = y_1 \mathbf{y}_2 \mathbf{y}_3^T y_4 \quad (\text{E.1})$$

e expressando-se (E.1) sob a forma escalar (ZWILLINGER, 1996), tem-se que

$$b_{i,j} = y_1 y_{2i} y_{3j} y_4 \quad (\text{E.2})$$

sendo que i representa a i -ésima linha e j a j -ésima coluna da matriz quadrada \mathbf{B} . Tomando-se o valor esperado de (E.2), tem-se que

$$E\{b_{i,j}\} = E\{y_1 y_{2i} y_{3j} y_4\}. \quad (\text{E.3})$$

O teorema de fatoração de variáveis Gaussianas (HAYKIN, 1996) estabelece a seguinte relação de valores esperados para variáveis reais Gaussianas:

$$\begin{aligned} E\{x_1 x_2 x_3 x_4\} &= E\{x_1 x_2\} E\{x_3 x_4\} + E\{x_1 x_3\} E\{x_2 x_4\} \\ &+ E\{x_1 x_4\} E\{x_2 x_3\}. \end{aligned} \quad (\text{E.4})$$

Substituindo (E.4) em (E.3) e rearranjando os termos resultantes, resulta em

$$\begin{aligned} E\{b_{i,j}\} &= E\{y_1 y_{2i}\} E\{y_{3j} y_4\} + E\{y_1 y_{3j}\} E\{y_{2i} y_4\} \\ &+ E\{y_1 y_4\} E\{y_{2i} y_{3j}\}. \end{aligned} \quad (\text{E.5})$$

Voltando à forma original, tem-se que

$$\begin{aligned} E\{y_1 \mathbf{y}_2 \mathbf{y}_3^T y_4\} &= E\{y_1 \mathbf{y}_2\} E\{\mathbf{y}_3^T y_4\} + E\{\mathbf{y}_2 y_4\} E\{y_1 \mathbf{y}_3^T\} \\ &+ E\{y_1 y_4\} E\{\mathbf{y}_2 \mathbf{y}_3^T\}. \end{aligned} \quad (\text{E.6})$$

ANEXO C – DETERMINAÇÃO PARA $E\{\mathbf{y}_1\mathbf{y}_2^T\mathbf{w}_1\mathbf{w}_2^T\mathbf{y}_3\mathbf{y}_4^T\}$

A seguinte derivação é baseada no Anexo 3 de Costa (2001) e resulta na fatoração de $E\{\mathbf{y}_1\mathbf{y}_2^T\mathbf{w}_1\mathbf{w}_2^T\mathbf{y}_3\mathbf{y}_4^T\}$ em termos de valores esperados de no máximo duas variáveis aleatórias.

Considerando-se a independência entre os vetores \mathbf{y} e \mathbf{w} , e supondo que \mathbf{y}_1 , \mathbf{y}_2 , \mathbf{y}_3 e \mathbf{y}_4 sejam vetores de variáveis aleatórias reais, conjuntamente Gaussianas de média zero, e ainda que \mathbf{w}_1 e \mathbf{w}_2 são vetores de elementos reais constantes, é possível atribuir uma nova matriz \mathbf{C} que represente a operação entre estas variáveis, ou seja,

$$\mathbf{C} = \mathbf{y}_1\mathbf{y}_2^T\mathbf{w}_1\mathbf{w}_2^T\mathbf{y}_3\mathbf{y}_4^T \quad (\text{F.1})$$

sendo possível expandir os elementos da matriz como

$$\begin{aligned} \{c_{ij}\} &= \sum_{l=1}^N \sum_{p=1}^N (\mathbf{y}_1\mathbf{y}_2^T)_{i,l} (\mathbf{w}_1\mathbf{w}_2^T)_{l,p} (\mathbf{y}_3\mathbf{y}_4^T)_{p,j} \\ &= \sum_{l=1}^N \sum_{p=1}^N y_{1,i}y_{2,l}w_{1,l}w_{2,p}y_{3,p}y_{4,j} \end{aligned} \quad (\text{F.2})$$

em que i representa a i -ésima linha e j a j -ésima coluna da matriz \mathbf{C} de tamanho $N \times N$.

Tomando-se o valor esperado de (F.2), reagrupando os elementos escalares com base na suposição de independência entre estas variáveis, tem-se que

$$E\{c_{ij}\} = \sum_{l=1}^N \sum_{p=1}^N E\{y_{1,i}y_{2,l}y_{3,p}y_{4,j}\} E\{w_{1,l}w_{2,p}\}. \quad (\text{F.3})$$

Desta forma, o primeiro valor esperado do lado direito da equação (F.3) pode ser simplificado através do *teorema da fatoração dos momentos Gaussianos* (HAYKIN, 2002, p. 69), onde para um conjunto de variáveis reais Gaussianas, tem-se que

$$\begin{aligned} E\{\mathbf{x}_1\mathbf{x}_2^T\mathbf{x}_3\mathbf{x}_4^T\} &= E\{\mathbf{x}_1\mathbf{x}_2^T\} E\{\mathbf{x}_3\mathbf{x}_4^T\} + E\{\mathbf{x}_1\mathbf{x}_3^T\} E\{\mathbf{x}_2\mathbf{x}_4^T\} \\ &\quad + E\{\mathbf{x}_1\mathbf{x}_4^T\} E\{\mathbf{x}_2\mathbf{x}_3^T\}. \end{aligned} \quad (\text{F.4})$$

Substituindo a equação (F.4) em (F.3), tem-se que

$$\begin{aligned}
 E\{c_{ij}\} &= \sum_{l=1}^N \sum_{p=1}^N E\{y_{1,i}y_{2,l}\}w_{1,l}w_{2,p}E\{y_{3,p}y_{4,j}\} \\
 &+ \sum_{l=1}^N \sum_{p=1}^N E\{y_{1,i}y_{3,p}\}w_{2,p}w_{1,l}E\{y_{2,l}y_{4,j}\} \\
 &+ E\{y_{1,i}y_{4,j}\}E\left\{\sum_{l=1}^N y_{2,l}w_{1,l} \sum_{p=1}^N y_{3,p}w_{2,p}\right\}
 \end{aligned} \tag{F.5}$$

ou na forma matricial,

$$\begin{aligned}
 E\{\mathbf{C}\} &= E\{\mathbf{y}_1\mathbf{y}_2^T\}\mathbf{w}_1\mathbf{w}_2^T E\{\mathbf{y}_3\mathbf{y}_4^T\} \\
 &+ E\{\mathbf{y}_1\mathbf{y}_3^T\}\mathbf{w}_2\mathbf{w}_1^T E\{\mathbf{y}_2\mathbf{y}_4^T\} \\
 &+ E\{\mathbf{y}_1\mathbf{y}_4^T\}\mathbf{w}_1^T E\{\mathbf{y}_2\mathbf{y}_3^T\}\mathbf{w}_2.
 \end{aligned} \tag{F.6}$$

ANEXO D – AVALIAÇÃO DA FUNÇÃO DE CORRELAÇÃO DE UM SINAL AUTOREGRESSIVO

Segue a reprodução modificada das seções de introdução e desenvolvimento do relatório técnico de Costa (2011) apud (NICOLAU, 2010).

INTRODUÇÃO

Este relatório tem como objetivo determinar de forma teórica a função de correlação de um sinal autoregressivo. O desenvolvimento aqui apresentado baseia-se em (NICOLAU, 2010).

DESENVOLVIMENTO

Um sinal autoregressivo é definido como:

$$x(n) = \sum_{i=1}^P a_i x(n-i) + \eta(n). \quad (G.1)$$

Multiplicando-se ambos os lados da equação (G.1) por $x(n-l)$

$$x(n)x(n-l) = \sum_{i=1}^P a_i x(n-i)x(n-l) + \eta(n)x(n-l). \quad (G.2)$$

Tomando-se o valor esperado de (G.2),

$$r_{xx}(l) = \sum_{i=1}^P a_i r_{xx}(l-i) + r_{\eta x}(l) \quad (G.3)$$

onde $r_{xx}(l) = E\{x(n)x(n-l)\}$ e $r_{\eta x}(l) = E\{\eta(n)x(n-l)\}$

A partir de (G.3) obtém-se o seguinte conjunto de equações:

$$\left\{ \begin{array}{l} r_{xx}(0) = a_1 r_{xx}(-1) + a_2 r_{xx}(-2) + a_3 r_{xx}(-3) \dots + a_{p-1} r_{xx}(-P+1) + a_p r_{xx}(-P) + r_{\eta x}(0) \\ r_{xx}(1) = a_1 r_{xx}(0) + a_2 r_{xx}(-1) + a_3 r_{xx}(-2) \dots + a_{p-1} r_{xx}(-P+2) + a_p r_{xx}(-P+1) + r_{\eta x}(1) \\ r_{xx}(2) = a_1 r_{xx}(1) + a_2 r_{xx}(0) + a_3 r_{xx}(-1) \dots + a_{p-1} r_{xx}(-P+3) + a_p r_{xx}(-P+2) + r_{\eta x}(2) \\ r_{xx}(3) = a_1 r_{xx}(2) + a_2 r_{xx}(1) + a_3 r_{xx}(0) \dots + a_{p-1} r_{xx}(-P+4) + a_p r_{xx}(-P+3) + r_{\eta x}(3) \\ \dots \\ r_{xx}(P) = a_1 r_{xx}(P-1) + a_2 r_{xx}(P-2) \dots + a_{p-1} r_{xx}(1) + a_p r_{xx}(0) + r_{\eta x}(P) \\ r_{xx}(P+1) = a_1 r_{xx}(P) + a_2 r_{xx}(P-1) \dots + a_{p-1} r_{xx}(2) + a_p r_{xx}(1) + r_{\eta x}(P+1) \\ \dots \\ r_{xx}(L-1) = a_1 r_{xx}(L-2) + a_2 r_{xx}(L-3) + a_3 r_{xx}(L-4) \dots + a_p r_{xx}(L-P+1) + r_{\eta x}(L-1) \\ r_{xx}(L) = a_1 r_{xx}(L-1) + a_2 r_{xx}(L-2) + a_3 r_{xx}(L-3) \dots + a_p r_{xx}(L-P) + r_{\eta x}(L) \end{array} \right. \quad (G.4)$$

em que L é o comprimento desejado da função de correlação.

O conjunto de equações apresentado em (G.4) pode ser agrupado matricialmente da seguinte forma:

$$\begin{bmatrix} r_{xx}(0) \\ r_{xx}(1) \\ \vdots \\ r_{xx}(P) \\ r_{xx}(P+1) \\ \vdots \\ r_{xx}(L) \end{bmatrix} = \begin{bmatrix} \dots & 0 & a_p & \dots & a_2 & a_1 & 0 & \dots \\ & \dots & 0 & a_p & \dots & a_2 & a_1 & 0 & \dots \\ & & \ddots \\ & & & \dots & 0 & a_p & \dots & a_2 & a_1 & 0 & \dots \\ & & & \dots & \dots & 0 & a_p & \dots & a_2 & a_1 & 0 & \dots \\ & & & & \dots & \dots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ & & & & & \dots & 0 & a_p & \dots & a_2 & a_1 & 0 & 0 \\ & & & & & & \dots & 0 & a_p & \dots & a_2 & a_1 & 0 \end{bmatrix} \begin{bmatrix} r_{xx}(-L) \\ \vdots \\ r_{xx}(-P-1) \\ r_{xx}(-P) \\ \vdots \\ r_{xx}(-1) \\ r_{xx}(0) \\ r_{xx}(1) \\ \vdots \\ r_{xx}(P) \\ r_{xx}(P+1) \\ \vdots \\ r_{xx}(L) \end{bmatrix} + \begin{bmatrix} r_{\eta x}(0) \\ r_{\eta x}(1) \\ \vdots \\ r_{\eta x}(P) \\ r_{\eta x}(P+1) \\ \vdots \\ r_{\eta x}(L) \end{bmatrix} \quad (\text{G.5})$$

Sabendo-se que para sinais reais $r_{xx}(l) = r_{xx}(-l)$ então (G.5) torna-se:

$$\begin{bmatrix} r_{xx}(0) \\ r_{xx}(1) \\ r_{xx}(2) \\ r_{xx}(3) \\ r_{xx}(4) \\ \vdots \\ r_{xx}(P) \\ r_{xx}(P+1) \\ \vdots \\ r_{xx}(L-2) \\ r_{xx}(L-1) \\ r_{xx}(L) \end{bmatrix} = \begin{bmatrix} 0 & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & \dots & a_p & 0 & \dots & 0 \\ a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & \dots & a_p & 0 & 0 & \dots & 0 \\ a_2 & (a_1+a_3) & a_4 & a_5 & a_6 & \dots & a_p & 0 & 0 & 0 & \dots & 0 \\ a_3 & (a_2+a_4) & (a_1+a_5) & a_6 & \dots & a_p & 0 & 0 & 0 & 0 & \dots & 0 \\ a_4 & (a_3+a_5) & (a_2+a_6) & (a_1+a_7) & \dots & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots \\ a_p & a_{p-1} & a_{p-2} & a_{p-3} & a_{p-4} & a_{p-5} & a_{p-6} & \dots & a_1 & 0 & \dots & 0 \\ 0 & a_p & a_{p-1} & a_{p-2} & a_{p-3} & a_{p-4} & a_{p-5} & \dots & a_1 & 0 & \dots & 0 \\ \vdots & \vdots \\ 0 & 0 & a_p & a_{p-1} & a_{p-2} & a_{p-3} & a_{p-4} & \dots & a_1 & 0 & \dots & 0 \\ \vdots & \vdots \\ 0 & \dots & 0 & 0 & a_p & \dots & a_6 & a_5 & a_4 & a_3 & a_2 & a_1 \\ 0 & \dots & 0 & a_p & \dots & a_6 & a_5 & a_4 & a_3 & a_2 & a_1 & 0 \end{bmatrix} \begin{bmatrix} r_{xx}(0) \\ r_{xx}(1) \\ r_{xx}(2) \\ r_{xx}(3) \\ r_{xx}(4) \\ \vdots \\ r_{xx}(P) \\ r_{xx}(P+1) \\ \vdots \\ r_{xx}(L-2) \\ r_{xx}(L-1) \\ r_{xx}(L) \end{bmatrix} + \begin{bmatrix} r_{\eta x}(0) \\ r_{\eta x}(1) \\ r_{\eta x}(2) \\ r_{\eta x}(3) \\ r_{\eta x}(4) \\ \vdots \\ r_{\eta x}(P) \\ r_{\eta x}(P+1) \\ \vdots \\ r_{\eta x}(L-2) \\ r_{\eta x}(L-1) \\ r_{\eta x}(L) \end{bmatrix} \quad (\text{G.6})$$

A equação (G.6) pode ser representada por

$$\mathbf{r}_{xx} = \mathbf{\Psi} \cdot \mathbf{r}_{xx} + \mathbf{r}_{\eta x}. \quad (\text{G.7})$$

Os elementos da matriz $\mathbf{\Psi}$ podem ser calculados através do seguinte algoritmo:

Seja um vetor \mathbf{v}_0 de tamanho $2L+1$ e índices associados, conforme a Figura 28a, preenche-se inicialmente o vetor com os coeficientes do modelo AR nos respectivos índices do vetor. Os demais elementos são nulos. Defina-se ainda o vetor \mathbf{v}_i dado pelo vetor original \mathbf{v}_0 cujo conteúdo é sujeito a um deslocamento de i posições para a esquerda com inserção de zeros à direita (Figuras 1b e 1c).

Os elementos de cada linha da matriz Ψ são definidos pelo vetor linha:

$$\Psi_i = [v_0 \quad (v_{-1} + v_{+1}) \quad (v_{-2} + v_{+2}) \quad (v_{-3} + v_{+3}) \quad \dots \quad (v_{-L} + v_{+L})] \quad (\text{G.8})$$

onde i relaciona-se à linha da matriz Ψ .

0	...	0	0	0	0	0	0	a_1	a_2	...	a_P	0	...	0
$-L$...	-5	-4	-3	-2	-1	0	1	2	3	4	5	...	L

(a) \mathbf{v}_0

0	...	0	0	0	0	0	a_1	a_2	...	a_P	0	...	0	0
$-L$...	-5	-4	-3	-2	-1	0	1	2	3	4	5	...	L

(b) \mathbf{v}_1

0	...	0	0	0	0	a_1	a_2	...	a_P	0	...	0	0	0
$-L$...	-5	-4	-3	-2	-1	0	1	2	3	4	5	...	L

(c) \mathbf{v}_2

Figura 29 – Vetor base (a) e deslocados (b e c) do algoritmo para determinação dos elementos da matriz Ψ .

Em decorrência da característica de ruído branco, o vetor $\mathbf{r}_{\eta\mathbf{x}}$ é dado por:

$$\mathbf{r}_{\eta\mathbf{x}} = [\sigma_\eta^2 \quad 0 \quad \dots \quad 0]^T \quad (\text{G.9})$$

onde σ_η^2 é a potência da inovação do processo AR.

Conhecidos Ψ e $\mathbf{r}_{\eta\mathbf{x}}$ obtém-se $\mathbf{r}_{\mathbf{xx}}$ da seguinte forma:

$$\mathbf{r}_{\mathbf{xx}} = [\mathbf{I} - \Psi]^{-1} \mathbf{r}_{\eta\mathbf{x}}. \quad (\text{G.10})$$

ANEXO E – INFERÊNCIA ESTATÍSTICA SOBRE A VARIÂNCIA DE UMA POPULAÇÃO

Este anexo apresenta um procedimento para definir o tamanho de uma amostra de forma a inferir a variância populacional e o intervalo de confiança, através de teste de hipóteses. Considera-se amostra um conjunto de R elementos da população de uma variável Gaussiana.

Conceitualmente, uma hipótese é uma afirmação acerca de um ou mais parâmetros θ da distribuição de probabilidades de uma variável aleatória, portanto, não deve ser aplicada a uma amostra. Assim, um valor constante θ_0 é utilizado para determinar a condição de teste para o caso de uma única variável populacional. O conjunto de testes em (H.1) contém a *hipótese nula* H_0 , que desejamos testar, e a *hipótese alternativa bilateral* H_1 , que corresponde à rejeição da hipótese nula.

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0. \end{cases} \quad (\text{H.1})$$

Um teste hipotético alternativo unilateral pode ser obtido através de H_1 , considerando apenas valores maiores ou menores que θ_0 . O procedimento consiste em aplicar uma *estatística de teste* em uma amostra aleatória, de forma a tomar uma decisão em relação à hipótese nula (MONTGOMERY e RUNGER, 1999, p. 140). Dessa forma, duas regiões são definidas com base na estimativa do parâmetro considerando a distribuição empregada no teste: a *região de aceitação*, onde parâmetros estimados contidos nela suportam a hipótese nula, e a *região de rejeição* que rejeita H_0 . Os limiares dessas regiões são denominados de *valores críticos*. Na Figura 30.a, a área branca é a região de aceitação e a área sombreada é a região de rejeição.

A rejeição de H_0 quando a estimativa amostral do parâmetro está contida na região crítica, mas o parâmetro é válido, é denominado de *erro de tipo I*, enquanto que o *tipo II* é a falha em rejeitar a hipótese nula quando a estimativa amostral está contida na região de aceitação, sendo H_0 falsa. Os símbolos α e β correspondem à probabilidade de cada erro, respectivamente. Três pontos importantes são definidos:

1. O tamanho da região crítica e, conseqüentemente, a probabilidade do erro tipo I, α , sempre pode ser reduzido através da seleção apropriada dos valores críticos;
2. Os erros tipo I e tipo II estão relacionados. Uma diminuição na probabilidade de um tipo de erro sempre resulta em um aumento

na probabilidade do outro, desde que o tamanho da amostra R não varie;

3. Um aumento no tamanho da amostra geralmente reduzirá α e β , desde que os valores críticos sejam mantidos constantes. (MONTGOMERY e RUNGER, 1999, p. 146)

O ajuste dos valores críticos é realizado através do controle da probabilidade α , controlando dessa forma a probabilidade de rejeitar erroneamente H_0 . O parâmetro β depende do tamanho da amostra e da extensão onde H_0 é falsa, ou seja, a falha em rejeitar H_0 . Isso não significa necessariamente que há uma alta probabilidade de que a hipótese nula seja verdadeira, mas que talvez seja necessário um maior número de elementos da amostra.

Três parâmetros podem ser extraídos da inferência de um estimador não-tendencioso: (a) média com variância conhecida (avaliação baseada na distribuição Gaussiana), (b) média com variância desconhecida (avaliação baseada na distribuição t-student) e (c) variância desconhecida (avaliação baseada na distribuição qui-quadrada). Esta análise, para o valor quadrático, se baseia no item (c).

O estimador não tendencioso da variância amostral s^2 de σ^2 da população x é expresso por (PAPOULIS e PILLAI, 2002, p. 307)

$$s^2 = \frac{1}{R-1} \sum_{i=1}^R (x_i - \bar{x})^2 \quad (\text{H.2})$$

sendo que x_i é i -ésima amostra e \bar{x} a sua média. Deseja-se testar o seguinte conjunto de hipóteses, conforme (H.1),

$$\begin{cases} H_0 : \sigma^2 = \sigma_o^2 \\ H_1 : \sigma^2 \neq \sigma_o^2. \end{cases} \quad (\text{H.3})$$

Considerando que a variável é Gaussiana, tem-se que para um conjunto de R elementos da amostra, a distribuição da estatística de teste é dada por uma distribuição qui-quadrada com $R-1$ graus de liberdade, denotada por χ^2_{R-1} ,

$$\chi^2 = \frac{(R-1)s^2}{\sigma^2}. \quad (\text{H.4})$$

Conforme $R \rightarrow \infty$, a distribuição tende a uma Gaussiana, e o parâmetro estimado à σ^2 . A Figura 30.a apresenta a curva de uma distribuição qui-

quadrada para um dado valor de $R-1$ graus de liberdade. A área sombreada corresponde à probabilidade α e os pontos percentuais superior $\chi^2_{\alpha/2, R-1}$ e inferior $\chi^2_{1-\alpha/2, R-1}$ da distribuição representam os valores críticos citados anteriormente, sendo obtidos através de uma tabela, como a Tabela 7.

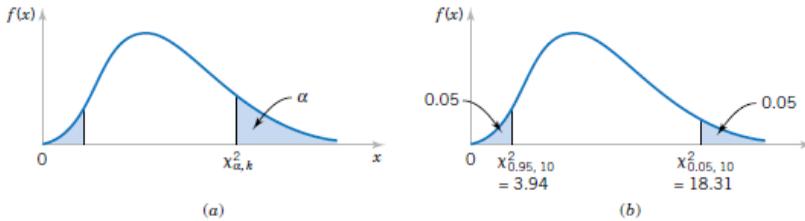


Figura 30 – Pontos percentuais superior e inferior da distribuição qui-quadrada.

O intervalo de confiança, desenvolvido inicialmente por Jerzy Neyman (1894-1981) em 1937, é definido em (MONTGOMERY e RUNGER, 1999, p. 163) como a probabilidade,

$$P\left[\chi^2_{\alpha/2, R-1} < \chi^2 < \chi^2_{1-\alpha/2, R-1}\right] = (1-\alpha) \quad (\text{H.5})$$

sendo que $(1-\alpha)$ é o coeficiente de confiança e α é o percentual de confiança (*erro de tipo I*). Na área de probabilidade frequencista, o intervalo de confiança é uma medida que expressa o grau de incerteza associado a uma estimativa. Quanto mais próximo $(1-\alpha)$ estiver do valor unitário, reduz-se a área sombreada na Figura 30 e, portanto, é mais provável que as estimativas de θ estejam contidos na área não sombreada (região de aceitação). Ressalta-se que o IC difere do intervalo de tolerância, o qual pode ser visto como uma concessão para imperfeições nas medidas de uma dada variável.

Em (MONTGOMERY e RUNGER, 1999, p. 140), cita-se que “intervalos de confiança estão intimamente relacionados a uma outra técnica estatística de tomada de decisão, chamada de *teste de hipóteses*”, pois podem ser tomadas com base na faixa de valores do intervalo de confiança. Assim, substituindo-se (H.4) em (H.5), tem-se que

$$P\left\{\chi^2_{\alpha/2, R-1} < \frac{(R-1)s^2}{\sigma^2} < \chi^2_{1-\alpha/2, R-1}\right\} = (1-\alpha). \quad (\text{H.6})$$

Tabela 7- Pontos percentuais da distribuição qui-quadrada.

	α	$\chi^2_{\alpha/2, 499}$	$\chi^2_{1-\alpha/2, 499}$
99%	0,01	584,1	421,4
95%	0,05	562,8	439
90%	0,10	552,1	448,2

Dessa forma, o intervalo de confiança de $100(1 - \alpha)\%$ para σ^2 é

$$\frac{(R-1)s^2}{\chi_{1-\alpha/2, R-1}^2} < \sigma^2 < \frac{(R-1)s^2}{\chi_{\alpha/2, R-1}^2}. \quad (\text{H.7})$$

O valor médio da diferença entre esses dois limites corresponde ao IC. Em escala logarítmica, o IC relativo (IC_R) é função dos pontos percentuais da distribuição qui-quadrada,

$$\begin{aligned} IC_R &= 10/2 \left[\log_{10} \frac{(R-1)s^2}{\chi_{\alpha/2, R-1}^2} - \log_{10} \frac{(R-1)s^2}{\chi_{1-\alpha/2, R-1}^2} \right] \\ &= 5 \cdot \log_{10} \left(\frac{\chi_{\alpha/2, R-1}^2}{\chi_{1-\alpha/2, R-1}^2} \right). \end{aligned} \quad (\text{H.8})$$

A probabilidade de erro tipo II, ou a *potência* ($1-\beta$), pode ser utilizada para determinar a quantidade de elementos da amostra necessária para detectar um particular valor de σ^2 , que difere do valor hipotético σ_0^2 , sendo medido pela razão λ ,

$$\lambda = \frac{\sigma^2}{\sigma_0^2}. \quad (\text{H.9})$$

Portanto, a potência é uma medida de sensibilidade em um teste estatístico, ou seja, da capacidade em detectar diferenças nos parâmetros (MONTGOMERY e RUNGER, 1999, p. 146). Gráficos de curvas características operacionais são usados para determinar o valor de β (erro tipo II) em função de R , λ e α . A equação que relaciona esses valores é difícil de ser encontrada. Uma análise aproximada, baseada nas equações para a média como figura de mérito, é expressa por

$$\beta = \chi_{[1/\lambda^2]\varphi_{\alpha/2, R-1, R-1}^2} - \chi_{[1/\lambda^2]\varphi_{1-\alpha/2, R-1, R-1}^2} \quad (\text{H.10})$$

sendo que φ^2 é a distribuição inversa da qui-quadrada. O comportamento de β para diferentes valores de R e λ é apresentado numericamente na Figura 31 para $\alpha = 0,01$ e na Figura 32 para $\alpha = 0,05$. Esses gráficos obtidos pela equação (H.10) são aproximações dos obtidos em VI.i e VI.j de (MONTGOMERY e RUNGER, 1999, p. 410).

Observa-se através dessas figuras que conforme o valor de λ se aproxima do valor unitário, menor a potência do teste para um determinado valor de R . Isso é esperado, pois não há margem para variação (pouca sensibilidade). Por outro lado, quanto maior o valor de R , para um dado valor de α e λ , menor o β e consequentemente maior a potência do teste. Claramente, quanto maior o tamanho da amostra, mais sensível ele é para detectar a diferença entre σ^2 e σ_0^2 , ou seja, rejeitar corretamente uma hipótese nula falsa.

Dessa forma, o procedimento para obter o tamanho da amostra para um teste estatístico, consiste em encontrar R utilizando as curvas características de (H.10) através de algum algoritmo de otimização, sendo função de λ , α e $(1-\beta)$. A equação (H.8) determina o intervalo de confiança através de R e α , e pode ser usado para reduzir a incerteza do teste. Quanto menor α e β , e mais próximo do valor unitário for λ , maior o número de promediações e menor o IC.

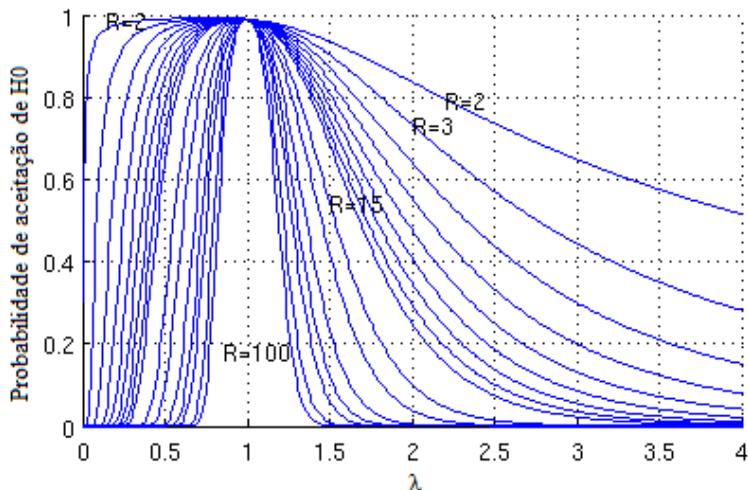


Figura 31 – Curvas Características Operacionais para $\alpha = 0,01$.

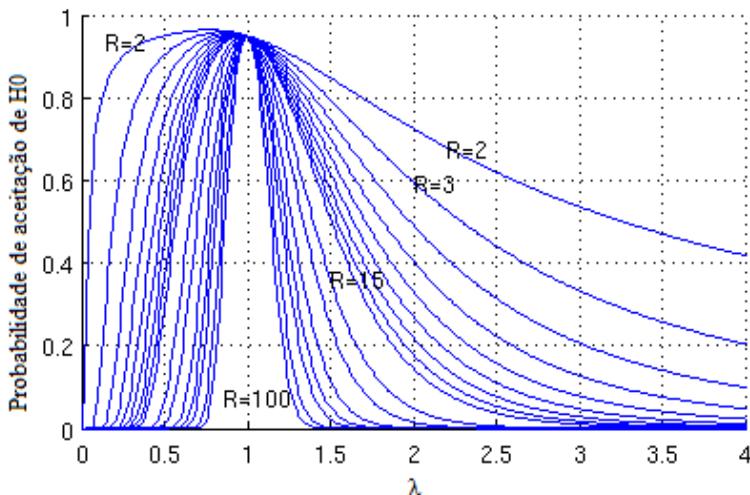


Figura 32 – Curvas Características Operacionais para $\alpha = 0,05$.