

MINISTÉRIO DA EDUCAÇÃO E DO DESPORTO
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA METALÚRGICA E DOS
MATERIAIS - PPGEMM

FILTROS ADAPTATIVOS APLICADOS À SEPARAÇÃO DE SINAIS

por

Marcelo Negreiros

Engenheiro Eletricista (UFRGS, 1992)

Trabalho realizado no Departamento de Engenharia Elétrica da Escola de Engenharia da UFRGS, dentro do Programa de Pós-Graduação em Engenharia Metalúrgica e dos Materiais - PPGEMM.

Porto Alegre

1994

Título da dissertação:

FILTROS ADAPTATIVOS APLICADOS À SEPARAÇÃO DE SINAIS

Apresentada ao Programa de Pós-Graduação em Engenharia Metalúrgica e dos Materiais - PPGEMM, como parte dos requisitos para a obtenção do título de

Mestre em Engenharia

Área de Concentração: Instrumentação Eletro-Eletrônica

por

Marcelo Negreiros, Engenheiro Eletricista

1994

Esta DISSERTAÇÃO foi julgada adequada para a obtenção do título de Mestre em Engenharia, Área de Concentração em Instrumentação Eletro-Eletrônica e aprovada na sua forma final pelo Orientador e pela Banca Examinadora do Curso de Pós-Graduação.

Orientador: _____

Altamiro Amadeu Suzim

Doutor em Informática

Professor do Programa de Pós-Graduação em Engenharia

Metalúrgica e dos Materiais da UFRGS

Banca examinadora:

Renato Machado de Brito

Doutor em Engenharia pela UFRGS

Professor do Programa de Pós-Graduação em Engenharia

Metalúrgica e dos Materiais da UFRGS

Álvaro Augusto Almeida de Salles

PhD em Engenharia Elétrica pela Universidade de Londres

Professor do Departamento de Engenharia Elétrica da UFRGS

Liliane Basso Barichello

Doutor em Engenharia pela UFRGS

Professora do Departamento de Matemática Pura e Aplicada do

Instituto de Matemática da UFRGS

Coordenador do PPGEMM

Prof. Dr. Telmo Strohaecker

Esta DISSERTAÇÃO somente pôde ser concluída graças ao trabalho de co-orientação desenvolvido pelo prof. Thomas Weihmann (28/09/67-23/08/94), Mestre em Engenharia e professor do Departamento de Engenharia Elétrica da UFRGS.

Agradecimentos:

Ao professor e orientador Dr. Altamiro Amadeu Suzim, pelo exemplo, pelo permanente apoio e pela dedicação a seus ideais.

Ao professor e co-orientador Msc. Thomas Weihmann (28/09/67-23/08/94), pelo apoio e dedicação, pelo exemplo de vida, pela ajuda prestada no desenvolvimento do trabalho. Concordo com o colega Cícero R. Lorenzi: "... acho que alguém se enganou e pegou o último da fila".

Ao colega de graduação Msc. Joel A. Luft, por estar sempre presente no acompanhamento deste trabalho, sempre com disposição para oferecer críticas e ajuda. Barbada!

Ao professor Msc Alceu Heinke Frigeri, por sua disposição permanente para colaborar.

Aos colegas Cícero Renato Lorenzi e Fernando Martins Bauer, pela colaboração e companheirismo.

A todos os demais colegas do Laboratório de Instrumentação Eletro-Eletrônica e professores do departamento de Eng^a. Elétrica, que souberam fazer do laboratório um ambiente ímpar, em todos os aspectos, principalmente de boa-vontade, trabalho e dedicação.

Aos demais professores, funcionários e colegas que, de uma forma ou outra, colaboraram neste trabalho.

A minha família, pelo apoio incondicional, pela formação a mim propiciada, pela compreensão. Em especial à minha mãe, pelo exemplo de dedicação, pela paciência e boa-vontade infinitas. Agradeço também a minha avó, por estar sempre disposta a ouvir e por não se importar em perder tempo cozinhando, sempre preocupada...

Ao CNPq e Banco do Brasil, pela bolsa de mestrado e pelo apoio financeiro.

Em especial, a todos aqueles que, de um modo ou outro, colaboram para o surgimento e manutenção de locais de pesquisa e ensino em nosso país e no exterior.

SUMÁRIO

LISTA DE FIGURAS.....	xiii
LISTA DE TABELAS.....	xvii
LISTA DE ABREVIATURAS	xviii
RESUMO.....	xix
ABSTRACT.....	xx
1 <u>INTRODUÇÃO</u>	1
<u>1.1</u> <u>O problema das entradas interferentes em um sistema de medição</u>	1
<u>1.2</u> <u>O cancelador adaptativo de ruído como uma solução</u>	3
<u>1.3</u> <u>Conceitos de filtragem ótima e filtros adaptativos</u>	6
<u>1.4</u> <u>Motivação para a abordagem do problema</u>	9
<u>1.5</u> <u>Organização do trabalho</u>	10
2 <u>FUNDAMENTAÇÃO</u>	12
<u>2.1</u> <u>Introdução</u>	12
<u>2.2</u> <u>Processos estocásticos</u>	13
2.2.1 Sinais amostrados versus processos estocásticos	13
2.2.2 Representação frequencial de processos estocásticos.....	15
2.2.3 Efeito de sistemas lineares invariantes no tempo	17
<u>2.3</u> <u>Estruturas para implementação de filtros digitais</u>	18
<u>2.4</u> <u>O problema de filtragem ótima linear e sua solução</u>	20
2.4.1 Introdução	20
2.4.2 Filtros de Wiener	22
2.4.2.1 História	22

2.4.2.2	O problema de filtragem ótima linear do ponto de vista estatístico	22
2.4.2.3	Função de custo	23
2.4.2.4	Princípio da ortogonalidade	24
2.4.2.5	Equações de Wiener-Hopf.....	26
2.4.2.6	Superfície de performance do erro	27
2.4.2.7	Valor mínimo da função de custo J.....	27
2.4.2.8	Formulação matricial das equações de Wiener-Hopf.....	28
2.4.2.9	Forma matricial do erro mínimo.....	30
2.4.2.10	Forma canônica da superfície de performance do erro.....	31
2.4.2.11	O filtro de Wiener não-causal	32
2.4.3	Método dos mínimos quadrados.....	33
2.4.3.1	Introdução	33
2.4.3.2	Janelamento de dados.....	34
2.4.3.3	Princípio da ortogonalidade.....	35
2.4.3.4	Equações normais	36
<u>2.5</u>	<u>O cancelador adaptativo de ruído</u>	39
2.5.1	Conceito.....	39
2.5.2	Soluções estacionárias utilizando a teoria de filtros de Wiener.....	40
2.5.2.1	Análise usando relação sinal-ruído	42
2.5.3	Efeitos de componentes do sinal na referência.....	44
2.5.3.1	Relação sinal-ruído na saída do cancelador ($\rho_{out}(z)$)	45
2.5.3.2	Distorção do sinal na saída do cancelador.....	46
2.5.3.3	Espectro do ruído na saída do cancelador.....	47
2.5.4	Problemas práticos.....	47
<u>2.6</u>	<u>Predição linear</u>	48
2.6.1	Preditores <i>lattice</i>	51
<u>3</u>	<u>FILTROS ADAPTATIVOS</u>	54
<u>3.1</u>	<u>Introdução</u>	54

3.2	<u>Classificação</u>	57
3.3	<u>Método de <i>steepest descent</i></u>	59
3.3.1	Análise de estabilidade	61
3.4	<u>Algoritmo LMS</u>	63
3.4.1	Derivação	64
3.4.2	Análise de estabilidade	66
3.4.3	Análise para ambiente não-estacionário	69
3.4.4	Algoritmos relacionados.....	70
3.5	<u>Algoritmo <i>Recursive Least Squares</i> ou RLS</u>	71
3.5.1	Algoritmo RLS com peso de dados exponencial.....	73
3.5.2	Operação em ambiente não-estacionário.....	76
3.5.3	Algoritmos <i>fast RLS</i>	76
3.6	<u>Algoritmos para filtros <i>lattice</i></u>	77
3.6.1	Derivação do algoritmo <i>lattice</i> LMS ou GAL para o caso real	77
3.7	<u>Algoritmo <i>fast RLS</i> para filtros <i>lattice</i></u>	81
3.8	<u>Algoritmos para filtros adaptativos IIR</u>	83
3.8.1	Algoritmo IIR LMS	83
3.8.1	Algoritmo <i>Recursive</i> LMS	85
3.9	<u>Algoritmo para filtro adaptativo IIR baseado em hiper-estabilidade</u>	87
3.10	<u>Problemas de filtros IIR</u>	88
3.11	<u>Comparação dos algoritmos através de um exemplo</u>	89
4	<u>APLICAÇÃO AO PROBLEMA DE CANCELAMENTO DE ECO</u>	95
4.1	<u>Introdução</u>	95
4.2	<u>O problema de eco em telefonia</u>	95
4.2.1	Noções de telefonia.....	95
4.2.2	Ecos e sistemas para cancelamento.....	96
4.3	<u>O Módulo Compressor de Voz -MCV</u>	99
4.3.1	O projeto Digitel-LaPSi	99
4.3.2	O problema de eco específico do MCV	100

<u>4.4</u>	<u>Abordagem do problema</u>	101
4.4.1	Abordagem inicial.....	101
4.4.2	Porque a atenuação não alcança níveis maiores ?.....	104
4.4.2.1	Identificação da duração da resposta ao impulso do conjunto "híbrida+telefone".....	104
4.4.2.2	Análise das medidas de atenuação.....	106
4.4.2.3	Análise do efeito de perda de amostra.....	108
4.4.2.4	Conclusões.....	110
4.4.3	Qual a impressão do usuário sobre o funcionamento do cancelador implementado ?.....	110
4.4.4	Resultados de implementações alternativas.....	111
4.4.4.1	Algoritmos IIR - SHARF e <i>Recursive LMS</i>	111
4.4.4.2	Algoritmo <i>lattice</i> simplificado.....	113
<u>4.5</u>	<u>Outros resultados experimentais</u>	113
4.5.1	Efeitos de variação do fator de convergência e do número de blocos.....	113
4.5.2	Diferença de atenuação para sinais de voz e ruído.....	116
<u>4.6</u>	<u>Conclusões</u>	117
<u>5</u>	<u>CONCLUSÃO</u>	120
<u>ANEXO A - IDENTIFICAÇÃO DA FUNÇÃO DE TRANSFERÊNCIA DO CONJUNTO "HÍBRIDA+TELEFONE"</u>		
		124
<u>A.1</u>	<u>Introdução</u>	124
<u>A.2</u>	<u>Preparação do módulo</u>	125
<u>A.3</u>	<u>Ensaio</u>	125
<u>A.4</u>	<u>Aplicação do algoritmo e simulação dos modelos</u>	126
<u>ANEXO B - MEDIDAS DE ATENUAÇÃO SUBJETIVAS</u>		
		129
<u>B.1</u>	<u>Objetivos</u>	129
<u>B.2</u>	<u>Metodologia de teste</u>	129
<u>B.3</u>	<u>Metodologia de análise</u>	130

<u>B.4 Resultados</u>	132
<u>ANEXO C - MEDIDAS DE ATENUAÇÃO ABSOLUTAS DO SINAL DE ECO</u>	134
<u>C.1 Objetivos</u>	134
<u>C.2 Metodologia</u>	134
<u>C.3 Configuração usada para a medição de nível e tempo de atenuação</u>	135
<u>C.4 Resultados obtidos</u>	136
<u>ANEXO D - O ALGORITMO BLOCK UPDATE LMS</u>	139
<u>ANEXO E - NOTAÇÃO VETORIAL</u>	141
<u>REFERÊNCIAS BIBLIOGRÁFICAS</u>	143
<u>APÊNDICE I - PROBABILIDADE, VARIÁVEIS ALEATÓRIAS E</u> <u>PROCESSOS ESTOCÁSTICOS</u>	148
<u>I.1 Introdução</u>	148
<u>I.2 Probabilidade e variáveis aleatórias</u>	148
I.2.1 Função distribuição	149
I.2.2 Função densidade	150
I.2.3 Função de variável aleatória	150
I.2.4 Caracterização de variáveis aleatórias	150
I.2.5 Duas variáveis aleatórias	151
<u>I.3 Processos estocásticos</u>	153
I.3.1 Propriedades gerais	155
I.3.2 Processos estacionários	156
I.3.3 Processos estocásticos discretos	156
<u>APÊNDICE II - ESTIMATIVA MÉDIA DO QUADRADO LINEAR E O</u> <u>PRINCÍPIO DA ORTOGONALIDADE</u>	159
<u>II.1 Introdução</u>	159
<u>II.2 Estimação de uma variável aleatória por uma função de outra variável</u> <u>aleatória</u>	159

<u>II.3</u> <u>Estimação de uma variável aleatória por uma função de N outras</u> <u>variáveis aleatórias</u>	161
------------------------------------------------------------------------------------------------------------------------	-----

LISTA DE FIGURAS

1.1	Diagrama em blocos genérico de um instrumento de medição.....	1
1.2	Configuração geral de entrada / saída de sistemas de medição	2
1.3	Modelo de sistema de medição com dois sensores em posições distintas.....	4
1.4	Modelo do problema das entradas interferentes e um sistema "caixa-preta" para sua solução	5
1.5	O problema com funções de transferência e sistema de cancelamento com filtro fixo	5
1.6	O cancelador adaptativo de ruído	6
1.7	O problema de estimação linear.....	7
2.1	Efeito de sistema LTI sobre sinais determinísticos e estocásticos.....	17
2.2	Representação de um sistema LTI, onde $y[n]=h[n]*x[n]$	18
2.3	Filtro digital com estrutura de forma direta.....	19
2.4	Filtro IIR implementado com 2 filtros FIR.....	20
2.5	Filtro IIR utilizando estrutura <i>lattice</i>	20
2.6	Representação do problema de filtragem ótima linear em diagrama de blocos.....	21
2.7	Filtro transversal usado na solução do problema de filtragem linear	22
2.8	Gráfico da função de custo J para os coeficientes reais w_0 e w_1	27
2.9	Modelo para a resolução do problema de filtragem ótima linear por mínimos quadrados.....	33
2.10	Modelo de regressão linear múltipla para formação da seqüência $d[n]$	34
2.11	O cancelador adaptativo de ruído.....	40
2.12	O cancelador adaptativo de ruído com ruídos não-correlacionados nas entradas	42
2.13	O problema de <i>leakage</i> do sinal desejado	44
2.14	Modelo de geração de sinal.....	48
2.15	Preditor "para frente"	49

2.16	Filtro de erro de predição para frente	49
2.17	Diagrama em blocos do filtro de erro de predição para frente	50
2.18	Preditor "para trás"	50
2.19	Filtro de erro de predição para trás.....	50
2.20	Diagrama em blocos do filtro de erro de predição para trás	51
2.21	Filtro de erro de predição <i>lattice</i>	52
2.22	Preditor <i>lattice</i>	53
3.1	Estrutura geral de um filtro adaptativo e sinais envolvidos.....	55
3.2	As quatro configurações básicas de filtros adaptativos.....	55
3.3	Filtro adaptativo FIR.....	59
3.4	<i>Steepest descent</i> mostrando $W[n]$ e $C[n]$	61
3.5	Realimentação de v_k	62
3.6	Curva de aprendizagem de v_k	62
3.7	Estrutura de realimentação no algoritmo LMS	65
3.8	Desajustamento M e seus componentes devido a E_1 (ruído) e E_2 (atraso).....	70
3.9	Sinais correspondentes ao algoritmo RLS	71
3.10	Diagrama de fluxo de sinais no algoritmo RLS com peso exponencial	75
3.11	Estrutura do filtro <i>lattice</i> para estimação de processo conjunto	80
3.12	Filtro IIR de estrutura direta	83
3.13	Filtro IIR com algoritmo SHARF.....	87
3.14	Trechos dos arquivos de sinais utilizados para a simulação e sua formação.....	89
3.15	Resposta ao impulso de $1/H(z)$	90
3.16	Curvas de aprendizagem instantâneas para o algoritmo LMS.....	90
3.17	Curvas de aprendizagem instantâneas para o algoritmo <i>lattice</i> LMS ou GAL	91
3.18	Curvas de aprendizagem instantâneas para o algoritmo RLS	91
3.19	Curvas de aprendizagem instantâneas para o algoritmo LSL.....	92
3.20	Curvas de aprendizagem instantâneas para o algoritmo IIR LMS	92
3.21	Curvas de aprendizagem instantâneas para o algoritmo <i>Recursive</i> LMS.....	93
3.22	Curvas de aprendizagem instantâneas para o algoritmo SHARF	93

3.23	<i>Root locus</i> para o algoritmo SHARF (10 segundos)	94
3.24	<i>Root locus</i> para algoritmo SHARF - <i>zoom</i> do primeiro segundo	94
4.1	Sistema telefônico local a dois fios	95
4.2	A anti-local	96
4.3	Sistema telefônico para longa distância	96
4.4	Tipos de eco	97
4.5	Supressor de eco	98
4.6	Cancelador de eco	98
4.7	O modelo digital para o problema de cancelamento de eco	99
4.8	O cancelador adaptativo de ruído como um cancelador de eco	99
4.9	Utilização do módulo compressor de voz	100
4.10	Ligação entre aparelhos telefônicos pelos módulos MCV	100
4.11	Diagrama em blocos do Módulo Compressor de Voz	101
4.12	Atenuação do eco de retorno em função do número de coeficientes do filtro	103
4.13	Carga computacional do cancelador de eco	103
4.14	Resposta ao salto do conjunto "híbrida+telefone"	105
4.15	Atenuação do eco de retorno em função do número de coeficientes do filtro	106
4.16	Carga computacional do cancelador de eco	106
4.17	Quadro comparativo de atenuações	107
4.18	Quadro comparativo de atenuações em escala linear	107
4.19	Curvas de aprendizagem instantâneas sem e com perda de amostra (simulação)	108
4.20	Comparação do tempo para atingir cerca de 70% da atenuação máxima	109
4.21	Curva de aprendizagem instantânea do módulo MCV digitalizada	109
4.22	Níveis de atenuação absolutos para algoritmos IIR implementados	112
4.23	Carga computacional dos algoritmos IIR	112
4.24	Atenuação em função do número de coeficientes e do fator de convergência	114
4.25	Tempo para o cancelador atingir cerca de 70% da sua atenuação máxima	114
4.26	Atenuação máxima obtida em função do número de blocos	115
4.27	Carga computacional versus número de blocos	115

4.28 Médias das atenuações dos sinais de voz e ruído para módulos síncronos e sem compressão.....	116
4.29 Médias das atenuações máximas para módulos assíncronos e com compressão.....	117
A.1 Funções de transferência em Z do filtro adaptativo e do conjunto "híbrida+telefone".....	124
A.2 Ensaio ao salto digitalizado.....	126
A.3 Resposta ao salto digitalizada.....	128
A.4 Resposta ao salto simulada em S	128
A.5 Resposta ao salto simulada em Z	128
B.1 Respostas para as questões 1, 2 e 3, para os filtros 1 (superior), 2, 3 e 4 (inferior).....	130
C.1 Configuração usada para medição de parâmetros.....	135
C.2 Sistema usado para excitar o telefone.....	135
D.1 Processo de adaptação dos coeficientes do filtro usando o algoritmo <i>block update</i> LMS.....	140
I.1 O espaço S e dois eventos A e B , onde $A \cap B \neq \emptyset$	149
I.2 Exemplo de funções distribuição e densidade de variáveis aleatórias.....	150
I.3 Conjunto (família) de funções que formam um processo estocástico contínuo no tempo.....	154
I.4 Interpretação da função distribuição de um processo estocástico, dados x e t	155
I.5 Conjunto (família) de funções que formam um processo estocástico discreto no tempo.....	157
II.1 Interpretação geométrica do princípio da ortogonalidade.....	161
II.2 Interpretação geométrica para o caso de $N=2$	162

LISTA DE TABELAS

2.1	Propriedades da função densidade de potência espectral.....	16
2.2	Propriedades da função densidade de potência espectral cruzada.....	16
2.3	Propriedades em Z das funções de densidade de potência espectral	17
2.4	Relações de transformação válidas em sistemas LTI para processos WSS.....	18
3.1	Algoritmo LMS	65
3.2	Algoritmo <i>Recursive Least Squares</i> com peso de dados exponencial	75
3.3	Comparação da ordem da carga computacional de algoritmos de mínimos quadrados.....	77
3.4	Algoritmo <i>Gradient Adaptive Lattice</i> (GAL) ou <i>Lattice LMS</i> - caso real	81
3.5	Algoritmo <i>Recursive Least Squares Lattice using a priori errors with error feedback</i>	82
3.6	Algoritmo IIR LMS	85
3.7	Algoritmo <i>Recursive LMS</i> - caso real	86
3.8	Algoritmo SHARF.....	88
4.1	Níveis subjetivos de amplitude do eco residual e atenuação do sinal	110
4.2	Desempenho subjetivo do cancelador em relação ao cancelador anterior (com menos coeficientes).....	111
B.1	Resultados para a pergunta 4 em termos do número de respostas.....	131
B.2	Mapeamento de conceitos para números	131
B.3	Medidas de desempenho subjetivas do cancelador de eco para 19 pessoas.....	133
C.2	Medidas de atenuação para o algoritmo <i>lattice LMS</i>	136
C.3	Medidas de atenuação dos algoritmos SHARF e <i>Recursive LMS</i>	137
C.1	Medidas de atenuação para o algoritmo <i>block update LMS</i>	138
D.1	Algoritmo <i>block update LMS</i> utilizado	140
I.1	Caracterização de processos estocásticos contínuos no tempo ($\tau=t_1-t_2$).....	157
I.2	Caracterização de processos estocásticos discretos no tempo ($m=n_1-n_2$)	158

LISTA DE ABREVIATURAS

ADPCM	<i>Adaptive Differential Pulse Code Modulation</i>
ANC	<i>Adaptive Noise Canceller</i>
A	Alto
AD, A/D	<i>Analog to Digital</i>
AR	<i>Autoregressive</i>
ARMA	<i>Autoregressive Moving Average</i>
MMB	Baixíssimo
B	Baixo / Bom (adequado)
CMA	<i>Constant Modulus Algorithm</i>
DSP	<i>Digital Signal Processing / Processor</i>
DA, D/A	<i>Digital to Analog</i>
FAEST	<i>FAst a posteriori Error Sequential Technique</i>
FIR	<i>Finite Impulse Response</i>
GAL	<i>Gradient Adaptive Lattice</i>
HARF	<i>Hyperstable Adaptive Recursive Filter</i>
I	Igual
IIR	<i>Infinite Impulse Response</i>
CCITT	<i>International Telegraph and Telephone Consultive Committee</i>
LMS	<i>Least Mean Squares</i>
LSL	<i>Least-Squares Lattice</i>
LPC	<i>Linear Predictive Coding</i>
LTl	<i>Linear Time Invariant (system)</i>
MCV	Módulo Compressor de Voz
MCVcod	Módulo Compressor de Voz com algoritmo de codificação e assíncrono
MCVncod	Módulo Compressor de Voz sem algoritmo de codificação e síncrono
MA	<i>Moving Average</i> / Muito Alto
MB	Muito Baixo
MM	Muito Melhor
MP	Muito Pior
PM	Pouco Melhor
PP	Pouco Pior
RLS	<i>Recursive Least-Squares</i>
RLSL	<i>Recursive Least-Squares Lattice</i>
ReclMS	<i>Recursive LMS</i>
R	Ruim (não-adequado)
SHARF	<i>Simplified HARF</i>
WSS	<i>Wide Sense Stationary</i>
SSS	<i>Strict Sense Stationary</i>
BIBO	<i>Bounded Input Bounded Output</i>
QRD LSL	<i>QR-Decomposition based LSL</i>

RESUMO

O uso de técnicas de processamento digital de sinais e de processamento adaptativo de sinais possibilita a resolução de problemas que dificilmente seriam resolvidos do modo analógico convencional. O desenvolvimento da microeletrônica e da teoria de análise de sistemas discretos no tempo permitiu a aplicação destas técnicas para a resolução de problemas em "tempo real", com custo economicamente viável e com algumas características vantajosas em relação a implementações analógicas.

O presente trabalho evidencia a relação entre o problema de separação de sinais e a Instrumentação. Apresenta-se em detalhes o "cancelador adaptativo de ruído". É estudada a base teórica para o desenvolvimento de estruturas e algoritmos de filtros adaptativos. As principais estruturas e algoritmos de filtros adaptativos de resposta ao impulso de duração finita (FIR) são examinadas. São apresentados algoritmos para filtros adaptativos de resposta ao impulso de duração infinita (IIR) e estrutura direta. Problemas que estas estruturas enfrentam são salientados.

Mostra-se uma aplicação da técnica de separação de sinais com filtros adaptativos ao problema de cancelamento de eco em linha telefônica. É implementado um cancelador de eco em um equipamento baseado no microprocessador TMS320C25. Os problemas de implementação em tempo real são discutidos. São feitas comparações de desempenho de algumas estruturas e algoritmos de filtros adaptativos para este problema. Medidas de desempenho do sistema implementado e sugestões para a sua melhoria são apresentadas.

ABSTRACT

The use of digital signal processing and adaptive signal processing techniques makes it possible to solve problems that would hardly be solved by conventional analog techniques. The development of microeletronic and of discrete-time systems theory permitted the application of these techniques to the solution of real-time problems, with an economically viable cost and with some advantageous characteristics over analog technology implementations.

The present work evinces the relation between the signals separation problem and Instrumentation. The adaptive noise canceller is presented in details. It is studied the theoretical basis for the development of structures and algorithms of adaptive filters. The main structures and algorithms for adaptive filters with finite duration impulse response (FIR) are examined. Algorithms are presented for adaptive filters with infinite duration impulse response (IIR) and direct structure. Problems that these structures face are pointed out.

An application of the technique of signals separation with adaptive filters to the problem of echo canceling in telephone lines is showed . An echo canceller is implemented on a TMS320C25 microprocessor based equipment. Problems of real-time implementation are discussed. Comparisons of some adaptive filter structures and algorithms performance for this problem are made. Measures of the implemented system performance and suggestions for its improvement are showed.

1 INTRODUÇÃO

1.1 O problema das entradas interferentes em um sistema de medição

A área de Instrumentação possui um caráter abrangente, tanto devido à complexidade dos sistemas reais quanto a sua larga utilização. Neste sentido se configura um caráter multidisciplinar, onde diversas áreas do conhecimento humano unem-se para a realização de uma tarefa que normalmente envolve monitoramento e controle de processos. Quando se deseja observar uma característica de um processo, devem ser escolhidos os transdutores que melhor se adaptem à situação presente e deve existir um sistema de condicionamento que conforme o sinal de saída destes transdutores para adequá-lo ao uso posterior, seja por uma pessoa ou processador.

Segundo Borchardt e Brito [3], "Instrumentação é todo o conjunto de dispositivos que interligam um processo a um processador, capaz de registrar as informações correspondentes às variáveis de interesse, com a finalidade de monitoração ou controle". Esta definição coloca a instrumentação como um elo de ligação entre um processo e a informática.

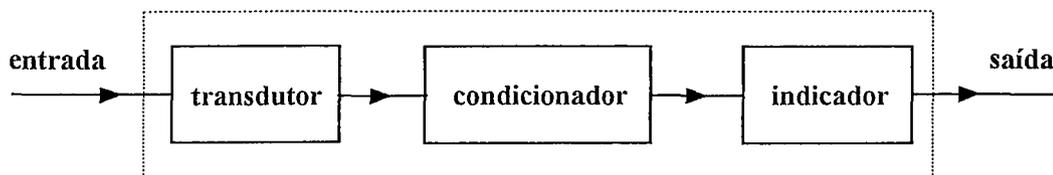


FIGURA 1.1 : Diagrama em blocos genérico de um instrumento de medição

Pode-se esquematizar os instrumentos de medição em geral como sendo constituídos por 3 estágios (blocos) principais [3] (figura 1.1). O primeiro estágio é constituído pelo transdutor sensor. Este recebe um estímulo (energia) de alguma grandeza física de interesse (temperatura, pressão, força, etc...) e geralmente realiza uma conversão deste estímulo para uma outra grandeza, como deslocamento ou tensão elétrica, por exemplo. Em geral, a grandeza de saída deste estágio é escolhida de modo a ser utilizável pelo próximo estágio. Como há mais interesse neste trabalho em instrumentos eletro-eletrônicos, a grandeza de saída do sensor será considerada elétrica. O segundo estágio é chamado de condicionador de sinal, que no caso recebe um sinal elétrico proveniente do transdutor e realiza um processamento sobre este sinal de modo a torná-lo adequado para o estágio indicador. Este processamento inclui todas as funções necessárias para a adequação do sinal, como amplificação, linearização, outras operações matemáticas, eliminação de interferências, etc. O terceiro estágio é o que

apresenta o resultado da medida, podendo ser um indicador analógico, registrador gráfico ou um sistema de aquisição de dados computadorizado.

O sistema acima toma como base a existência de uma grandeza mensurável pelo transdutor e trata de condicioná-la para que a medida seja utilizável. No entanto, quando o sensor é colocado num sistema real, ele sofre simultaneamente a influência de diversas grandezas presentes nesse sistema. Na prática, é difícil que o sensor apresente uma resposta somente a estímulos de uma determinada grandeza. Doebelin [6] modela de modo geral a configuração de entrada/saída de instrumentos/sistemas de medida conforme a figura 1.2.

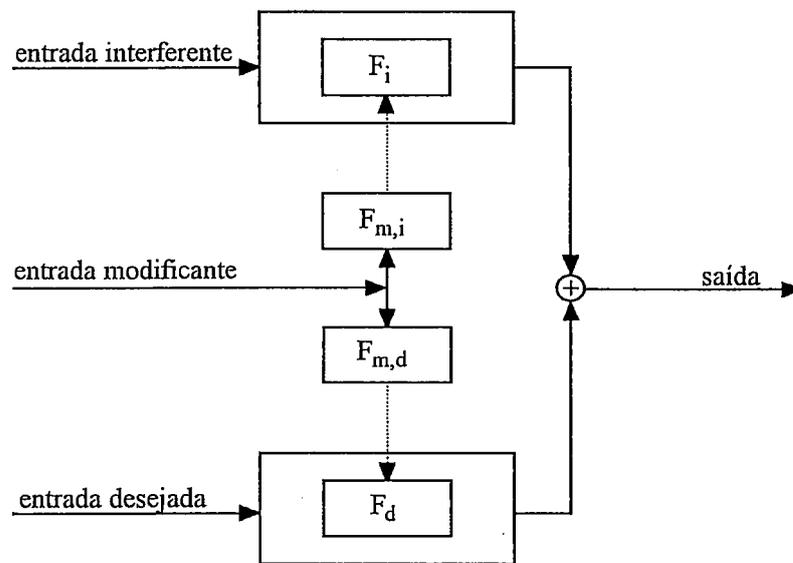


FIGURA 1.2: Configuração geral de entrada / saída de sistemas de medição

As grandezas de entrada são classificadas em 3 categorias: desejadas, interferentes e modificantes. As grandezas desejadas são aquelas que o sistema pretende medir. As interferentes são aquelas às quais o sistema é sensível, mesmo sem esta intenção. A influência destas grandezas na saída do sistema é modelada pelas funções de transferência F_d e F_i respectivamente. As grandezas de entrada modificantes são aquelas que causam uma mudança nas relações de entrada-saída para as entradas desejadas e interferentes; em outras palavras, afetam as funções de transferência F_d e F_i . Essa influência é diferente para F_d e F_i , sendo modelada por $F_{m,d}$ e $F_{m,i}$.

Esta configuração constitui um modelo para o tratamento de sinais interferentes e modificantes no sistema de medição. As entradas modificantes provocam variações nas funções de transferência dos sinais interferentes e desejados. Por outro lado, a saída do sistema é a soma das respostas às entradas interferentes e desejadas. O problema básico é obter o sinal de resposta à entrada desejada na saída do sistema, sem a influência das outras entradas. Para resolver o problema causado pela entrada interferente é preciso "separar" o sinal resultante desta entrada do sinal resultante da entrada desejada, pois os dois foram misturados na saída do sistema. A entrada modificante atua sobre a função de transferência F_d , função de

transferência do sensor para a entrada desejada; portanto, a ação da entrada modificante altera o "caminho" do sinal desejado. Fala-se então que esse é um problema de separação de sinais e de correção de caminho.

Doebelin [6] apresenta algumas soluções gerais para este problema. O "método das correções calculadas de saída" requer que se meça ou estime as magnitudes das entradas interferentes e/ou modificantes e se saiba quantitativamente como elas afetam a saída. Com esta informação é possível calcular correções que podem ser somadas ou subtraídas da saída indicada de modo a deixar (idealmente) somente o componente associado com a entrada desejada. É sugerido o uso de sensores para as entradas espúrias e um sistema eletrônico computadorizado para implementar o método de forma automática.

Outro método, filtragem de sinal, se baseia na possibilidade de introdução de "filtros" no instrumento, que bloqueiem os sinais espúrios, de modo que seus efeitos na saída sejam eliminados ou reduzidos. O filtro pode ser aplicado a qualquer sinal, seja ele de entrada, saída ou sinal intermediário. O estágio condicionador de sinais deverá minimizar a influência de entradas interferentes e modificantes na saída do instrumento. Para isso, normalmente são utilizados filtros que atuam sobre características do sinal de saída do sensor, baseados no conhecimento de características do sinal de interesse ou do(s) sinal(is) interferente(s). Os filtros usados geralmente são seletivos em frequência, ou seja, atuam sobre as características espectrais dos sinais. Com o conhecimento do espectro dos sinais interferentes e de interesse é possível projetar um filtro que atenua as frequências presentes na interferência.

Busca-se uma solução para o problema das entradas interferentes, considerado a seguir.

1.2 O cancelador adaptativo de ruído como uma solução

Pode-se considerar o problema do seguinte modo: ao ser colocado no processo, o sensor será submetido à influência da grandeza de interesse e de outras grandezas físicas presentes. O sensor gera então um sinal de saída em resposta à essa influência, segundo o modelo da figura 1.2. Considera-se essa resposta como o somatório das respostas às influências individuais das grandezas físicas presentes. Assume-se que é possível colocar um segundo sensor no sistema e que seja sensível apenas às demais grandezas influentes, chamadas interferentes por não serem de interesse. Na saída deste sensor, chamado de referência, tem-se um sinal proporcional às grandezas interferentes. A figura 1.3 ilustra essa idéia.

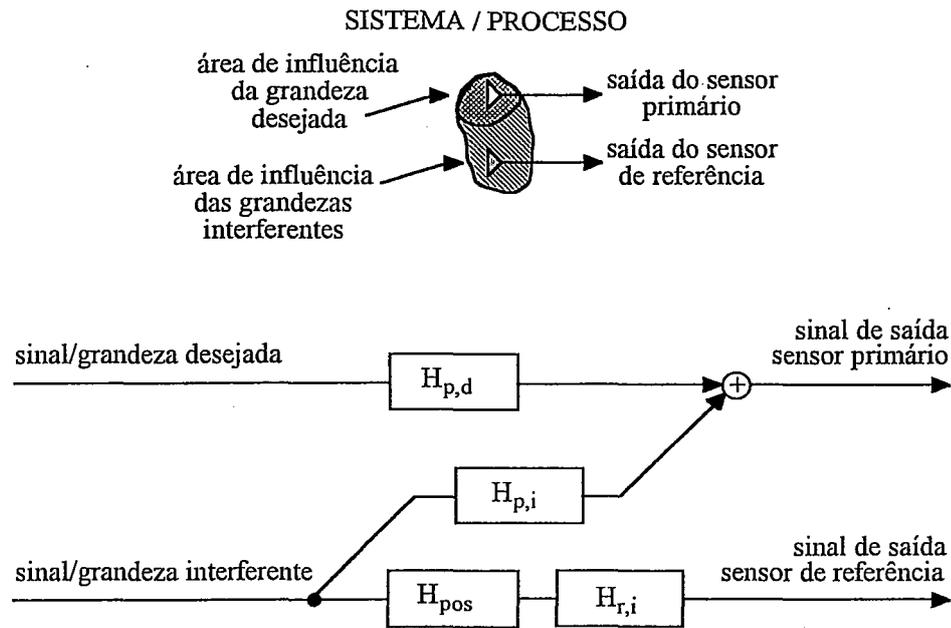


FIGURA 1.3: Modelo de sistema de medição com dois sensores em posições distintas

O modelo com funções de transferência da figura acima traduz a situação anterior de modo simplificado, indicando diferenças dos sensores e de suas posições no sistema. $H_{p,d}$ é a função de transferência do sensor primário para o sinal desejado. $H_{p,i}$ é a função de transferência do sensor primário para o sinal interferente. $H_{r,i}$ é a função de transferência do sensor de referência para o sinal interferente. H_{pos} modela a diferença de posição dos sensores no sistema físico.

Este modelo pode ser simplificado ainda mais se forem levados em consideração outros fatores. O primeiro diz respeito à função $H_{p,d}$, que deve ser conhecida pelo usuário ou projetista do instrumento de medição, já que ela representa a resposta do sensor à grandeza de interesse. A resposta deste bloco é a convolução do sinal desejado com a resposta ao impulso de $H_{p,d}$, que é o sinal esperado pelos estágios seguintes do sistema de medição. Deste modo, por sinal desejado de agora em diante se entenderá o sinal de resposta do bloco $H_{p,d}$. Outro fator é que o sinal interferente presente nos dois sensores é diferente. Essa diferença pode ser modelada por uma função de transferência $H = H_{pos}H_{r,i}/H_{p,i}$, considerando que o sinal interferente passa a ser representado pela resposta de $H_{p,i}$ ao sinal interferente da figura 1.3. O novo modelo pode ser visto na figura 1.4.

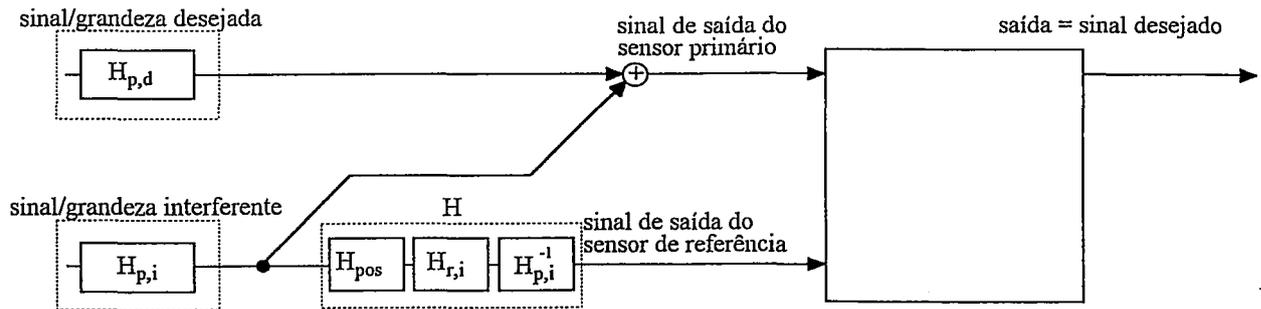


FIGURA 1.4: Modelo do problema das entradas interferentes e um sistema "caixa-preta" para sua solução

Procura-se resolver o problema das entradas interferentes através de um sistema hipotético "caixa-preta", capaz de separar o sinal de entrada interferente presente no sensor primário com o auxílio de um sensor de referência que receba apenas sinal interferente. A saída deste sistema hipotético é o sinal desejado. Um sistema deste tipo é uma implementação do método das correções calculadas para sinais interferentes. Algumas questões que podem ser feitas a seu respeito: o sensor que mede a interferência mede somente a interferência ou terá outros sinais? Será que a combinação (mistura) dos sinais que chega no sensor primário será independente do tempo? H é constante? Como sintonizar o sistema?

Se fosse possível conhecer as características da função de transferência que leva o ruído aos sensores seria possível, em geral, projetar um filtro fixo capaz de transformar o ruído presente no sensor de referência no ruído presente no sensor primário. Neste caso, o sinal de saída do filtro poderia ser subtraído do sinal primário e a saída do sistema seria o sinal desejado apenas. Este sistema poderia ser chamado de cancelador de ruído fixo.

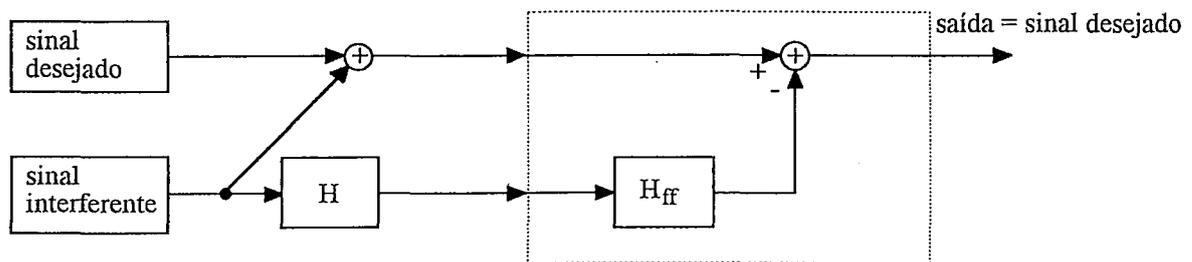


FIGURA 1.5: O problema com funções de transferência e sistema de cancelamento com filtro fixo

A figura 1.5 ilustra a idéia do "cancelador de ruído fixo", onde H_{ff} é a função de transferência do sistema hipotético para o sinal proveniente do sensor de referência, que visa transformar o sinal interferente presente no sensor de referência no mesmo sinal interferente presente no sensor primário. Para que isso seja possível deve-se ter $H_{ff}=1/H$. Assim pode-se subtrair o ruído do sinal presente no sensor primário, obtendo apenas o sinal desejado na saída

do sistema. No entanto, geralmente as características da função de transferência H são desconhecidas ou aproximadamente conhecidas e de natureza variável no tempo ($H(t)$), o que impede o uso de filtros fixos. Ainda mais, mesmo que fosse possível projetar um filtro fixo, suas características deveriam ser ajustadas com precisão e o menor erro resultaria num aumento da potência do ruído de saída, pois estaria sendo injetado sinal interferente na saída do sistema.

No sistema da figura 1.6, o sinal proveniente do sensor de referência é processado por um filtro adaptativo que ajusta automaticamente sua função de transferência H_{fa} , procurando atingir $H_{fa}(t)=1/H(t)$. Assim, com o algoritmo adequado, o filtro pode operar sobre condições variáveis e pode reajustar a si mesmo continuamente para minimizar o sinal de erro. Essa propriedade confere ao sistema cancelador de ruído a possibilidade de seguir variações de $H(t)$. Ainda mais, não é necessário conhecimento algum de $H(t)$ *a priori*, já que o algoritmo ajusta automaticamente a resposta do filtro. Essa configuração recebe o nome de cancelador adaptativo de ruído, conforme definição de Widrow [46].

Em sistemas de cancelamento de ruído o objetivo prático é produzir uma saída do sistema que é a melhor estimativa do sinal de interesse (desejado), no sentido dos mínimos quadrados. Este objetivo é atingido realimentando a saída do sistema para o filtro adaptativo e ajustando o filtro por um algoritmo adaptativo que minimize a potência total do sinal de saída do sistema. Será mostrado no capítulo 2 que estes processos são equivalentes.

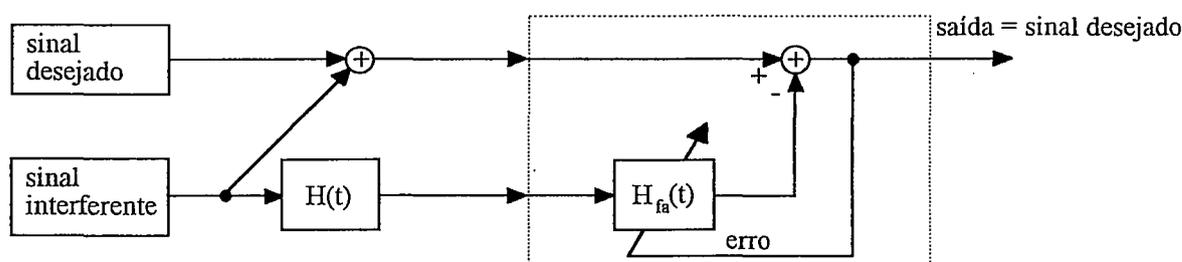


FIGURA 1.6: O cancelador adaptativo de ruído

1.3 Conceitos de filtragem ótima e filtros adaptativos

Como principal componente do cancelador adaptativo de ruído, muitas das características do sistema são dependentes das características do filtro adaptativo usado. O entendimento das características básicas desses filtros e seus algoritmos proporciona uma visão geral sobre o comportamento do sistema cancelador para as mais diversas situações. Segundo Haykin [18], filtro é um dispositivo que é aplicado a um conjunto de dados com ruído a fim de extrair informação sobre uma dada quantidade de interesse. Existem três operações, ou problemas de estimação, básicos:

1. Filtragem - extração da informação sobre a quantidade de interesse no instante t , usando informação medida até o instante t inclusive;
2. *Smoothing* (interpolação) - a informação sobre a quantidade de interesse não precisa estar disponível no instante t , e dados obtidos após este instante podem ser usados para obter esta informação. Isto significa que há um atraso na produção do resultado de interesse;
3. Predição - determinação aproximada da quantidade de interesse num instante $t + \tau$ no futuro, para $\tau > 0$, usando informação medida até o tempo t , inclusive.

Se a quantidade de interesse na saída do dispositivo for uma função linear das observações (do sinal) aplicadas à entrada do filtro diz-se que o filtro é linear.

Antes de estudar filtros adaptativos é importante que se tenha idéia de conceitos gerais sobre filtros fixos. De modo geral, não se está interessado no problema de filtros seletivos em frequência. O problema de interesse aqui é, segundo Kailath [23], como determinar a *estimativa mínima quadrática linear e causal* do sinal de um processo corrompido por ruído branco aditivo, conforme a figura 1.7. Isto é, dado um sinal desejado, u , ao qual foi adicionado ruído branco, n , deseja-se projetar um filtro para recuperar o sinal. O filtro deve ser linear, por motivos de simplicidade de análise matemática, e além disso deve ser causal, para que seja possível implementá-lo. Usa-se como critério de avaliação do filtro uma função de custo que é função do erro da estimativa feita pelo filtro, e , definido como a diferença entre o sinal desejado, u , e a saída atual do filtro, y . Essa função de custo pode ser definida usando métodos probabilísticos ou de mínimos quadrados. Este é um problema-chave de estimação, já que sua solução contém a solução de muitos outros problemas. Estudado inicialmente nas áreas de engenharia de comunicação e de teoria de informação, originou a teoria sobre filtros adaptativos.

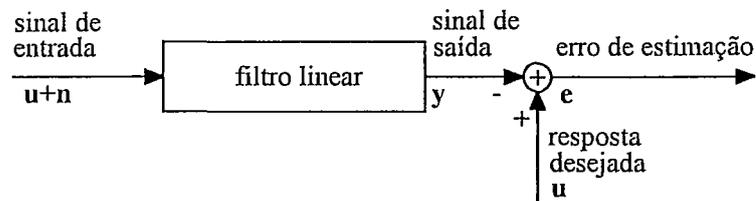


FIGURA 1.7: O problema de estimação linear

O problema pode ser atacado do ponto de vista probabilístico por dois lados. Engenheiros de comunicação e pessoas ligadas à teoria da informação são mais acostumados a casos em que informação sobre covariância é dada sobre sinal e ruído, que são usualmente chamados problemas de filtragem de Wiener. Engenheiros de controle lidam mais frequentemente com problemas onde o sinal e o ruído são descritos por *modelos de espaço de estado* (*state-space models*), usualmente chamados problemas de filtragem de Kalman [23].

Na resolução deste problema é assumido o conhecimento (disponibilidade) de certos parâmetros estatísticos (média, correlação) do sinal de interesse u e do ruído adicionado n . O objetivo é projetar um filtro linear que use este sinal ruidoso como entrada e minimize os efeitos do ruído na saída y , segundo algum critério estatístico. Um critério útil consiste em usar como função de custo o valor médio do quadrado do sinal de erro. Essa função de custo deve ser minimizada para que se tenha um filtro otimizado segundo o critério de mínima média dos quadrados (*least mean squares*).

Para sinais de entrada estacionários, a solução resultante é comumente conhecida como filtro de Wiener. O filtro de Wiener não é adequado para lidar com situações onde a não-estacionaridade do sinal de interesse ou do ruído é intrínseca ao problema. Nessas situações o filtro deve assumir uma forma variante com o tempo (Kalman).

O projeto de um filtro de Wiener necessita do conhecimento prévio das características do sinal a ser processado. O filtro é ótimo somente quando as características estatísticas dos dados de entrada coincidem com a informação prévia na qual o projeto do filtro foi baseado. Quando esta informação não é conhecida completamente pode não ser possível projetar o filtro de Wiener ou o projeto pode não ser ótimo. Neste caso, uma abordagem onde é feita uma estimativa dos parâmetros estatísticos dos sinais relevantes pode ser usada. Esta estimativa é usada para obter uma fórmula não-recursiva para calcular os parâmetros do filtro.

Outra abordagem é usar um filtro adaptativo. É um dispositivo "auto-projetante", i.e., ajusta sua própria resposta ao impulso, já que se baseia num algoritmo recursivo que lhe permite operar satisfatoriamente num ambiente onde conhecimento completo das características do sinal de interesse não está disponível. O algoritmo parte de um predeterminado conjunto de condições iniciais, representando completo desconhecimento do meio. Num ambiente estacionário, após sucessivas iterações do algoritmo, ele converge à solução ótima de Wiener. Em um ambiente não-estacionário, o algoritmo oferece uma capacidade de rastreamento (*tracking*), podendo seguir variações do comportamento estatístico do sinal de entrada desde que estas sejam lentas o suficiente.

Como conseqüência direta da aplicação de um algoritmo recursivo, onde os parâmetros do filtro adaptativo são atualizados de uma iteração para a próxima, os parâmetros se tornam dependentes do sinal. Isto significa que um filtro adaptativo é um dispositivo não-linear no sentido que ele não obedece ao princípio da superposição. De outro modo, um filtro adaptativo é freqüentemente referido como linear no sentido que a estimativa da quantidade de interesse é obtida adaptativamente (na saída do filtro) como uma combinação linear do conjunto de observações (dados) aplicado à entrada do filtro [18].

A operação de um algoritmo de filtragem adaptativo envolve dois processos básicos: um processo de filtragem, destinado a produzir uma saída em resposta a uma seqüência de dados de entrada, e um processo adaptativo, cujo propósito é proporcionar um controle adaptativo de um conjunto de parâmetros ajustáveis usados no processo de filtragem. Estes

processos trabalham iterativamente um com o outro. A escolha da estrutura para o processo de filtragem tem um profundo efeito na operação do algoritmo como um todo.

1.4 Motivação para a abordagem do problema

O estágio condicionador de sinal do instrumento de medição, ao operar sobre o sinal proveniente do sensor, realiza operações definidas no campo de processamento de sinais. Este campo está preocupado com a representação, transformação e manipulação de sinais e da informação que eles contêm [33].

A área de processamento de sinais sempre se beneficiou de uma proximidade entre teoria, aplicações e tecnologias para implementação de sistemas de processamento de sinais. Até os anos 60, a tecnologia para processamento de sinais era quase que exclusivamente analógica de tempo contínuo. A rápida evolução dos computadores digitais, microprocessadores e da microeletrônica em geral, aliados com alguns desenvolvimentos teóricos importantes, deu origem ao campo de processamento digital de sinais (DSP - *Digital Signal Processing*).

Em DSP, os sinais são considerados como uma seqüência de amostras de números de precisão finita e o processamento é implementado usando computação digital. No caso de sinais contínuos, é necessária uma conversão para o sistema digital (conversão A/D) e uma posterior conversão de volta (conversão D/A). É desejável a operação de tais sistemas em "tempo real", onde as amostras de saída são calculadas à mesma taxa de amostragem do sinal de entrada, sendo tais sistemas bastante difundidos hoje em dia graças ao avanço da tecnologia de microeletrônica e microprocessadores [33].

O uso de técnicas de DSP possui diversas vantagens sobre implementações analógicas, como tolerância aos componentes e reproducibilidade, insensibilidade ao meio (p.ex., umidade, temperatura), precisão controlada, etc. Como desvantagens, pode-se citar a limitação da freqüência do sinal (banda) e a necessidade de conversão A/D e D/A. No entanto, com o avanço da tecnologia digital, a limitação em freqüência começa a ter seus limites cada vez mais ampliados.

Outro campo importante ligado a esta área é o de processamento adaptativo de sinais. Sistemas adaptativos representam uma classe de sistemas variantes no tempo e, de certa forma, não-lineares. Apesar disso, possuem grande aplicação atualmente, além de serem baseados em técnicas bem estabelecidas de projeto e análise [49].

Até o momento, procurou-se mostrar como está ligado com a área de instrumentação o assunto de separação de sinais usando filtros adaptativos. No entanto as aplicações desta técnica não se resumem a esta área e nem a este contexto. Alguns exemplos de aplicações são: condicionamento de sinais em instrumentação, cancelamento de eco em linhas telefônicas,

cancelamento de interferência acústica em gravações, cancelamento ativo de interferência acústica, cancelamento de interferência em antenas, detecção de sinais e análise espectral, etc... Diversos autores [49] [18] apresentam aplicações de filtros adaptativos.

A implementação dos filtros adaptativos estudados será tratada de forma digital. A simulação de tais sistemas com o uso de computadores fica facilitada, pois se modela um sistema de mesma natureza. Neste aspecto, deve ser ressaltado que é muito difícil a implementação de certos algoritmos em tecnologia analógica, ao passo que a tecnologia digital traz facilidades que permitem a implementação de sistemas em "tempo real" que utilizam o computador como plataforma.

A implementação de sistemas adaptativos desta forma possibilita a análise de diversos algoritmos. Os sistemas mais importantes e estudados são os de resposta ao impulso finita (FIR - *finite impulse response*). Sistemas de resposta ao impulso de duração infinita (IIR - *infinite impulse response*) possuem problemas próprios que impedem sua aplicação em larga escala.

Com estes fatores em mente, o presente trabalho busca (1) salientar e divulgar a importância do uso de processamento digital de sinais e processamento adaptativo de sinais, (2) apresentar um panorama sobre o desenvolvimento na área de processamento adaptativo voltado à implementação e análise de filtros adaptativos e (3) apresentar estes resultados na implementação de um método de separação de sinais e aplicação a um caso prático. Embora o caso escolhido seja da área de telecomunicações, sua transposição para a Instrumentação é imediata e direta, sendo indicada em casos onde interferências de características não conhecidas *a priori* atuem sobre o sinal a ser condicionado e onde seja possível usar um sensor de referência.

Outras técnicas utilizando "máxima semelhança" (*maximum likelihood*) e "redes neurais" relacionadas ao problema de cancelamento de interferência são citadas na literatura [12] [7]. Formas e implementações singulares de filtros adaptativos podem ser vistas em diversas referências [43] [25] [44] [51].

1.5 Organização do trabalho

Este trabalho está organizado em cinco partes principais. No capítulo 1, introdução, motiva-se o estudo de filtros adaptativos e são expostas algumas idéias básicas do assunto de separação de sinais aplicadas ao cancelamento de interferências em instrumentos de medição. No capítulo 2, fundamentação, busca-se expor a base teórica indispensável para o entendimento do capítulo seguinte. Apresenta-se em maiores detalhes o cancelador adaptativo de ruído, conforme definido por Widrow [46] em 1975, que é a uma das principais estruturas usadas em processamento adaptativo de sinais. Também é estudada a base teórica para o

desenvolvimento de estruturas e algoritmos de filtros adaptativos. Esta consiste de conceitos nos campos de filtros e algoritmos de busca. Conceitos relevantes de teoria de probabilidades são definidos.

No capítulo 3 tenta-se dar uma visão global das principais estruturas e algoritmos de filtros adaptativos. São examinadas estruturas direta e *lattice*, algoritmos *Recursive Least Squares* e *Gradient Descent*, para filtros de resposta ao impulso de duração finita (FIR). Para filtros de resposta ao impulso de duração infinita (IIR) são apresentados alguns algoritmos para a forma direta e problemas que estas estruturas enfrentam. Busca-se expor as vantagens e desvantagens associadas a cada um destes casos. É feita uma comparação através de exemplo para todos os casos.

O capítulo 4, implementação, mostra uma aplicação da técnica ao problema de cancelamento de eco em tempo real em linha telefônica. É implementado um cancelador de ruído adaptativo sendo ressaltados os problemas de implementação em tempo real com o equipamento utilizado. São feitas medidas de performance de vários tipos de filtros adaptativos usados neste problema. Problemas de outras estruturas de filtros são citados. No capítulo 5, conclusão, são feitas sugestões para a melhoria de sistemas de cancelamento de eco. Os apêndices e anexos contêm assuntos que complementam o texto ou que, de outro modo, desviariam o enfoque dado.

2 FUNDAMENTAÇÃO

2.1 Introdução

Este capítulo busca expor alguns conceitos básicos que serão necessários para o desenvolvimento da teoria de filtros adaptativos e do cancelador adaptativo de ruído. Como esta teoria é, em grande parte, desenvolvida usando ferramentas de teoria de probabilidade e estatística, torna-se difícil seu entendimento sem algumas destas noções. O apêndice I apresenta uma visão sobre os aspectos de interesse neste trabalho. A extensão para a análise de sinais e o uso de transformadas para a análise frequencial é apresentada neste capítulo. Estruturas de filtros digitais que são de interesse para o processamento adaptativo são abordadas com o objetivo de familiarização e estabelecimento de notação.

Além destas noções é apresentado o problema de filtragem ótima linear. É mostrada a resolução do problema do ponto de vista estatístico e pelo método dos mínimos quadrados. A solução é apresentada para filtros transversais FIR por ser mais simples e ter resultados mais significativos. O importante e básico conceito de superfície de performance da função de custo é estabelecido para este caso. A forma dessa superfície é de importância para o desenvolvimento de algoritmos de filtros adaptativos porque possui um único mínimo global e uma dependência quadrática para os coeficientes do filtro. O chamado princípio da ortogonalidade [18] [34] [42] também é introduzido. Induz-se o desenvolvimento das equações de Wiener-Hopf para o caso discreto. Estas equações serão usadas na análise do cancelador adaptativo de ruído e também na análise e desenvolvimento de algoritmos para filtros adaptativos. Esse desenvolvimento é feito por dois métodos equivalentes, mas que originam algoritmos de características diferentes para o processo de adaptação de filtros. O estudo das características destes algoritmos, baseados na busca do mínimo pelo gradiente ou na inversão da matriz de autocorrelação, será feito no capítulo 3, com base nas definições do presente capítulo.

O cancelador adaptativo de ruído é apresentado e são estudadas as influências de ruídos não-correlacionados e de *leakage* (presença de componentes do sinal desejado na entrada de referência). Fórmulas são derivadas para prover uma estimativa do comportamento do sistema nestas condições. Conceitos de predição linear são definidos para prover uma base ao entendimento de filtros adaptativos de estrutura *lattice*.

2.2 Processos estocásticos

2.2.1 Sinais amostrados versus processos estocásticos

No contexto de processamento digital de sinais, a primeira etapa de um sistema que trabalhe com sinais preocupa-se com o condicionamento e com a conversão analógico-digital dos mesmos. Como resultado final deste processo freqüentemente resulta uma seqüência de números que representa o sinal em instantes de tempo igualmente espaçados (período de amostragem constante). Esta seqüência pode então ser escrita como $x[t_0]$, $x[t_1]$, $x[t_2]$, ..., $x[t_{N-1}]$ ou, simplesmente $x[0]$, $x[1]$, $x[2]$, ..., $x[N-1]$.

No caso de sinais ditos determinísticos, ou seja, que podem ser previstos exatamente por alguma função matemática, o comportamento do sinal e sua análise são simplificados. No entanto, para sinais que não podem ser descritos por uma função matemática devido à sua complexidade ou ao desconhecimento de sua função geradora, é necessário o uso de ferramentas de análise mais poderosas. Este caso é o mais encontrado em aplicações práticas, como análise de sinais de voz, radar e sinais que contém informação em geral [42]. O uso de ferramentas estatísticas permite analisar o problema de acordo com modelos probabilísticos e com uma teoria bem estabelecida. No entanto, a aplicação direta desta teoria não é possível e exige o uso de simplificações, levando em conta as situações reais. O uso desta teoria com sinais presentes nestas situações é tratado a seguir.

Uma seqüência amostrada em particular pode ser vista como fazendo parte de uma seqüência definida para todo o tempo, de $-\infty$ a $+\infty$. Escreve-se então $x[n]$, para $-\infty \leq n \leq +\infty$, e chama-se esta seqüência de série temporal. Uma série temporal é vista, do ponto de vista estatístico, como uma realização de um processo estocástico, ou seja, ela é considerada como um dos membros da família de funções que compõem um processo estocástico [APÊNDICE I].

Um processo estocástico é caracterizado por um conjunto de funções de distribuição ou densidade de probabilidade. Para a sua análise, é necessário que estas funções sejam estimadas. Como não é possível determinar na prática, por meio de medidas adequadas, estas funções [18], procede-se a uma caracterização parcial do processo estocástico, especificando os seus momentos de primeira e segunda ordem, tais como média e correlação [APÊNDICE I]. No entanto, estes momentos ou médias são definidos "através do processo", ou seja, para que sejam calculados, se deve ter diversas realizações do processo. Isto impede o uso de apenas uma seqüência para a análise. Na prática é interessante que se possa estimar estes parâmetros de apenas uma única realização ou seqüência.

Essa simplificação baseia-se no uso de médias temporais, ou seja, médias definidas ao longo do processo estocástico para cada função membro. Estas médias fazem uso da seqüência dada para todo o tempo, de $-\infty$ a $+\infty$. Para processos estocásticos ditos ergódicos, as médias temporais igualam as médias do conjunto (através do processo). Portanto, para essa classe de processos estocásticos é válido usar médias temporais calculadas de uma única seqüência em lugar de médias probabilísticas. Processos ergódicos devem ser estacionários. No entanto, nem todos processos estacionários são ergódicos [34] [42]. Para o caso de processos não-estacionários (e, portanto, não ergódicos) não é possível substituir a média probabilística pela média temporal. Para uma certa classe de processos não-estacionários conhecidos como processos localmente estacionários, entretanto, torna-se razoável estimar estes parâmetros com uma média curta no tempo [30]. Exemplo de tais processos são sinais de voz e EEG (eletroencefalograma).

O conceito de médias temporais, por ser definido para todo o tempo, impossibilita o seu emprego direto em aplicações práticas. Para tanto são construídos "estimadores" para estas médias, usando um conjunto finito de amostras. A análise destes estimadores é um problema de estatística [42] [34]. A seguir são apresentadas as definições de médias temporais e estimadores, onde estimativas são denotadas pelo acento circunflexo e médias temporais por $\langle \rangle$ [33].

i) média:

$$\langle x[n] \rangle = \lim_{L \rightarrow \infty} \frac{1}{2L+1} \sum_{n=-L}^{+L} x[n] \quad (2.1)$$

$$\hat{\eta}_x = \frac{1}{L} \sum_{n=0}^{L-1} x[n] \quad (2.2)$$

ii) variância:

$$\hat{\sigma}_x^2 = \frac{1}{L} \sum_{n=0}^{L-1} |x[n] - \hat{\eta}_x|^2 \quad (2.3)$$

iii) autocorrelação:

$$\langle x[n+m] x^*[n] \rangle = \lim_{L \rightarrow \infty} \frac{1}{2L+1} \sum_{n=-L}^{+L} x[n+m] x^*[n] \quad (2.4)$$

$$\hat{R}_x[m] = \langle x[n+m] x^*[n] \rangle_L = \frac{1}{L} \sum_{n=0}^{L-1} x[n+m] x^*[n] \quad (2.5)$$

Makhoul [30] omite a divisão por L , no contexto de predição linear.

2.2.2 Representação freqüencial de processos estocásticos

Embora sinais estocásticos não sejam absolutamente somáveis ou quadrado somáveis (o somatório de $-\infty$ a $+\infty$ do valor ou do seu quadrado não é garantidamente menor que $\pm\infty$) e, conseqüentemente, não possuem transformada de Fourier, muitas de suas propriedades podem ser resumidas em termos das seqüências de autocovariância e autocorrelação (seqüências aperiódicas), para as quais a transformada de Fourier freqüentemente existe. A transformada de Fourier da seqüência de autocorrelação tem uma interpretação útil em termos da distribuição freqüencial da potência do sinal. Outra vantagem é que o efeito de um sistema linear discreto no tempo sobre um sinal estocástico pode ser descrito convenientemente pelo efeito deste sistema sobre a seqüência de autocorrelação [33].

A transformada de Fourier da seqüência de autocorrelação de um processo estocástico é chamada de função densidade espectral de potência (*power spectrum density function*) e é denotada por $S_x(e^{j\omega})$. A seguir é definida a função e a sua inversa.

$$S_x(e^{j\omega}) = \sum_{l=-\infty}^{+\infty} R_x[l] e^{-j\omega l} \quad (2.6)$$

$$R_x[l] = \frac{1}{2\pi} \int_{-\pi}^{+\pi} S_x(e^{j\omega}) e^{j\omega l} d\omega \quad (2.7)$$

O gráfico de $S_x(e^{j\omega})$ por ω é chamado de espectro de densidade de potência (*power density spectrum*) (ou simplesmente espectro de potência ou espectro) do processo estocástico, e seu valor numa dada freqüência ω é chamado de densidade espectral de potência (*power spectral density*). A razão para o termo densidade de potência torna-se clara notando que a potência média no processo estocástico $x[n]$ é dada por:

$$\mathcal{E}\{[x[n]]^2\} = R_x[0] = \frac{1}{2\pi} \int_{-\pi}^{+\pi} S_x(e^{j\omega}) e^{j\omega l} d\omega \quad (2.8)$$

Como a integral representa potência, $S_x(e^{j\omega})$ deve ser densidade de potência. Se a integral é tomada sob uma banda de freqüências, o resultado é proporcional à potência do sinal naquela banda. As propriedades fundamentais estão listadas na tabela abaixo.

TABELA 2.1: Propriedades da função densidade de potência espectral

Propriedades de $S_x(e^{j\omega})$
$S_x(e^{j\omega})$ é real
$S_x(e^{j\omega}) \geq 0$
Se o processo é real:
$S_x(e^{j\omega}) = S_x^*(e^{-j\omega})$

Uma descrição das relações estatísticas entre dois processos estocásticos é dada pela função densidade de potência espectral cruzada (*cross-power spectral density function*), definida como a transformada de Fourier da função de correlação cruzada. A definição da função e de sua inversa é apresentada a seguir.

$$S_{xy}(e^{j\omega}) = \sum_{l=-\infty}^{+\infty} R_{xy}[l]e^{-j\omega l} \quad (2.9)$$

$$R_{xy}[l] = \frac{1}{2\pi} \int_{-\pi}^{+\pi} S_{xy}(e^{j\omega}) e^{j\omega l} d\omega \quad (2.10)$$

O valor desta função é chamado de espectro de potência cruzado (*cross-power spectrum*) ou espectro cruzado dos dois processos estocásticos $x[n]$ e $y[n]$. Observando a definição da função inversa (2.10), pode-se dar a interpretação de que $S_{xy}(e^{j\omega_0})$ mede a correlação entre dois processos estocásticos numa dada frequência ω_0 [42]. Para processos estocásticos reais (não-complexos), a magnitude de $S_x(e^{j\omega})$ é par e a fase é ímpar. Outras propriedades são dadas na tabela abaixo.

TABELA 2.2: Propriedades da função densidade de potência espectral cruzada

Propriedades de $S_{xy}(e^{j\omega})$
$S_{xy}(e^{j\omega})$ é complexa
$S_{xy}(e^{j\omega}) = S_{yx}^*(e^{j\omega})$
Se o processo é real:
$S_{xy}(e^{j\omega}) = S_{xy}^*(e^{-j\omega})$

Estas propriedades podem ser estendidas para o domínio Z , considerando que as avaliações são tomadas sobre o círculo de raio unitário. A tabela 2.3 lista as propriedades destas funções e suas definições.

TABELA 2.3: Propriedades em Z das funções de densidade de potência espectral

Propriedades de $S_x(z)$	Propriedades de $S_{xy}(z)$
$S_x(z) = \sum_{l=-\infty}^{+\infty} R_x[l]z^{-l}$	$S_{xy}(z) = \sum_{l=-\infty}^{+\infty} R_{xy}[l]z^{-l}$
$S_x(z) = S_x^*(1/z^*)$	$S_{xy}(z) = S_{yx}^*(1/z^*)$
Se o processo é real: $S_x(z) = S_x(z^{-1})$	Se o processo é real: $S_{xy}(z) = S_{yx}(z^{-1})$

2.2.3 Efeito de sistemas lineares invariantes no tempo

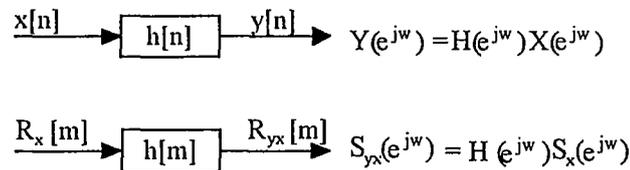


FIGURA 2.1: Efeito de sistema LTI sobre sinais determinísticos e estocásticos

Considere um sistema linear invariante no tempo (LTI - *linear time-invariant*) [33] e estável com resposta ao impulso $h[n]$. A entrada para este sistema é um processo estocástico $x[n]$ discreto no tempo e estacionário no sentido amplo (WSS). A saída do sistema, $y[n]$, é uma função única de um processo estocástico, relacionado à entrada pela seguinte transformação linear:

$$y[n] = \sum_{k=-\infty}^{+\infty} h[n-k] x[k] = \sum_{k=-\infty}^{+\infty} h[k] x[n-k] = h[n] * x[n] \quad (2.11)$$

Para um sistema LTI excitado por um processo WSS [APÊNDICE I], a saída também é WSS [42] [33] [34]. Calculando a autocorrelação do sinal de saída do filtro e aplicando Fourier, obtém-se:

$$S_y(e^{j\omega}) = |H(e^{j\omega})|^2 S_x(e^{j\omega}) \quad (2.12)$$

onde $|H|$ representa o módulo de H . Observando a correlação cruzada entre a entrada e a saída de um sistema linear invariante no tempo, nota-se que esta é a convolução da resposta ao impulso com a seqüência de autocorrelação [42] [33] [34]. Aplicando Fourier, chega-se a:

$$S_{xy}(e^{j\omega}) = H^*(e^{j\omega}) S_x(e^{j\omega}) \quad (2.13)$$

Com estas equações observa-se que é possível estimar a resposta ao impulso (e a função de transferência) de um sistema LTI se for possível observar sua resposta à ruído branco de média zero e variância σ^2 , pois o espectro de potência cruzada é proporcional à resposta em frequência do filtro, com constante de proporcionalidade σ^2 [33]. Os resultados importantes dessa análise podem ser resumidos na figura 2.1 e na tabela 2.4 [42].

TABELA 2.4: Relações de transformação válidas em sistemas LTI para processos WSS

$R_{yx}[l]=h[l]*R_x[l]$	$S_{yx}(e^{j\omega})=H(e^{j\omega})*S_x(e^{j\omega})$	$S_{yx}(z)=H(z)S_x(z)$
$R_{xy}[l]=h^*[-l]*R_x[l]$	$S_{xy}(e^{j\omega})=H^*(e^{j\omega})S_x(e^{j\omega})$	$S_{xy}(z)=H^*(1/z^*)S_x(z)$
$R_y[l]=h[l]*R_{xy}[l]$	$S_y(e^{j\omega})=H(e^{j\omega})S_{xy}(e^{j\omega})$	$S_y(z)=H(z)S_{xy}(z)$
$R_y[l]=h[l]*h^*[-l]*R_x[l]$	$S_y(e^{j\omega})= H(e^{j\omega}) ^2S_x(e^{j\omega})$	$S_y(z)=H(z)H^*(1/z^*)S_x(z)$

2.3 Estruturas para implementação de filtros digitais

Esta seção está interessada no estabelecimento de notação para as estruturas de filtros digitais mais utilizadas em processamento adaptativo, que são as estruturas direta e *lattice*. Estruturas de filtros digitais em geral são tratadas em Oppenheim [33]. Estruturas *Systolic arrays* são vistas por Haykin [18].

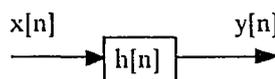


FIGURA 2.2: Representação de um sistema LTI, onde $y[n]=h[n]*x[n]$

Um sistema LTI discreto no tempo tem sua relação de entrada/saída caracterizada de modo equivalente pela equação de diferenças, pela resposta ao impulso e pela função de transferência. Considera-se que a relação de entrada/saída do filtro (sistema) pode ser expressa por

$$y[n] = \sum_{k=1}^N a_k y[n-k] + \sum_{k=0}^M b_k x[n-k] \quad (2.14)$$

onde $y[n]$ é a saída do sistema, $x[n]$ é a sua entrada e a_k e b_k representam os seus parâmetros. Estes parâmetros podem ser melhor interpretados se for tomada a transformada Z da relação acima

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^M b_k z^{-k}}{1 - \sum_{k=1}^N a_k z^{-k}} \quad (2.15)$$

onde nota-se que a função de transferência é uma razão de polinômios em z^{-1} , sendo as raízes do numerador chamadas de zeros e as raízes do denominador de pólos do sistema. Essa função pode ser expressa de modo não-causal, mas equivalente, pela relação

$$y[n] = \sum_{k=-\infty}^{+\infty} h[k]x[n-k] = h[n] * x[n] \quad (2.16)$$

onde os coeficientes $h[k]$ representam a resposta ao impulso do sistema. Para uma representação causal, o limite inferior do somatório deve ser feito igual a zero. Neste caso, o sistema possui uma resposta ao impulso de duração infinita (IIR) e causal. Se, além disso, o limite superior for definido, o filtro passa a ter uma resposta ao impulso de duração finita (FIR).

A chamada "estrutura de forma direta" ou linha de atraso (*tapped delay line*) é uma estrutura que implementa um sistema de resposta ao impulso de duração finita (FIR). A função de transferência deste tipo de filtro possui somente zeros. A resposta à seqüência de entrada é dada por (2.17)

$$y[n] = \sum_{k=0}^M b_k x[n-k] \quad (2.17)$$

e sua implementação pode ser vista na figura (2.3). Neste caso, a resposta ao impulso do sistema é igual aos coeficientes b_k , ou seja, $h[k]=b_k$ para $0 \leq k \leq M$.

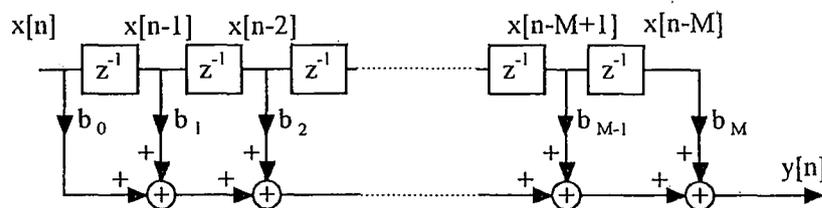


FIGURA 2.3: Filtro digital com estrutura de forma direta

Esta estrutura é a mais comumente encontrada em processamento adaptativo de sinais devido à simplicidade de implementação, existência de algoritmos simples para o ajuste dos coeficientes e porque sistemas FIR são implicitamente estáveis. O uso desta estrutura para a

implementação de sistemas IIR adaptativos [43] [13] [48] [22] [35] é possível de acordo com o seguinte esquema:

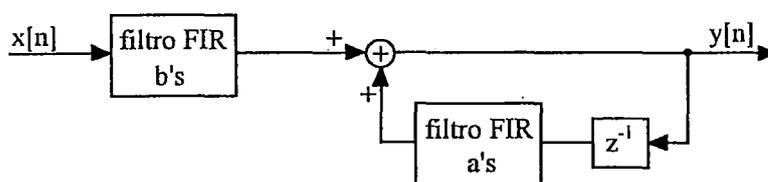


FIGURA 2.4: Filtro IIR implementado com 2 filtros FIR

Uma outra forma de implementação de filtros é através da estrutura *lattice* [15] [33] [1]. Apesar de estas estruturas serem matematicamente menos eficientes, pois requerem um volume de cálculo maior, possuem propriedades interessantes como baixa sensibilidade à variação de parâmetros e efeitos de quantização. Do ponto de vista adaptativo outras propriedades também são de interesse. A forma genérica para a implementação de uma função de transferência com pólos e zeros baseada neste tipo de estrutura é mostrada na figura (2.5). Cada estágio ou seção pode ser implementado de diversas maneiras [15] e a estrutura pode ser normalizada para que os sinais que se propagam e os parâmetros do filtro não excedam a unidade em módulo. Estruturas similares a esta podem ser utilizadas em filtros IIR adaptativos [37] [36] [9].

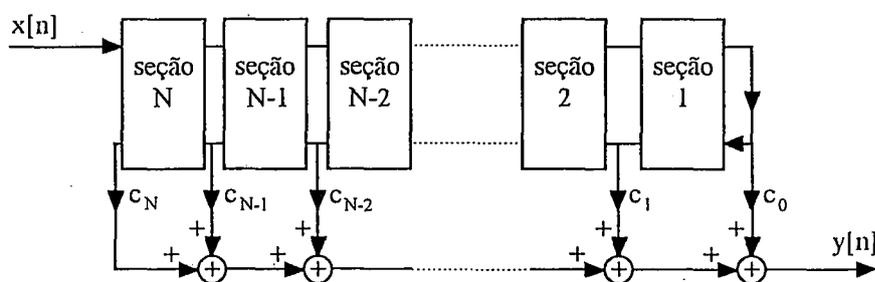


FIGURA 2.5: Filtro IIR utilizando estrutura *lattice*

2.4 O problema de filtragem ótima linear e sua solução

2.4.1 Introdução

Considere o diagrama da figura 2.6, construído em torno de um filtro linear discreto no tempo. As amostras de entrada do filtro constituem uma série temporal $u[0], u[1], u[2], \dots$ e o filtro é caracterizado por uma resposta ao impulso w_0, w_1, w_2, \dots . Num certo instante discreto de tempo n , o filtro produz uma saída, $y[n]$, usada para prover uma estimativa de uma resposta desejada, $d[n]$. A diferença entre $d[n]$ e $y[n]$ é chamada de erro de estimação. A princípio não

são feitas restrições sobre o tipo de filtro, estrutura e causalidade. O objetivo é projetar um filtro que forneça uma saída, $y[n]$, o mais próximo possível da resposta desejada, $d[n]$. Para isso procura-se minimizar uma função do sinal de erro, $e[n]$, chamada função de custo.

Dependendo de como assumimos ser os sinais da figura 2.6 teremos dois modos de resolver o problema. O primeiro modo assume que é conhecido um número finito de amostras do sinal $u[n]$ e dos demais sinais. O objetivo é encontrar a melhor estimativa para a resposta desejada assumindo como função de custo a soma dos quadrados do sinal de erro. Esta é a solução pelo método dos mínimos quadrados. O segundo método supõe que a seqüência de entrada é conhecida para todo tempo e não somente num intervalo finito, e assume também que os sinais são parte de um processo estocástico. A função de custo é definida como a média do quadrado do sinal de erro. Esta é a solução de Wiener. A solução destes problemas possui muito em comum. Em ambos os casos o erro quadrado médio pode ser escrito em termos das funções de autocorrelação e correlação cruzada dos sinais $u[n]$ e $d[n]$, tendo pequena diferença na maneira de como são definidas estas funções. Dadas estas funções e \mathbf{R} e \mathbf{p} (definidos adiante), os métodos para calcular a resposta ao impulso ótima são iguais.

O problema principal é que o critério dos mínimos quadrados não nos permite quantificar quão similares duas formas de onda determinísticas são [43]. Isto quer dizer que para um desvio de uma amostra na janela (conjunto de amostras observado) não há garantias de que os vetores de coeficientes ótimos estejam perto um do outro, isto é, de que as solução sejam semelhantes. É virtualmente impossível generalizar sobre esta solução [43]. No mundo real tem-se apenas uma realização e, portanto, se deveria considerar somente a solução dos mínimos quadrados. Cria-se então o artifício analítico de um processo estocástico, uma realização do qual é nossa seqüência observada. Assim pode-se prever a solução ótima de um grande número de possíveis realizações, incluindo a observada. De modo genérico, prova-se teoremas sobre o problema da média do quadrado, mas, na prática, se resolve o problema da soma dos quadrados [43].

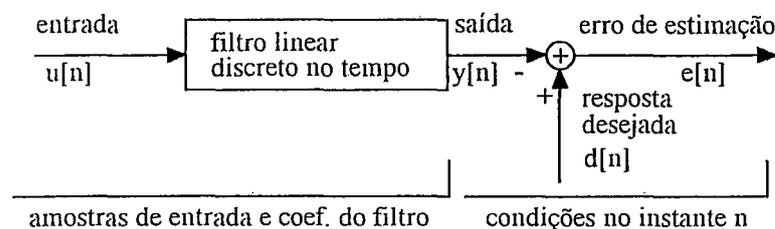


FIGURA 2.6: Representação do problema de filtragem ótima linear em diagrama de blocos

2.4.2 Filtros de Wiener

2.4.2.1 História

Os primeiros estudos sobre estimação da mínima média do quadrado (*minimum mean-square estimation*) em processos estocásticos foi feita por Komolgorov, Krein e Wiener nas décadas de 30 e 40. Wiener formulou o problema de predição linear para tempo contínuo e derivou uma fórmula explícita para o preditor ótimo. Wiener também considerou o problema de estimar um processo corrompido por um processo ruidoso aditivo. A fórmula explícita para a estimativa ótima requer a solução de uma equação integral conhecida como *Equação de Wiener-Hopf*. Em 1947, Levinson formulou o problema de filtragem de Wiener em tempo discreto e a equação de Wiener-Hopf toma uma forma matricial simples [18] [23].

2.4.2.2 O problema de filtragem ótima linear do ponto de vista estatístico

Considere o diagrama da figura 2.6, construído em torno de um filtro linear discreto no tempo. Visando a facilidade matemática e a implementação digital, são impostas restrições iniciais como o uso de um filtro causal, de estrutura transversal e resposta ao impulso de duração finita. Assume-se que a entrada do filtro $u[n]$ e a resposta desejada são realizações únicas de processos estocásticos conjuntamente WSS e discretos no tempo, ambos com média zero [APÊNDICE I]. O erro de estimação é uma amostra de uma variável aleatória. O problema consiste em fazer este erro tão pequeno quanto possível em um dado senso estatístico. O critério estatístico usado na otimização deve levar a uma matemática tratável, sendo escolhido a média do quadrado do erro de estimação.

O problema é, então, como projetar um filtro linear causal discreto no tempo, de estrutura transversal, cuja saída $y[n]$ forneça uma estimativa da resposta desejada $d[n]$, dado um conjunto de amostras de entrada $u[n]$, tais que o valor da média do quadrado do erro de estimação $e[n]$, definido como a diferença entre a resposta desejada $d[n]$ e a resposta atual $y[n]$, seja minimizada.

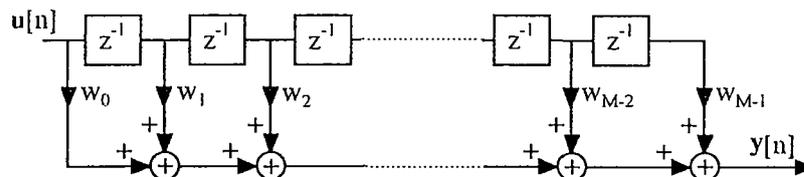


FIGURA 2.7: Filtro transversal usado na solução do problema de filtragem linear

2.4.2.3 Função de custo

Assumindo que a série temporal e os coeficientes do filtro podem ter valores complexos e que o filtro possui M coeficientes (figura 2.7), a sua saída é dada por:

$$y[n] = \sum_{k=0}^{M-1} w_k u[n-k] \quad (2.18)$$

onde $n=0,1,2,\dots$

Para otimizar o projeto do filtro minimiza-se a função de custo J , definida como o valor médio do quadrado do erro $e[n]$:

$$J = \varepsilon\{|e[n]|^2\} \quad (2.19)$$

Denote o k -ésimo coeficiente do filtro por $w_k = a_k + jb_k$, onde $k=0,1,2,\dots,M-1$. Substituindo a fórmula para a saída do filtro (2.18) na definição do erro de estimação:

$$e[n] = d[n] - \sum_{k=0}^{M-1} w_k u[n-k] \quad (2.20)$$

Desenvolvendo a função de custo J (2.19) utilizando 2.20:

$$J = \varepsilon\{e[n] e^*[n]\} \quad (2.21)$$

$$J = \varepsilon\{|d[n]|^2\} - \sum_{k=0}^{M-1} w_k \varepsilon\{u[n-k] d^*[n]\} - \sum_{k=0}^{M-1} w_k^* \varepsilon\{u^*[n-k] d[n]\} + \sum_{k=0}^{M-1} \sum_{i=0}^{M-1} w_k w_i^* \varepsilon\{u[n-k] u^*[n-i]\} \quad (2.22)$$

As quatro esperanças estatísticas desta equação representam:

- i) variância da resposta desejada $d[n]$, assumida com média zero:

$$\varepsilon\{|d[n]|^2\} \triangleq \sigma_d^2 \quad (2.23)$$

- ii) correlações cruzadas entre o sinal de entrada $u[n]$ e a resposta desejada:

$$\varepsilon\{u[n-k] d^*[n]\} = R_{ud}[-k] \quad (2.24)$$

$$\varepsilon\{\mathbf{u}^*[n-k] \mathbf{d}[n]\} = R_{ud}^*[-k] \quad (2.25)$$

iii) função de autocorrelação do sinal de entrada $\mathbf{u}[n]$ para um atraso $i-k$:

$$\varepsilon\{\mathbf{u}[n-k] \mathbf{u}^*[n-i]\} = R_{uu}[i-k] \quad (2.26)$$

Reescrevendo J levando-se em conta estas definições:

$$J = \sigma_d^2 - \sum_{k=0}^{M-1} w_k R_{ud}[-k] - \sum_{k=0}^{M-1} w_k^* R_{ud}^*[-k] + \sum_{k=0}^{M-1} \sum_{i=0}^{M-1} w_k w_i^* R_{uu}[i-k] \quad (2.27)$$

Se o sinal de entrada do filtro $\mathbf{u}[n]$ e a resposta desejada $\mathbf{d}[n]$ são conjuntamente estacionários [APÊNDICE I] (a variância é constante e as funções de correlação para um mesmo atraso k também são constantes), a função de custo J é precisamente uma função de segunda ordem dos coeficientes do filtro.

2.4.2.4 Princípio da ortogonalidade

Defina o operador gradiente ∇ , cujo k -ésimo elemento é composto pelas derivadas parciais de primeira ordem com respeito à parte real a_k e imaginária b_k , para o k -ésimo coeficiente do filtro, $w_k = a_k + jb_k$ [18]:

$$\nabla_k(J) = \frac{\partial J}{\partial a_k} + j \frac{\partial J}{\partial b_k} \quad (2.28)$$

Para J atingir o valor mínimo deve-se ter $\nabla_k(J) = 0$ para todo k .

$$\nabla_k(J) = \frac{\partial J}{\partial a_k} + j \frac{\partial J}{\partial b_k} = 0 \quad (2.29)$$

Resolvendo:

$$\nabla_k(J) = \varepsilon\left\{ \mathbf{e}^*[n] \frac{\partial}{\partial a_k} \mathbf{e}[n] + \mathbf{e}[n] \frac{\partial}{\partial a_k} \mathbf{e}^*[n] + j \mathbf{e}^*[n] \frac{\partial}{\partial b_k} \mathbf{e}[n] + j \mathbf{e}[n] \frac{\partial}{\partial b_k} \mathbf{e}^*[n] \right\} \quad (2.30)$$

As derivadas parciais são:

$$\frac{\partial}{\partial a_k} e[n] = \frac{\partial}{\partial a_k} (d[n] - y[n]) = -\frac{\partial}{\partial a_k} \sum_{k=0}^{M-1} w_k u[n-k] = -u[n-k] \quad (2.31)$$

$$\frac{\partial}{\partial a_k} e^*[n] = \frac{\partial}{\partial a_k} (d^*[n] - y^*[n]) = -\frac{\partial}{\partial a_k} \sum_{k=0}^{M-1} w_k^* u^*[n-k] = -u^*[n-k] \quad (2.32)$$

$$\frac{\partial}{\partial b_k} e[n] = -\frac{\partial}{\partial b_k} y[n] = -\frac{\partial}{\partial b_k} \sum_{k=0}^{M-1} w_k u[n-k] = -j u[n-k] \quad (2.33)$$

$$\frac{\partial}{\partial b_k} e^*[n] = -\frac{\partial}{\partial b_k} y^*[n] = -\frac{\partial}{\partial b_k} \sum_{k=0}^{M-1} w_k^* u^*[n-k] = j u^*[n-k] \quad (2.34)$$

Substituindo e cancelando termos comuns, chega-se a:

$$\nabla_k (J) = -2 \varepsilon \{ e[n] u^*[n-k] \} \quad (2.35)$$

Na condição ótima:

$$\varepsilon \{ e[n] u^*[n-k] \} = \varepsilon \{ e^*[n] u[n-k] \} = 0 \quad (2.36)$$

para $k=0, \dots, M-1$. A última expressão diz que a condição necessária e suficiente para que a função de custo J atinja seu valor mínimo é que o valor correspondente do erro de estimação $e[n]$ seja ortogonal à cada amostra que entra na estimativa da resposta desejada no tempo $n-k$. Esta interpretação compara variáveis aleatórias à vetores num espaço onde o produto interno é definido como a esperança estatística [42]. Se o produto interno de dois vetores é zero, então eles são ortogonais. Este é o chamado princípio da ortogonalidade [42] [34] [18]. O apêndice II apresenta o desenvolvimento deste princípio.

Observando a correlação entre a saída do filtro e o erro de estimação:

$$\varepsilon \{ y[n] e^*[n] \} = \varepsilon \left\{ \sum_{k=0}^{M-1} w_k u[n-k] e^*[n] \right\} = \sum_{k=0}^{M-1} w_k \varepsilon \{ u[n-k] e^*[n] \} \quad (2.37)$$

Para operação na condição ótima tem-se que a estimativa da resposta desejada produzida na saída do filtro, $y_0[n]$, é ortogonal ao correspondente erro de estimação, $e_0[n]$ (note que o conjugado também é válido):

$$\varepsilon \{ \mathbf{y}_o[n] \mathbf{e}_o^*[n] \} = 0 \quad (2.38)$$

2.4.2.5 Equações de Wiener-Hopf

O princípio da ortogonalidade pode ser usado para encontrar os coeficientes do filtro ótimo, através da simples substituição de (2.20) em (2.36).

$$\varepsilon \left\{ \left(\mathbf{d}[n] - \sum_{i=0}^{M-1} w_i \mathbf{u}[n-i] \right) \mathbf{u}^*[n-k] \right\} = 0 \quad (2.39)$$

Colocando em evidência o somatório:

$$\varepsilon \{ \mathbf{d}[n] \mathbf{u}^*[n-k] \} = \varepsilon \left\{ \sum_{i=0}^{M-1} w_i \mathbf{u}[n-i] \mathbf{u}^*[n-k] \right\} = \sum_{i=0}^{M-1} w_i \varepsilon \{ \mathbf{u}[n-i] \mathbf{u}^*[n-k] \} \quad (2.40)$$

Utilizando a definição das funções de correlação:

$$\sum_{i=0}^{M-1} w_i R_u[k-i] = R_{du}[k] \quad (2.41)$$

Tomando o conjugado complexo

$$\sum_{i=0}^{M-1} w_i^* R_u^*[k-i] = R_{du}^*[k] \quad (2.42)$$

Usando as propriedades das funções de correlação, chega-se a

$$\sum_{i=0}^{M-1} w_i^* R_u[i-k] = R_{ud}[-k] \quad (2.43)$$

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

Este sistema de equações define os coeficientes do filtro ótimo em termos da função de autocorrelação da entrada do filtro e da função de correlação cruzada entre a entrada do filtro e a resposta desejada, tendo o nome de equações de Wiener-Hopf (note que este desenvolvimento foi feito para o caso específico de um filtro transversal, linear e discreto no tempo, onde estas equações se reduzem a um sistema de M equações simultâneas). Como o sistema possui M equações e M coeficientes, sua solução é única, o que assegura um único ponto ótimo para a função de custo J (mínimo global).

2.4.2.6 Superfície de performance do erro

Conseqüentemente, é possível visualizar a dependência da função de custo J em relação aos coeficientes do filtro como uma superfície $(M+1)$ dimensional (hiperparabolóide), côncava em forma de um vale. Ela deve ter a concavidade voltada para cima e seu valor mínimo deve ser maior ou igual a zero devido ao fato de ser função do quadrado do erro. Essa superfície é caracterizada por um único mínimo global, onde o gradiente de J vale zero. A figura 2.8 mostra um gráfico da função de custo J para um filtro com dois coeficientes reais w_0 e w_1 .

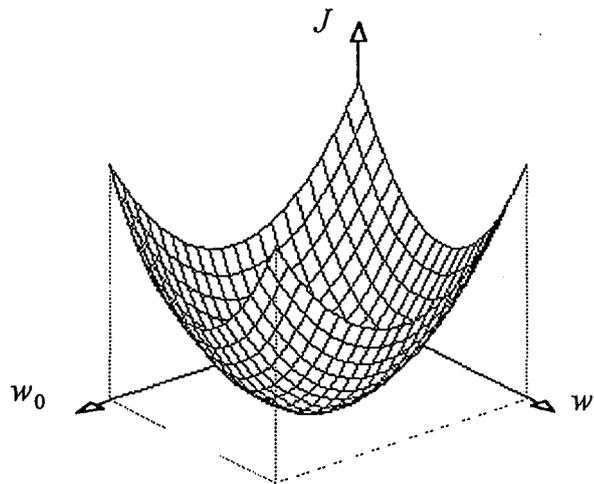


FIGURA 2.8: Gráfico da função de custo J para os coeficientes reais w_0 e w_1

2.4.2.7 Valor mínimo da função de custo J

Desenvolvendo J no ponto ótimo e denotando o valor mínimo por J_{\min} , tem-se:

$$J_{\min} = \varepsilon\{|\mathbf{e}_o[n]|^2\} = \varepsilon\{|\mathbf{d}[n] - \mathbf{y}_o[n]|^2\} \quad (2.44)$$

pois $\mathbf{e}_o[n] = \mathbf{d}[n] - \mathbf{y}_o[n]$. Continuando

$$J_{\min} = \varepsilon\{(\mathbf{d}[n] - \mathbf{y}_o[n])(\mathbf{d}^*[n] - \mathbf{y}_o^*[n])\} \quad (2.45)$$

$$J_{\min} = \varepsilon\{|\mathbf{d}[n]|^2\} + \varepsilon\{|\mathbf{y}_o[n]|^2\} - \varepsilon\{\mathbf{y}_o[n] \mathbf{d}^*[n]\} - \varepsilon\{\mathbf{y}_o^*[n] \mathbf{d}[n]\} \quad (2.46)$$

Notando que $\mathbf{d}[n] = \mathbf{e}_o[n] + \mathbf{y}_o[n]$, temos

$$J_{\min} = \varepsilon\{|d[n]|^2\} + \varepsilon\{|y_o[n]|^2\} - \varepsilon\{y_o[n] e_o^*[n] + y_o[n] y_o^*[n]\} - \varepsilon\{y_o^*[n] e_o[n] + y_o^*[n] y_o[n]\} \quad (2.47)$$

usando (2.38) e notando que o conjugado complexo também é válido:

$$J_{\min} = \varepsilon\{|d[n]|^2\} + \varepsilon\{|y_o[n]|^2\} \quad (2.48)$$

$$J_{\min} = \sigma_d^2 - \sigma_{y_o}^2 \quad (2.49)$$

O valor mínimo de J é igual à diferença da variância do sinal desejado pela variância do sinal de saída do filtro ótimo.

2.4.2.8 Formulação matricial das equações de Wiener-Hopf

Defina o vetor $M \times 1$ de sinal do filtro transversal $\mathbf{u}[n], \mathbf{u}[n-1], \dots, \mathbf{u}[n-M+1]$:

$$\mathbf{u}_n = [\mathbf{u}[n] \quad \mathbf{u}[n-1] \quad \dots \quad \mathbf{u}[n-M+1]]^T \quad (2.50)$$

Defina também o vetor $M \times 1$ de coeficientes do filtro transversal, w_0, w_1, \dots, w_{M-1} :

$$\mathbf{W} = [w_0 \quad w_1 \quad \dots \quad w_{M-1}]^T \quad (2.51)$$

A resposta de um filtro FIR a um sinal pode ser representada, no caso real, pelo produto interno do vetor de coeficientes do filtro, \mathbf{W} , e do vetor de sinal, \mathbf{u} . No entanto, a adoção desta definição no caso complexo não é adequada, pois teremos uma resposta ao impulso e coeficientes do filtro que serão o conjugado complexo do vetor de coeficientes \mathbf{W} . Este fato se põe em direta oposição à resposta que seria obtida pela definição de $y[n]$ em termos da soma de convolução representada por 2.16 [ANEXO E]. Assim, para o caso de vetores complexos, adota-se como representação da saída do filtro FIR o produto interno do vetor de coeficientes conjugado, \mathbf{W}^* , e do vetor de sinal, \mathbf{u}_n . Neste caso, a expressão do sinal de saída do filtro fica

$$y[n] = \mathbf{W}^* \cdot \mathbf{u}_n|_C = (\mathbf{W}^*)^H \cdot \mathbf{u}_n|_R = \mathbf{W}^T \mathbf{u}_n \quad (2.52)$$

Defina \mathbf{R} como a matriz $M \times M$ de correlação do sinal de entrada do filtro transversal:

$$\mathbf{R} = \mathcal{E}\{\mathbf{u}_n \mathbf{u}_n^H\} \quad (2.53)$$

onde H (hermitiano) representa as operações de conjugado complexo e transposição.

Em forma expandida:

$$\mathbf{R} = \begin{bmatrix} R_{uu}[0] & R_{uu}[1] & \cdots & R_{uu}[M-1] \\ R_{uu}^*[1] & R_{uu}[0] & \cdots & R_{uu}[M-2] \\ \vdots & \vdots & \ddots & \vdots \\ R_{uu}^*[M-1] & R_{uu}^*[M-2] & \cdots & R_{uu}[0] \end{bmatrix} \quad (2.54)$$

Defina \mathbf{p} como o vetor de correlação cruzada entre o sinal de entrada do filtro e a resposta desejada $\mathbf{d}[n]$:

$$\mathbf{p} = \mathcal{E}\{\mathbf{u}_n \mathbf{d}^*[n]\} \quad (2.55)$$

Em forma expandida:

$$\mathbf{p} = [R_{ud}[0] \quad R_{ud}[-1] \quad \cdots \quad R_{ud}[1-M]]^T \quad (2.56)$$

Reescrevendo as equações de Wiener-Hopf em forma matricial, temos:

$$\mathbf{R} \mathbf{W}_o^* = \mathbf{p} \quad (2.57)$$

onde o vetor de coeficientes ótimos conjugados do filtro, \mathbf{W}_o^* , é dado por

$$\mathbf{W}_o^* = [w_{o,0}^* \quad w_{o,1}^* \quad \cdots \quad w_{o,M-1}^*]^T \quad (2.58)$$

Assumindo que a matriz de autocorrelação é não-singular, encontra-se a equação que define os coeficientes ótimos do filtro transversal FIR para o caso em que se conhece a inversa da matriz de autocorrelação do sinal de entrada do filtro e o vetor de correlação cruzada entre este sinal ($\mathbf{u}[n]$) e o sinal desejado ($\mathbf{d}[n]$):

$$\mathbf{W}_o^* = \frac{\mathbf{p}}{\mathbf{R}} = \mathbf{R}^{-1} \mathbf{p} \quad (2.59)$$

Segundo Haykin [18], a matriz de autocorrelação $\mathbf{R}_{M \times M}$ de um processo estocástico discreto no tempo estacionário, com autovalores $\lambda_1, \dots, \lambda_M$ e autovetores $\mathbf{q}_1, \dots, \mathbf{q}_M$, possui as seguintes propriedades:

- i) A matriz de autocorrelação é hermitiana (igual ao seu transposto conjugado). Para sinais não-complexos, ela também é simétrica.
- ii) A matriz de autocorrelação é Toeplitz (todos os elementos de sua diagonal principal são iguais e todos os elementos das demais diagonais, paralelas à principal, também são iguais).
- iii) A matriz de autocorrelação é não-negativa definida e quase sempre positiva definida.
- iv) Os autovalores da matriz \mathbf{R}^k igualam $(\lambda_1^k, \dots, \lambda_M^k)$ para qualquer inteiro $k > 0$.
- v) Se os autovalores da matriz de correlação são distintos entre si, os autovetores associados são linearmente independentes.
- vi) Se os autovalores da matriz \mathbf{R} são distintos entre si, os autovetores associados são ortogonais entre si.
- vi) Os autovalores da matriz de correlação \mathbf{R} são reais e não-negativos.
- viii) A matriz de correlação \mathbf{R} é mal-condicionada se a razão do maior autovalor para o menor autovalor for grande.
- ix) Os autovalores da matriz de correlação \mathbf{R} de um processo discreto no tempo são limitados pelos valores mínimo e máximo da densidade espectral de potência do processo.

2.4.2.9 Forma matricial do erro mínimo

O valor mínimo da função de custo J é a diferença da variância do sinal desejado pela variância do sinal de saída do filtro na situação ótima (equação 2.49). A variância do sinal de saída do filtro na situação ótima pode ser descrito em forma matricial:

$$\sigma_{y_o}^2 = \varepsilon\{|y_o[n]|^2\} \quad (2.60)$$

$$\sigma_{y_o}^2 = \varepsilon\{y_o[n] y_o^*[n]\} \quad (2.61)$$

Usado 2.52 e considerando o filtro ótimo:

$$\sigma_{y_o}^2 = \varepsilon\{\mathbf{W}_o^T \mathbf{u}[n] \mathbf{u}^H[n] \mathbf{W}_o^*\} \quad (2.62)$$

Como os coeficientes do filtro ótimo são constantes:

$$\sigma_{y_o}^2 = \mathbf{W}_o^T \varepsilon \{ \mathbf{u}[n] \mathbf{u}^H[n] \} \mathbf{W}_o^* \quad (2.63)$$

Usando 2.53:

$$\sigma_{y_o}^2 = \mathbf{W}_o^T \mathbf{R} \mathbf{W}_o^* \quad (2.64)$$

Usando as equações de Wiener-Hopf em forma matricial (2.57), desenvolve-se :

$$\mathbf{W}_o^T \mathbf{R} \mathbf{W}_o^* = \mathbf{W}_o^T \mathbf{p} = \mathbf{p}^H \mathbf{W}_o^* \quad (2.65)$$

e

$$\mathbf{W}_o^* = \mathbf{R}^{-1} \mathbf{p} \quad (2.66)$$

portanto, 2.49 assume a forma:

$$J_{\min} = \sigma_d^2 - \mathbf{p}^H \mathbf{R}^{-1} \mathbf{p} \quad (2.67)$$

que é o valor mínimo de J em função da variância do sinal desejado, da matriz de autocorrelação \mathbf{R} e do vetor de correlação cruzada \mathbf{p} .

2.4.2.10 Forma canônica da superfície de performance do erro

Reescrevendo a equação 2.27 em forma matricial [18]:

$$J = \sigma_d^2 - \mathbf{W}^T \mathbf{p} - \mathbf{p}^H \mathbf{W}^* + \mathbf{W}^T \mathbf{R} \mathbf{W}^* \quad (2.68)$$

Utilizando 2.67:

$$J = J_{\min} + (\mathbf{W} - \mathbf{W}_o)^T \mathbf{R} (\mathbf{W} - \mathbf{W}_o)^* \quad (2.69)$$

Decompondo \mathbf{R} em termos de seus autovalores e autovetores, reescreve-se

$$\mathbf{R} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^H \quad (2.70)$$

onde $\mathbf{\Lambda}$ é a matriz diagonal cujos elementos são os autovalores de \mathbf{R} ($\lambda_1, \lambda_2, \lambda_3, \dots$) e \mathbf{Q} é a matriz cujas colunas são os autovetores ($\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \dots$) associados aos autovalores.

Assim,

$$J = J_{\min} + (\mathbf{W} - \mathbf{W}_o)^T \mathbf{Q} \Lambda \mathbf{Q}^H (\mathbf{W} - \mathbf{W}_o)^* \quad (2.71)$$

Definindo a matriz \mathbf{v} como a matriz "transformada" da diferença entre o vetor de coeficientes, \mathbf{W} , e o vetor ótimo de coeficientes, \mathbf{W}_o :

$$\mathbf{v} = \mathbf{Q}^H (\mathbf{W} - \mathbf{W}_o)^* \quad (2.72)$$

A forma canônica é:

$$J = J_{\min} + \mathbf{v}^H \Lambda \mathbf{v} \quad (2.73)$$

e sua relação com os autovalores da matriz de correlação é:

$$J = J_{\min} + \sum_{k=1}^M \lambda_k v_k v_k^* \quad (2.74)$$

$$J = J_{\min} + \sum_{k=1}^M \lambda_k |v_k|^2 \quad (2.75)$$

onde v_k é o k -ésimo componente do vetor \mathbf{v} . Os componentes do vetor \mathbf{v} constituem os eixos principais da superfície de performance do erro.

2.4.2.11 O filtro de Wiener não-causal

Se for utilizado um filtro IIR não-causal na figura 2.6, o filtro resultante é dito filtro de Wiener não-causal. Embora não tenha aplicação prática direta, este desenvolvimento provê uma perspectiva da solução e será usado no desenvolvimento do cancelador de ruído adaptativo.

A estimativa para o sinal desejado é da forma

$$y[n] = \sum_{k=-\infty}^{+\infty} w[n-k]u[k] \quad (2.76)$$

e a aplicação do princípio da ortogonalidade (2.36) leva à equação de Wiener-Hopf:

$$\sum_{l=-\infty}^{+\infty} R_x[l-i]w^*[l] = R_{dx}^*[i] \quad (2.77)$$

para $-\infty \leq i \leq +\infty$. Esta equação deve ser válida para todos os infinitos valores de i . Se for feito o conjugado complexo desta equação e a sua transformada z , chega-se a [42]:

$$H(z)S_x(z) = S_{dx}(z) \quad (2.78)$$

Esta equação pode ser resolvida diretamente para a $H(z)$ ótima.

2.4.3 Método dos mínimos quadrados

2.4.3.1 Introdução

O método dos mínimos quadrados pode ser usado para resolver o problema de filtragem linear ótima sem o uso de ferramentas estatísticas. A idéia básica do método é, supondo que se tenha um conjunto de N medidas u_i feitas nos instantes de tempo t_i , construir uma curva que aproxime estes pontos num senso ótimo. O senso ótimo é atingido minimizando a soma dos quadrados da diferença entre os valores estimados e os valores medidos para todos os N instantes de tempo. Este método pode ser visto como uma alternativa aos filtros de Wiener, com a diferença básica de que são usadas médias temporais ao invés de médias probabilísticas. Estas médias dependem então do número de amostras usadas em sua estimativa.

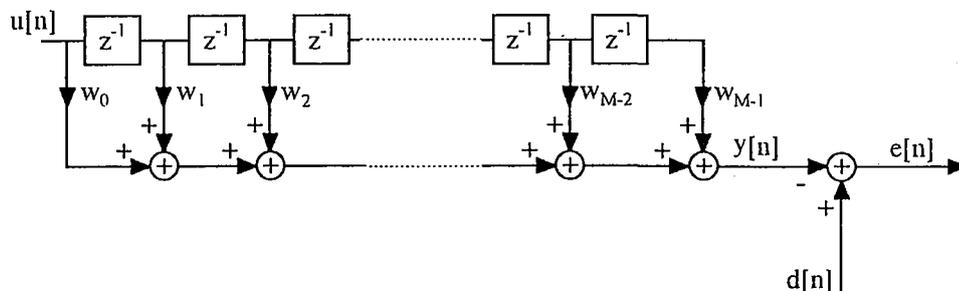


FIGURA 2.9: Modelo para a resolução do problema de filtragem ótima linear por mínimos quadrados

Considere a figura 2.9. Usa-se o mesmo filtro linear causal discreto no tempo, de estrutura transversal, considerado anteriormente para a resolução do problema de filtragem ótima. No entanto, os sinais agora não são assumidos como fazendo parte de um processo estocástico. Assume-se que o sinal $d[n]$ foi gerado por um modelo de regressão linear múltipla (figura 2.10) [18] [30], com um erro não-observável. Este erro é assumido como ruído branco de variância σ^2 e média zero (um processo estocástico discreto no tempo). O filtro da figura 2.9 deve estimar os parâmetros $w_{k,m}$ do modelo, baseado no conjunto de variáveis observáveis $\{u[i]\}$ e $\{d[i]\}$, para $i=1, \dots, N$. O erro de estimação ou resíduo é dado por:

$$e[i] = d[i] - y[i] \quad (2.79)$$

onde

$$y[i] = \sum_{k=0}^{M-1} w_k u[i-k] \quad (2.80)$$

e a função de custo a ser minimizada é a soma dos quadrados dos erros:

$$J_{ls} = \sum_{i=i_1}^{i_2} |e[i]|^2 \quad (2.81)$$

Nesta equação, i_1 e i_2 definem os limites onde a minimização do erro ocorre. Os valores destes limites dependem do tipo de janelamento de dados (*data windowing*) usado. O filtro resultante da minimização é chamado de filtro de mínimos-quadrados linear (*linear least-squares filter*). Há dois modos de descrever a condição ótima do filtro de mínimos quadrados: o princípio da ortogonalidade e o sistema de equações normais [18].

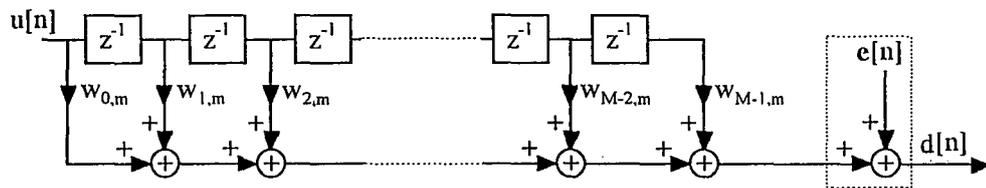


FIGURA 2.10: Modelo de regressão linear múltipla para formação da seqüência $d[n]$

2.4.3.2 Janelamento de dados

Pode ser construída uma matriz com os N dados de entrada disponíveis do filtro da figura 2.9, $u[1], \dots, u[N]$. Esta matriz vai depender de como são feitas as escolhas para os limites i_1 e i_2 . Definindo M como o número de coeficientes (*taps*) do filtro ($M \leq N$), podem ser definidos quatro métodos diferentes de janelamento de dados de entrada [18]:

- 1) Método da covariância: não faz considerações sobre dados fora do intervalo $[1, N]$. Pode-se formar a matriz de dados com $i_1=M$ e $i_2=N$:

$$\mathbf{A} = \begin{bmatrix} u[M] & u[M+1] & \cdots & u[N] \\ u[M-1] & u[M] & \cdots & u[N-1] \\ \vdots & \vdots & \ddots & \vdots \\ u[1] & u[2] & \cdots & u[N-M+1] \end{bmatrix} \quad (2.82)$$

- 2) Método da autocorrelação: assume que os dados antes de $i=1$ e depois de $i=N$ são zero. Definindo os limites $i_1=1$ e $i_2=N+M-1$:

$$\mathbf{A} = \begin{bmatrix} u[1] & u[2] & \cdots & u[M] & u[M+1] & \cdots & u[N] & 0 & \cdots & 0 \\ 0 & u[1] & \cdots & u[M-1] & u[M] & \cdots & u[N-1] & U[N] & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u[1] & u[2] & \cdots & U[N-M+1] & U[N-M] & \cdots & u[N] \end{bmatrix} \quad (2.83)$$

- 3) Método de pré-janelamento: assume que os dados antes de $i=1$ são zero, mas não faz consideração sobre dados depois de $i=N$. Neste caso $i_1=1$ e $i_2=N$.

$$\mathbf{A} = \begin{bmatrix} u[1] & u[2] & \cdots & u[M] & u[M+1] & \cdots & u[N] \\ 0 & u[1] & \cdots & u[M-1] & u[M] & \cdots & u[N-1] \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u[1] & u[2] & \cdots & U[N-M+1] \end{bmatrix} \quad (2.84)$$

- 4) Método de pós-janelamento: não faz considerações sobre dados antes de $i=1$, mas assume que os dados depois de $i=N$ são zero. Neste caso $i_1=M$ e $i_2=N+M+1$.

$$\mathbf{A} = \begin{bmatrix} u[M] & u[M+1] & \cdots & u[N] & 0 & \cdots & 0 \\ u[M-1] & u[M-1] & \cdots & u[N-1] & u[N] & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ u[1] & u[2] & \cdots & u[N-M+1] & u[N-M] & \cdots & u[N] \end{bmatrix} \quad (2.85)$$

O emprego destes nomes é mais tradicional do que propriamente matemático [18].

2.4.3.3 Princípio da ortogonalidade

A função de custo no método da covariância é dada por

$$J_{ls} = \sum_{i=M}^N |e[i]|^2 \quad (2.86)$$

Estes limites asseguram que, para cada valor de i , todos os sinais de entrada dos coeficientes do filtro transversal (*tap inputs*) têm valor diferente de zero. Para determinar os coeficientes w que minimizam a função de custo, calcula-se o gradiente de J_{ls} para a parte real e imaginária de w_k :

$$\nabla_k (J_{ls}) = \frac{\partial}{\partial a_k} J_{ls} + j \frac{\partial}{\partial b_k} J_{ls} \quad (2.87)$$

O desenvolvimento usando (2.86) leva a [18]:

$$\nabla_k (J_{ls}) = -2 \sum_{i=M}^N u^*[i-k]e[i] \quad (2.88)$$

Para que a função de custo atinja o seu valor mínimo, cada componente do gradiente deve ser igual a zero. Denotando o valor do erro na situação ótima em que o filtro opera por e_{\min} , erro na condição de mínimos quadrados, tem-se:

$$\sum_{i=M}^N u^*[i-k]e_{\min}[i] = 0 \quad (2.89)$$

para $k=0, \dots, M-1$. Esta é a versão temporal do princípio da ortogonalidade. Denotando o valor da saída do filtro na condição de mínimos quadrados por y_{\min} , a relação abaixo pode ser desenvolvida, usando (2.80) para coeficientes ótimos e (2.89), obtendo-se:

$$\sum_{i=M}^N y_{\min}^*[i]e_{\min}[i] = 0 \quad (2.90)$$

Ou seja, quando o filtro opera na condição de mínimos quadrados, as séries temporais de saída do filtro (y_{\min}) e do erro de estimação (e_{\min}) são ortogonais. Para obter o valor mínimo de J_{ls} , usa-se a igualdade válida na condição ótima de funcionamento do filtro

$$d[i] = y_{\min}[i] + e_{\min}[i] \quad (2.91)$$

Calculando a energia (soma dos quadrados) das séries temporais no intervalo $[M, N]$ e usando o corolário ao princípio da ortogonalidade (2.90), pode-se obter o valor mínimo da função de custo [18]:

$$J_{ls, \min} = \sum_{i=M}^N |e_{\min}[i]|^2 = \sum_{i=M}^N |d[i]|^2 - \sum_{i=M}^N |y_{\min}[i]|^2 \quad (2.92)$$

2.4.3.4 Equações normais

Usando (2.80) e substituindo em (2.79), tem-se

$$e_{\min}[i] = d[i] - \sum_{t=0}^{M-1} \hat{w}_t u[i-t] \quad (2.93)$$

onde \hat{w}_t são as estimativas ótimas para os coeficientes do modelo de regressão linear. Substituindo esta equação no princípio da ortogonalidade (2.89), chega-se a um sistema de M equações simultâneas

$$\sum_{t=0}^{M-1} \hat{w}_t \sum_{i=M}^N u^*[i-k]u[i-t] = \sum_{i=M}^N u^*[i-k]d[i] \quad (2.94)$$

onde $k = 0, 1, \dots, M-1$. Os somatórios sobre i podem ser vistos como estimativas para as médias temporais no intervalo de interesse. Definindo as funções temporais de autocorrelação e correlação cruzada [18]:

$$\phi(t, k) = \sum_{i=M}^N u[i-k]u^*[i-t] \quad (2.95)$$

para $0 \leq t, k \leq M-1$, e

$$\theta(-k) = \sum_{i=M}^N u[i-k]d^*[i] \quad (2.96)$$

para $0 \leq k \leq M-1$. Pode-se, então, reescrever o conjugado complexo da equação (2.94) como:

$$\sum_{t=0}^{M-1} \hat{w}_t^* \phi[t, k] = \theta[-k] \quad (2.97)$$

para $k=0, 1, \dots, M-1$. Este é o sistema expandido de equações normais de um filtro de mínimos quadrados linear, que representa a versão temporal das equações de Wiener-Hopf (2.43).

Para a formulação matricial das equações normais são necessárias as seguintes definições:

- i) a matriz de autocorrelação temporal $M \times M$ do sinal de entrada do filtro $u[i]$, $u[i-1]$, ..., $u[i-M+1]$:

$$\Phi = \begin{bmatrix} \phi[0,0] & \phi[1,0] & \dots & \phi[M-1,0] \\ \phi[0,1] & \phi[1,1] & \dots & \phi[M-1,1] \\ \vdots & \vdots & \ddots & \vdots \\ \phi[0,M-1] & \phi[1,M-1] & \dots & \phi[M-1,M-1] \end{bmatrix} \quad (2.98)$$

ii) o vetor de correlação cruzada temporal $M \times 1$ entre $u[i]$ e $d[i]$

$$\theta = [\theta[0] \quad \theta[-1] \quad \dots \quad \theta[-M+1]]^T \quad (2.99)$$

iii) o vetor de coeficientes $M \times 1$ do filtro de mínimos quadrados

$$\hat{\mathbf{W}}^* = [\hat{w}_0^* \quad \hat{w}_1^* \quad \dots \quad \hat{w}_{M-1}^*]^T \quad (2.100)$$

O sistema de equações normais pode ser escrito, então, como:

$$\Phi \hat{\mathbf{W}}^* = \theta \quad (2.101)$$

Se for assumido que Φ é não-singular e que, portanto, sua inversa existe, a última equação pode ser resolvida para os coeficientes do filtro linear de mínimos quadrados:

$$\hat{\mathbf{W}}^* = \Phi^{-1} \theta \quad (2.102)$$

A matriz de autocorrelação Φ (2.98) pode ser escrita de maneira equivalente como a média temporal do produto vetorial do vetor $\mathbf{u}[i]$ por seu transposto conjugado [ANEXO E]

$$\Phi = \sum_{i=M}^N \mathbf{u}[i] \mathbf{u}^H[i] \quad (2.103)$$

onde

$$\mathbf{u}[i] = [u[i] \quad u[i-1] \quad \dots \quad u[i-M+1]]^T \quad (2.104)$$

Outro modo de defini-la é pelo produto de duas matrizes, onde uma é o hermitiano da outra:

$$\Phi = \mathbf{A}^H \mathbf{A} \quad (2.105)$$

onde \mathbf{A} é a matriz de dados de entrada pelo método de covariância de janelamento (2.82). A matriz \mathbf{A} é uma matriz de ordem $(N-M+1) \times M$ com estrutura Toeplitz.

Haykin [18] lista as seguintes propriedades de Φ :

1) A matriz de correlação é hermitiana (igual à sua transposta conjugada)

$$\Phi = \Phi^H \quad (2.106)$$

2) A matriz de correlação é definida não-negativa, ou seja, para qualquer vetor $\mathbf{x}_{M \times 1}$

$$\mathbf{x}^T \Phi \mathbf{x} \geq 0 \quad (2.107)$$

3) Os autovalores da matriz de autocorrelação Φ são todos reais e não-negativos;

4) A matriz Φ é não-Toeplitz.

2.5 O cancelador adaptativo de ruído

2.5.1 Conceito

Deseja-se demonstrar o funcionamento do cancelador adaptativo de ruído (ANC - *Adaptive Noise Canceller*) através da análise da potência dos sinais envolvidos. O efeito da minimização da potência da saída do cancelador sobre o sinal de saída do filtro é tratado de modo a prover uma base intuitiva. A análise do problema usando a teoria de filtragem ótima de Wiener é feita para situações onde existam ruídos não-correlacionados presentes nas entradas e para o caso de *leakage*. São derivadas equações que permitem avaliar a performance do ANC nestas condições.

Considere a figura 2.11, onde s é o sinal desejado, i e i_2 são interferências, y é a saída do filtro e e é o erro ou saída do sistema. Assume-se que:

- 1) s , i , i_2 e y são estatisticamente estacionários (WSS) e com média zero;
- 2) s não é correlacionado com i ou i_2 (e, conseqüentemente com y);
- 3) i está correlacionado com i_2 .

Neste desenvolvimento assume-se que o filtro adaptativo minimiza a potência do sinal de saída. Essa minimização e seus mecanismos serão tratados no capítulo 3. A função de transferência do filtro adaptativo é denotada por $H_{fa}(z)$. O sinal de saída do cancelador é:

$$\mathbf{e} = \mathbf{s} + \mathbf{i} - \mathbf{y} \quad (2.108)$$

Elevando ao quadrado e tomando a média estatística de ambos os lados, e considerando a hipótese 2, chega-se a

$$\varepsilon \{ \mathbf{e}^2 \} = \varepsilon \{ \mathbf{s}^2 \} + \varepsilon \{ (\mathbf{i} - \mathbf{y})^2 \} \quad (2.109)$$

Que exprime a potência do sinal de saída do sistema. Quando o filtro adaptativo ajusta sua resposta para minimizar a potência do erro, devido à correlação dos sinais y e i , sua

atuação se dá sobre o segundo termo da direita da equação (2.109) e o primeiro termo da direita permanece inalterado. Isto equivale a dizer que, quando o filtro é ajustado para minimizar a potência do sinal de erro, sua saída é a melhor estimativa no sentido da mínima média do quadrado (*least mean squares*) da interferência i presente na entrada primária. Isto faz com que o sinal de saída do sistema seja a melhor estimativa do sinal desejado presente na entrada primária, s . No caso extremo, quando for possível tornar $y=i$, o sistema torna a saída completamente livre de ruído.

Num outro extremo, quando i e i_2 são completamente não-correlacionados, a saída do filtro y também fica não-correlacionada com a entrada primária e pode-se escrever a potência do sinal de saída como:

$$\varepsilon\{e^2\} = \varepsilon\{(s+i)^2\} + \varepsilon\{y^2\} \quad (2.110)$$

Neste caso a minimização da potência de saída implica em minimizar a potência do sinal y , o que é feito levando os coeficientes do filtro para zero (filtro FIR estrutura direta).

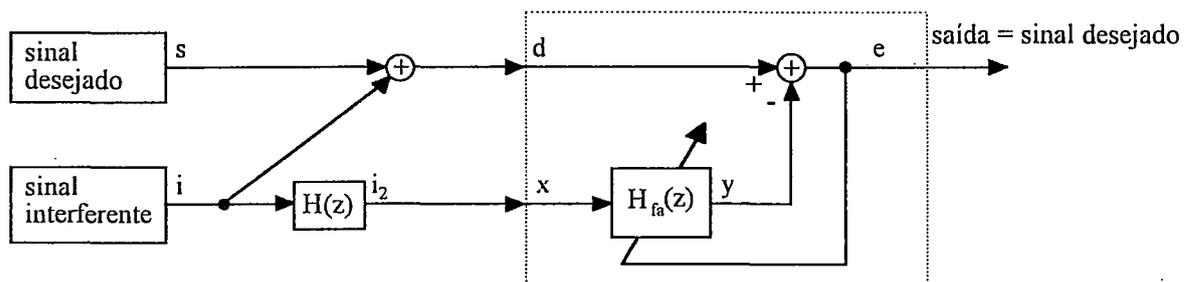


FIGURA 2.11: O cancelador adaptativo de ruído

2.5.2 Soluções estacionárias utilizando a teoria de filtros de Wiener

A função de transferência do filtro adaptativo da figura é assumida estável, ou seja, o filtro já adaptou os seus coeficientes e, como os sinais são estacionários, estes coeficientes permanecem inalterados de agora em diante. Nessa situação a função de transferência do filtro adaptativo pode ser substituída pela função de transferência do filtro de Wiener, sendo a semelhança com o problema de filtragem ótima evidenciado se for comparada a figura 2.6. Essa substituição apenas indica a resposta que o filtro deve atingir após o processo de adaptação.

Na figura 2.12 mostra-se o bloco principal do cancelador adaptativo de ruído para efeito de comparação com o problema de filtragem linear ótima. A entrada primária é o sinal d , a entrada de referência é o sinal x , o erro é e e y é a saída do filtro adaptativo. As hipóteses sobre os sinais anteriores são válidas ainda, e há a adição de ruído não-correlacionado entre si

e com os demais sinais em ambas as entradas, n_1 e n_2 . Deseja-se analisar a performance do sistema de cancelamento adaptativo de ruído através do uso da teoria de filtros de Wiener para esta situação.

$$|H(z)|^2 = H(z)H^*(1/z^*) \quad (2.111)$$

A função de transferência ótima no sentido de mínima média do quadrado é representada por $W_o(z)$. A transformada z é usada considerando que sua avaliação é feita sobre o círculo de raio unitário (onde (2.111) é válida) [33]. Esta é a solução que o filtro adaptativo deve atingir se os sinais presentes forem estacionários. Considera-se a solução para o caso de um filtro de resposta ao impulso de duração infinita não-causal, equação (2.78). Neste caso:

$$W_o(z) = \frac{S_{dx}(z)}{S_x(z)} \quad (2.112)$$

Os resultados da seção que trata dos efeitos de sistemas LTI sobre processos estocásticos (sinais) serão utilizados largamente nesta seção, especialmente a tabela 2.4.

O espectro de densidade de potência do sinal de entrada do filtro (d) é o espectro do ruído n_2 e da interferência i_2 , que é i chegando através de $H(z)$. Pode-se escrever, então:

$$S_x(z) = S_{n_2}(z) + S_i(z)|H(z)|^2 \quad (2.113)$$

O espectro de densidade de potência cruzado entre os sinais x e d depende somente dos componentes correlacionados presentes nestas duas entradas (provenientes de i)

$$S_{dx}(z) = S_{i_2}(z) \quad (2.114)$$

desenvolvendo usando a definição de S e as hipóteses assumidas sobre os sinais, chega-se a [46]

$$S_{dx}(z) = H^*(z)S_{i_1}(z) \quad (2.115)$$

Substituindo em (2.112) os termos equivalentes, tem-se a solução em termos de espectro de potência:

$$W_o(z) = \frac{S_i(z)H^*(z)}{S_{n_2}(z) + S_i(z)|H(z)|^2} \quad (2.116)$$

Note a independência da solução do espectro do sinal primário (s) e do espectro do ruído primário não-correlacionado (n_1). Para o caso onde não há ruído não-correlacionado na referência, tem-se a solução óbvia:

$$W_o(z) = \frac{1}{H(z)} \quad (2.117)$$

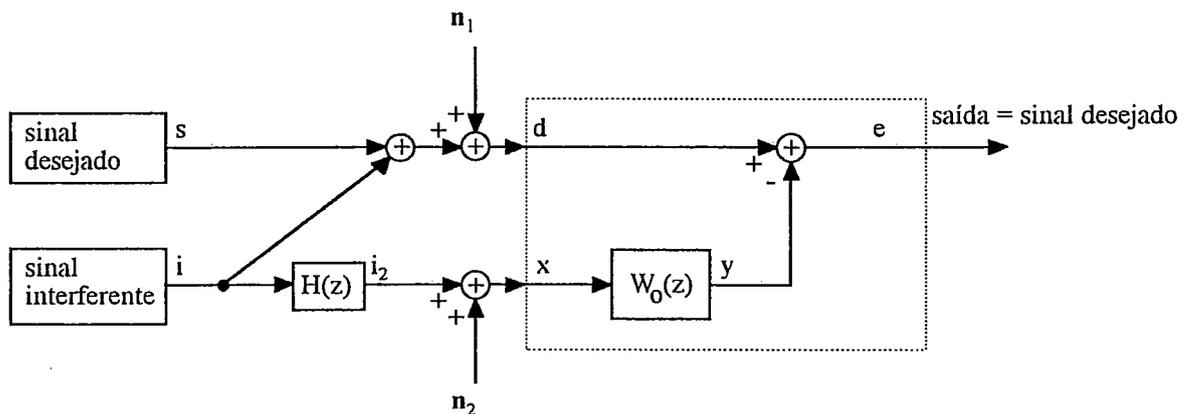


FIGURA 2.12: O cancelador adaptativo de ruído com ruídos não-correlacionados nas entradas

2.5.2.1 Análise usando relação sinal-ruído

Widrow [46] avalia a performance do ANC utilizando-se da relação sinal-ruído, definida aqui como a razão do espectro de potência do sinal pelo espectro de potência do ruído (sendo função da frequência) e denotada por $\rho(z)$. Assume-se que o espectro dos sinais é maior que zero para todas as frequências.

$$\rho(z) = \frac{S_{\text{sinal}}(z)}{S_{\text{ruído}}(z)} \quad (2.118)$$

A razão da relação sinal-ruído na saída do cancelador para a entrada primária permite avaliar a performance do cancelador de um ponto de vista mais prático. Desenvolvendo esta razão simplificando os termos do espectro do sinal primário, chega-se a [46]:

$$\frac{\rho_{\text{out}}(z)}{\rho_{\text{pri}}(z)} = \frac{S_{\text{ruído na entrada primária}}}{S_{\text{ruído na saída}}} = \frac{S_i(z) + S_{n_1}(z)}{S_{n_1}(z) + S_{n_2}(z)|W_o(z)|^2 + S_i(z)|1 - H(z)W_o(z)|^2} \quad (2.119)$$

O denominador é formado por 3 partes: propagação de n_1 diretamente para a saída, de n_2 via $W_o(z)$ e de i , diretamente e via $H(z)W_o(z)$, com defasagem de 180° . Se forem definidas as grandezas seguintes:

$$A(z) \triangleq \frac{S_{n_1}(z)}{S_i(z)} \quad (2.120)$$

$$B(z) \triangleq \frac{S_{n_2}(z)}{S_i(z)|H(z)|^2} \quad (2.121)$$

onde $A(z)$ é a relação ruído-ruído na entrada primária e $B(z)$ é a relação ruído-ruído na entrada de referência, pode-se reescrever (2.116) e (2.119) como

$$W_o(z) = \frac{1}{H(z)(B(z)+1)} \quad (2.122)$$

$$\frac{\rho_{out}(z)}{\rho_{pri}(z)} = \frac{(A(z)+1)(B(z)+1)}{A(z) + A(z)B(z) + B(z)} \quad (2.123)$$

Estas expressões permitem avaliar a performance do ANC com uma entrada primária e uma entrada de referência, com ruídos não correlacionados presentes em cada entrada. Uma razão das relações sinal-ruído alta (>1), indica melhora do sinal presente na entrada primária. Uma razão pequena (<1), indica introdução de ruído no sinal da entrada primária.

Pode-se relacionar os seguintes casos de interesse:

- i) pouco ruído não-correlacionado na entrada primária ($A(z)$ pequeno):

$$\left. \frac{\rho_{out}(z)}{\rho_{pri}(z)} \right|_{A(z) \rightarrow 0} = \frac{B(z)+1}{B(z)} \quad (2.124)$$

Se $B(z)$ for pequeno, isto é, se o efeito do ruído não-correlacionado for pequeno face ao ruído correlacionado, a razão é maior que 1 (um) e o cancelador melhora a relação sinal-ruído do sinal presente na entrada primária. Se $B(z)$ for grande, ou seja, quando houver muito ruído não-correlacionado, o cancelador não consegue melhorar a relação sinal-ruído, mas não a torna pior.

- ii) pouco ruído não-correlacionado na entrada de referência ($B(z)$ pequeno):

$$\left. \frac{\rho_{\text{out}}(z)}{\rho_{\text{pri}}(z)} \right|_{B(z) \rightarrow 0} = \frac{A(z)+1}{A(z)} \quad (2.125)$$

O mesmo comentário anterior é válido, com $B(z)$ substituído por $A(z)$.

iii) pouco ruído não-correlacionado em ambas entradas ($A(z)$ e $B(z)$ pequenos):

$$\left. \frac{\rho_{\text{out}}(z)}{\rho_{\text{pri}}(z)} \right|_{\substack{A(z) \rightarrow 0 \\ B(z) \rightarrow 0}} = \frac{1}{A(z)+B(z)} \quad (2.126)$$

Nota-se que, para pouco ruído não-correlacionado em ambas as entradas, a performance do cancelador não fica comprometida. Se $A(z)=B(z)=0$, implica em perfeito cancelamento da interferência. Na verdade, fatores práticos limitam essa performance.

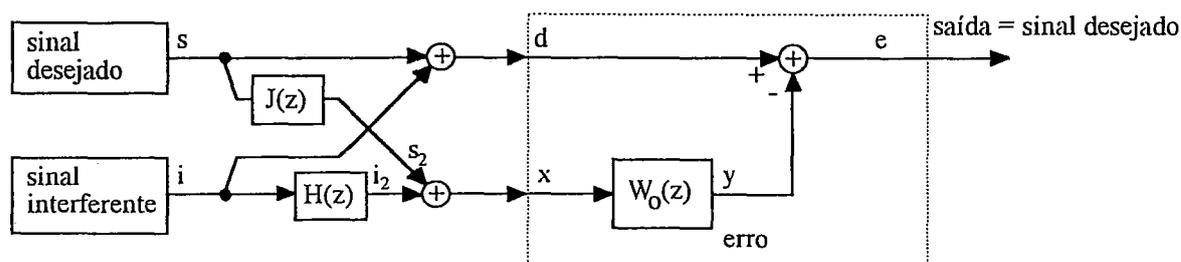


FIGURA 2.13 : O problema de *leakage* do sinal desejado

2.5.3 Efeitos de componentes do sinal na referência

Deseja-se saber qual o nível de sinal de interesse que pode estar presente na entrada de referência e, ainda assim, não tornar a aplicação do cancelador adaptativo de ruído inaceitável. A figura 2.13 ilustra esse problema, onde foram tomadas as transformadas z das funções de transferência e onde J representa a diferença do sinal desejado presente na entrada de referência e do presente na entrada primária. Não são considerados ruídos não-correlacionados por motivo de simplificação. Resolvendo (2.112) para este caso, nota-se que o espectro de potência da entrada de referência depende do sinal desejado chegando por $J(z)$ e do ruído chegando por $H(z)$, ou seja,

$$S_x(z) = S_s(z)|J(z)|^2 + S_i(z)|H(z)|^2 \quad (2.127)$$

O espectro de potência cruzado entre as entradas é desenvolvido usando a definição de espectro de potência cruzado como a transformada de Fourier da correlação cruzada dos sinais destas entradas:

$$S_{dx}(z) = S_{ii_2}(z) + S_{ss_2}(z) \quad (2.128)$$

Usando os resultados derivados na seção sobre efeitos de sistema LTI sobre um sinal estacionário (seção 2.2.3):

$$S_{dx}(z) = S_s(z)J^*(z) + S_i(z)H^*(z) \quad (2.129)$$

A função de transferência ótima será então:

$$W_o(z) = \frac{S_s(z)J^*(z) + S_i(z)H^*(z)}{S_s(z)|J(z)|^2 + S_i(z)|H(z)|^2} \quad (2.130)$$

2.5.3.1 Relação sinal-ruído na saída do cancelador ($\rho_{out}(z)$)

A razão das relações sinal-ruído na saída e na entrada de referência do cancelador possui uma semelhança importante, desenvolvida a seguir.

$$\rho_{out}(z) = \frac{S_{s_{out}}(z)}{S_{i_{out}}(z)} \quad (2.131)$$

Esta equação representa o espectro de potência do sinal na saída do cancelador, que resulta do sinal que chega diretamente e do que chega pelas funções de transferência J e W, defasado de 180°. O espectro de potência do ruído, por sua vez, depende do ruído que chega diretamente e do que chega através das funções de transferência H e W, com defasagem de 180°. Desenvolvendo usando (2.130) e cancelando termos comuns, chega-se a [46]:

$$\rho_{out}(z) = \frac{S_i(z)|H(z)|^2}{S_s(z)|J(z)|^2} \quad (2.132)$$

A relação da densidade de sinal para ruído na entrada de referência é definido de modo similar:

$$\rho_{ref}(z) = \frac{S_{s_{ref}}(z)}{S_{i_{ref}}(z)} = \frac{S_{s_2}(z)}{S_{i_2}(z)} = \frac{S_s(z)|J(z)|^2}{S_i(z)|H(z)|^2} \quad (2.133)$$

A razão entre as relações sinal-ruído na entrada de referência e na saída do cancelador pode ser expressa pela equação (2.134). Esse resultado afirma que a relação sinal-ruído na saída do ANC é o recíproco da relação sinal-ruído na entrada de referência. Segundo Widrow [46], este processo é chamado de inversão de potência.

$$\rho_{\text{out}}(z) = \frac{1}{\rho_{\text{ref}}(z)} \quad (2.134)$$

2.5.3.2 Distorção do sinal na saída do cancelador

Widrow [46] define a distorção como a razão adimensional entre o espectro do sinal desejado que se propaga para a saída pelo filtro adaptativo pelo espectro do sinal desejado que se propaga pelo caminho direto, conforme a equação (2.135). Note-se que este é um fenômeno linear relacionado à alteração da forma de onda do sinal da saída, diferente da distorção harmônica não-linear.

$$D(z) \triangleq \frac{S_s(z) |J(z)W_o(z)|^2}{S_s(z)} = |J(z)W_o(z)|^2 \quad (2.135)$$

A substituição de $W_o(z)$ considerando que $|J(z)|$ é pequeno em relação a $|H(z)|$ (simplificação aceitável na prática) leva a

$$D(z) \cong \left| \frac{J(z)}{H(z)} \right|^2 \quad (2.136)$$

A relação sinal-ruído na entrada primária é

$$\rho_{\text{pri}}(z) \triangleq \frac{S_s(z)}{S_i(z)} \quad (2.137)$$

com esta definição, a razão (2.136) pode ser expressa então por (2.138), indicando que uma baixa distorção do sinal é o resultado de uma relação sinal-ruído alta na entrada primária e baixa na entrada de referência, o que é intuitivamente razoável.

$$D(z) \cong \frac{\rho_{\text{ref}}(z)}{\rho_{\text{pri}}(z)} \quad (2.138)$$

2.5.3.3 Espectro do ruído na saída do cancelador

O espectro do ruído presente na saída do cancelador pode ser calculado considerando seus caminhos de propagação pelo cancelador, ou seja, pelo caminho direto e através das funções de transferência $H(z)$ e $W_o(z)$, com defasagem de 180° . Considerando $|J(z)|$ pequeno face a $|H(z)|$, o espectro do ruído na saída é dado por

$$S_{i_{out}}(z) \cong S_i(z) \left| \frac{S_s(z)J^*(z)}{S_i(z)H^*(z)} \right|^2 \quad (2.139)$$

Essa relação pode ser expressa em termos da relação sinal-ruído nas entradas primária e de referência como:

$$S_{i_{out}}(z) \cong S_i(z) \left| \rho_{ref}(z) \right| \left| \rho_{pri}(z) \right| \quad (2.140)$$

A interpretação desse resultado pode ser feita da seguinte maneira:

- i) a potência do ruído na saída é proporcional à potência do ruído na entrada;
- ii) para uma baixa relação sinal-ruído na referência (mais ruído que sinal), a potência do ruído na saída é baixa, pois o filtro se adapta melhor para cancelar o ruído correlacionado;
- iii) para uma baixa relação sinal-ruído na entrada primária, a potência do ruído de saída também é baixa. Este resultado quer dizer que o filtro será treinado para eliminar correlações entre os ruídos de modo mais efetivo, e não procurará correlações entre sinal desejado e ruído.

Assim, a presença de um baixo nível de sinal desejado na entrada de referência não inviabiliza o uso do cancelador adaptativo de ruído. Note-se que, neste caso, é necessário que exista ruído na entrada de referência; caso contrário, o sinal desejado será cancelado completamente. Essa situação, no entanto, não deve ocorrer com a escolha própria dos sinais de entrada do cancelador.

2.5.4 Problemas práticos

Apesar de não ser possível implementar filtros ideais não-causais com estrutura transversal infinitamente longa, podem ser feitas aproximações desses filtros com possibilidade de implementação utilizando técnicas de processamento digital de sinais. Este procedimento é conhecido como técnica de janelamento e é usado para o projeto de filtros FIR de estrutura direta [33], que consiste do truncamento da resposta ao impulso do filtro ideal e do atraso da

resposta truncada no tempo. A necessidade de inclusão deste atraso implica na introdução de um atraso no caminho direto da entrada primária para a saída do cancelador, para não mudar a resposta ótima do filtro calculado. Widrow [46], recomenda um atraso equivalente à metade do comprimento do filtro adaptativo implementado. O uso de mais de uma entrada de referência é também discutido por Widrow [46].

2.6 Predição linear

Neste tipo de problema, assume-se que um sinal observado é gerado por uma combinação linear de seus valores passados e de valores passados e presente de um outro sinal, entrada do sistema hipotético (modelo) que gerou o sinal.

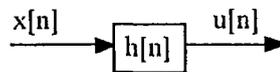


FIGURA 2.14: Modelo de geração de sinal

Na figura 2.14, $u[n]$ é o sinal a ser modelado, $x[n]$ é o sinal de entrada para o sistema hipotético cuja função de transferência é dada por H . O modelo assumido é

$$u[n] = -\sum_{k=1}^p a_k u[n-k] + G \sum_{l=0}^q b_l x[n-l] \quad (2.141)$$

onde a_k , b_l e G representam os parâmetros deste modelo. O sinal $u[n]$ é, então, predizível através de uma combinação linear de entradas e saídas, daí o nome predição linear [30]. Se for tomada a transformada Z deste sistema, verifica-se que a função de transferência $H(z)$ é a função de transferência de um sistema IIR, ou seja, possui pólos e zeros. O modelo recebe nome de acordo com a presença de pólos e zeros, ou seja, temos o modelo chamado de "somente pólos", "somente zeros" ou "pólo zero". Na literatura estatística estes modelos recebem os nomes *moving average model* (MA), *autoregressive model* (AR) e *autoregressive moving average model* (ARMA), respectivamente [30].

Se for assumido que o sinal $x[n]$ é completamente desconhecido, só é possível gerar uma aproximação ao sinal $u[n]$. Esta aproximação é acompanhada por um erro de predição definido por:

$$e[n] = u[n] - \hat{u}[n] \quad (2.142)$$

A estimativa dos parâmetros do modelo é feita assumindo um tipo de modelo (normalmente um modelo "somente pólos" [30]) e usando o método dos mínimos quadrados

para efetuar a minimização da função de custo baseada no erro de predição. Podem ser definidos dois tipos básicos de predição: "predição para frente" (*forward prediction*) e "predição para trás" (*backward prediction*). Estas definições podem ser melhor explicadas do seguinte modo: dado um conjunto de amostras do sinal $u[n]$, tais como $u[n-1], \dots, u[n-p]$, usamos estes pontos para fazer uma predição de como o sinal s vai ser no futuro, isto é, estimamos $u[n]$ (predição "para frente"). Também pode-se usar os mesmos pontos para estimar $u[n-p-1]$, um ponto no passado (predição "para trás"). A ordem do preditor é definida como o número de elementos de atraso necessários para armazenar o conjunto de amostras usado para fazer a predição [18]. O filtro que opera sobre o conjunto de amostras para produzir o erro de predição em sua saída é chamado de filtro de erro de predição "para frente" ou "para trás" (*forward/backward prediction error filter*) ou simplesmente filtro de predição.

O preditor para frente é definido por

$$\hat{u}[n] = \sum_{k=1}^p \alpha_k u[n-k] \quad (2.143)$$

e o filtro de erro de predição para frente é dado por

$$e_f[n] = u[n] - \hat{u}[n] = u[n] - \sum_{k=1}^p \alpha_k u[n-k] \quad (2.144)$$

cujas implementações usando filtros FIR de estrutura direta podem ser vistas nas figuras abaixo.

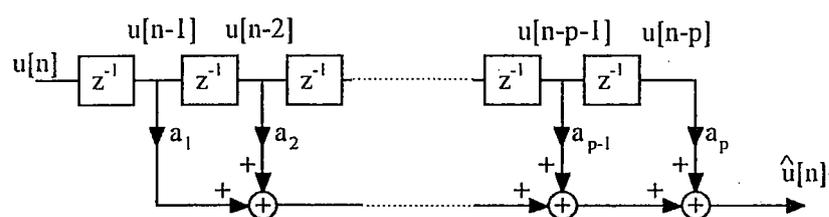


FIGURA 2.15: Preditor "para frente"

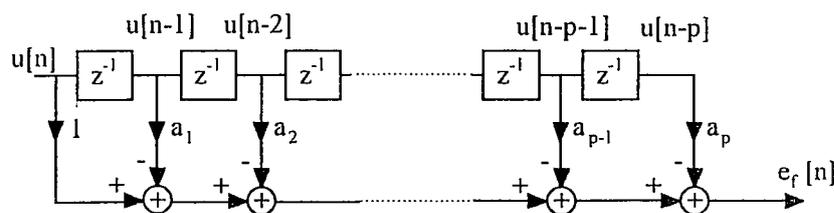


FIGURA 2.16: Filtro de erro de predição para frente

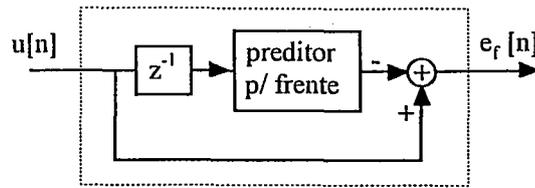


FIGURA 2.17: Diagrama em blocos do filtro de erro de predição para frente

O preditor para trás é definido por

$$\hat{u}[n-p] = \sum_{k=1}^p a_k u[n-k+1] \quad (2.145)$$

e o filtro de erro de predição para trás é dado por

$$e_b[n] = u[n-p] - \hat{u}[n-p] = u[n-p] - \sum_{k=1}^p a_k u[n-k+1] \quad (2.146)$$

cujas implementações usando filtros FIR de estrutura direta pode ser vista nas figuras abaixo.

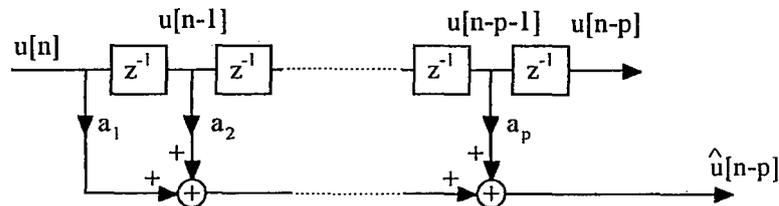


FIGURA 2.18: Preditor "para trás"

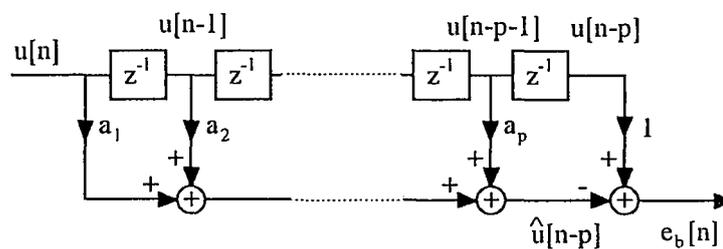


FIGURA 2.19: Filtro de erro de predição para trás

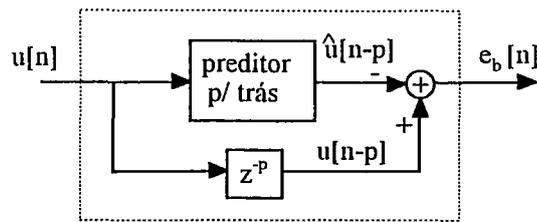


FIGURA 2.20: Diagrama em blocos do filtro de erro de predição para trás

Segundo Therrien [42], se a ordem do filtro de erro de predição crescer com cada nova amostra (observação) do sinal de entrada, a seqüência dos erros de predição resultante é ortogonal. Note-se que, para que o filtro cresça, um novo problema de predição linear deve ser resolvido para cada nova amostra. Esta propriedade de ortogonalização dos erros é conhecida como *whitening property*.

2.6.1 Preditores *lattice*

Embora preditores possam ser implementados com estrutura direta e *lattice*, preditores *lattice* possuem vantagens como filtros digitais. No contexto de predição linear, uma das vantagens do filtro de erro de predição *lattice* de ordem N é que ele contém todos os filtros de erro de predição de ordem inferior, ou seja, os sinais que se propagam pela estrutura são os erros de predição dos filtros de mais baixa ordem.

Os erros de predição para trás de ordem zero até a ordem M podem ser escritos em função dos filtros de erro de predição de forma direta. A relação escrita de forma matricial é

$$\mathbf{b}[n] = \mathbf{L} \mathbf{u}[n] \quad (2.145)$$

onde

$$\mathbf{b}[n] = [e_{b,0}[n] \quad e_{b,1}[n] \quad \dots \quad e_{b,M}[n]]^T \quad (2.146)$$

$$\mathbf{u}[n] = [u[n] \quad u[n-1] \quad \dots \quad u[n-M]]^T \quad (2.147)$$

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ a_{1,1} & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{M,M} & a_{M,M-1} & \dots & 1 \end{bmatrix} \quad (2.148)$$

onde o vetor $\mathbf{b}[n]$ é o vetor formado pelos erros de predição para trás de ordens sucessivas até M , $\mathbf{u}[n]$ é o vetor com os dados de entrada até o instante e \mathbf{L} é a matriz que representa a transformação efetuada pelos sucessivos filtros de erro de predição. Os elementos $a_{i,j}$ de \mathbf{L} são os coeficientes dos filtros de erro de predição conjugados. A transformação representada por (2.145) recebe o nome de algoritmo de ortogonalização de Gram-Schmidt [18]. Os erros de predição para trás são todos ortogonais para processos de entrada WSS [18].

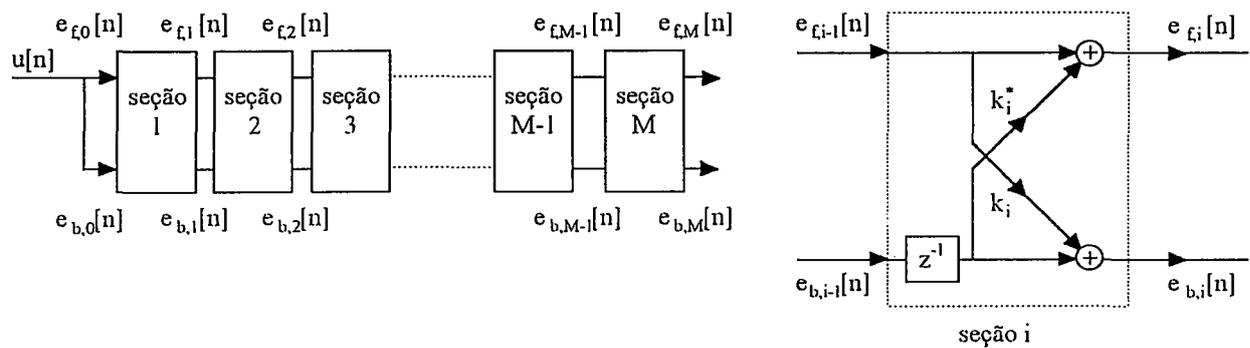
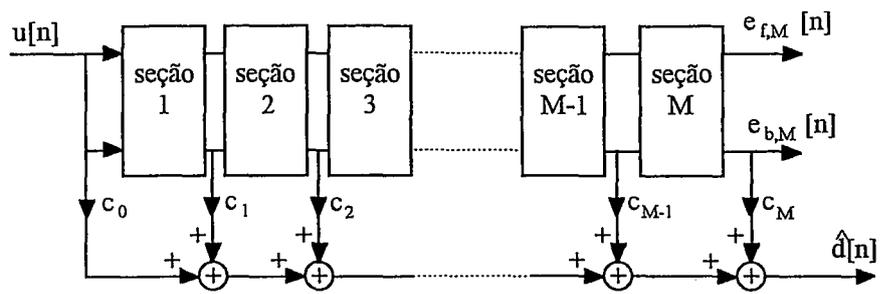


FIGURA 2.21: Filtro de erro de predição *lattice*

A implementação do algoritmo de Gram-Schmidt pode ser feita pelo empilhamento de $M+1$ filtros de erro de predição. Outro modo é usar um preditor *lattice* de ordem M , com M estágios, conforme a figura 2.21, onde $f_i[n]$ é o erro de predição para a frente, $b_i[n]$ é o erro de predição para trás e os parâmetros k_i são chamados de coeficientes de reflexão [30]. As principais vantagens do filtro de predição *lattice* são:

- i) a eficiência na geração simultânea dos erros de predição para a frente e para trás;
- ii) o desacoplamento entre os estágios;
- iii) o fato de que os erros de predição para trás são mutuamente ortogonais para processos WSS;
- iv) a modularidade, isto é, a adição de mais um módulo não afeta os demais;
- v) a similaridade entre os estágios.

O filtro de erro de predição produz sinais que são erros de predição. O preditor *lattice*, que gera uma estimativa a um sinal $d[n]$ (considerado como um processo conjuntamente estacionário com $u[n]$), utiliza-se de um filtro de predição *lattice* e de uma estrutura de regressão múltipla que usa como entrada os erros de predição para trás, conforme a figura 2.22 [18]. A esta estrutura pode ser aplicada a teoria de filtragem ótima. Detalhes dessa análise podem ser vistos em Haykin [18]. A derivação da estrutura *lattice* é vista em diversas referências [18] [42] [2].

FIGURA 2.22: Preditor *lattice*

3 FILTROS ADAPTATIVOS

3.1 Introdução

Um sistema adaptativo é um tipo de sistema que possui estrutura ajustável ou alterável de tal modo que o seu comportamento ou performance (de acordo com algum critério desejado) melhore quando em contato com o ambiente onde deve operar. Um filtro adaptativo é um tipo de sistema adaptativo que ajusta o seu comportamento, ou seja, sua resposta ao impulso, de modo automático. Este ajuste das características do filtro é feito com base num critério de performance pré-estabelecido pelo projetista. Este critério de performance define uma função de custo que é minimizada pelo processo de ajuste de parâmetros para atingir a melhor performance possível.

Um filtro adaptativo é usado em situações onde o projetista não conhece ou não pode prever as características do ambiente de operação do filtro. Neste caso o filtro se ajusta automaticamente ao meio, através do algoritmo de adaptação adequado, "aprendendo" as características desse meio. Note-se que, se o projetista soubesse as características do ambiente de funcionamento do filtro, ele mesmo poderia projetar um filtro fixo para a resolução do problema. Mesmo assim, se as características desse meio mudam, é necessário usar um filtro adaptativo para aprender as novas características do meio e ajustar-se novamente. A capacidade de acompanhamento das mudanças do ambiente pelo filtro é determinada pelo algoritmo e processos relacionados. O tipo de estrutura do filtro e seu algoritmo são determinados basicamente pelo tipo de aplicação em que este será usado.

A figura 3.1 mostra um diagrama genérico de como é estruturado um filtro adaptativo. O uso de um sinal desejado é comum, mas pode ser dispensado se não houver necessidade sua pelo algoritmo de adaptação (como o algoritmo CMA [43]). Em geral, este bloco é constituído pela simples subtração do sinal de saída do filtro do sinal desejado, tendo o sinal de erro como resultado. O filtro é o circuito que processa o sinal de entrada propriamente dito. Na utilização de técnicas de DSP, este bloco é a representação da estrutura do filtro digital. Conforme discutido no capítulo 2, estas estruturas dependem do tipo de resposta ao impulso do filtro. As regras de adaptação constituem o algoritmo de adaptação dos coeficientes. Esta é uma parte crítica do filtro adaptativo do ponto de vista computacional, pois realiza todo o processamento necessário para formar os novos coeficientes ou parâmetros do filtro. Normalmente, algoritmos de filtros adaptativos produzem resultados novos a cada nova amostra disponível. As limitações de *hardware* limitam o uso de algoritmos e estruturas em aplicações em tempo real.

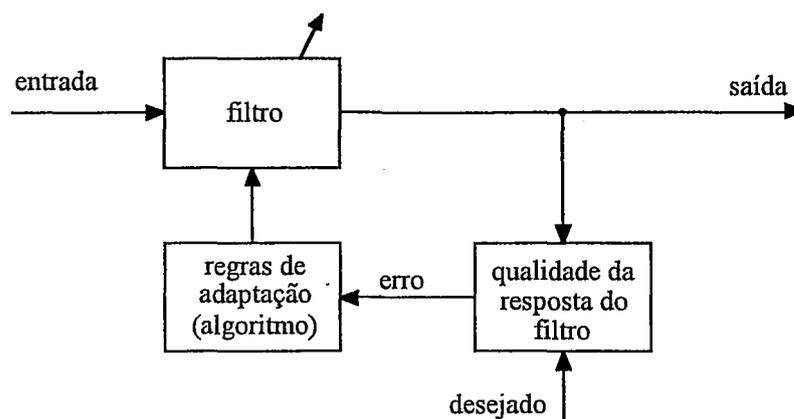


FIGURA 3.1: Estrutura geral de um filtro adaptativo e sinais envolvidos

As classes de aplicações de filtros adaptativos podem ser divididas em 4 esquemas básicos [49] [18]: identificação, modelamento inverso, predição e cancelamento de interferência. A figura 3.2 ilustra as estruturas destas aplicações.

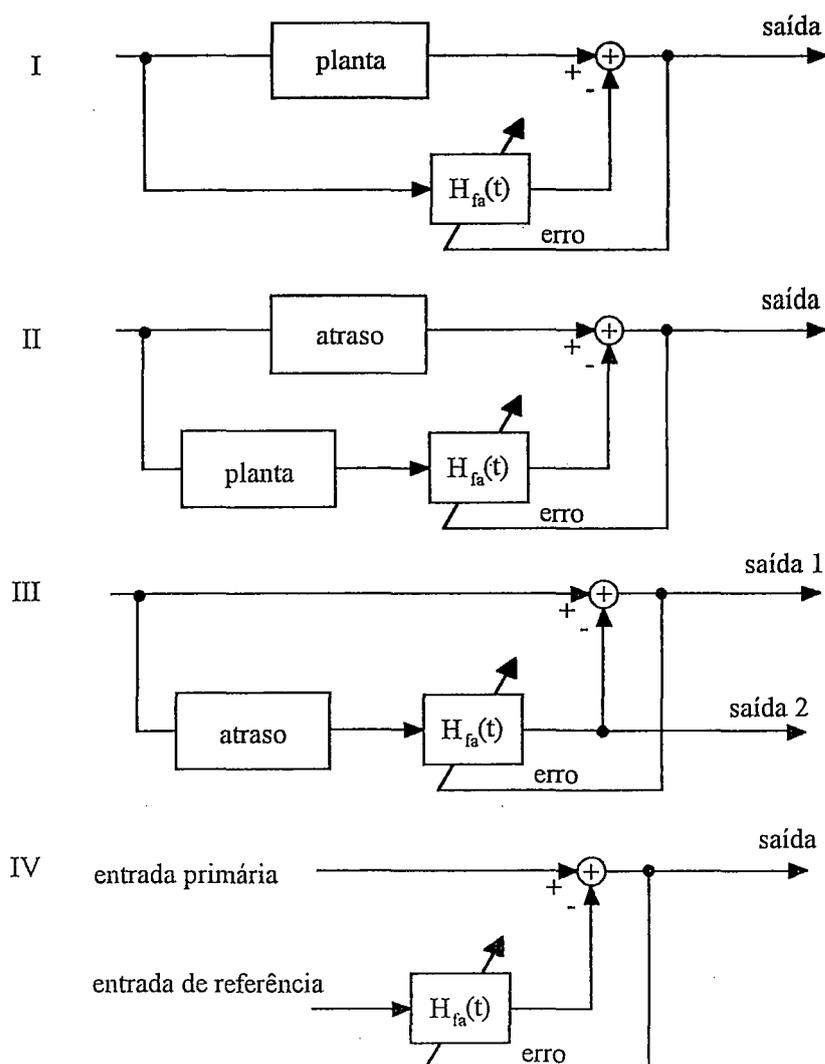


FIGURA 3.2: As quatro configurações básicas de filtros adaptativos

Na aplicação (i), identificação, deseja-se identificar a planta, utilizando o filtro adaptativo como modelo. A minimização do erro leva a função de transferência do filtro o mais próximo possível da função de transferência da planta, levando em conta a ordem e o tipo de resposta ao impulso. Os parâmetros do filtro são usados como saída neste caso. Este tipo de aplicação possibilita identificar plantas variáveis no tempo, sendo exemplos de aplicação encontrados em controle, na identificação de sistemas, em sismologia e comunicações.

O objetivo na aplicação (ii), modelamento inverso, é descobrir a função de transferência inversa (modelo inverso) da planta, ao invés de descobrir a própria função de transferência. O atraso pode ser nulo neste caso. Esta configuração é utilizada em equalização adaptativa. Para o sistema (iii), predição, o objetivo é prever o valor do sinal (saída 2) ou o erro de predição (saída 1). Exemplos de aplicações são em LPC (*Linear Predictive Coding*), ADPCM (*Adaptive Differential Pulse Code Modulation*) [45], detecção de sinais e análise espectral autoregressiva.

A última classe de aplicação, cancelamento de interferência (iv), tem por objetivo cancelar a interferência contida na entrada primária, baseado na estimativa da interferência fornecida pela entrada de referência. Esta é a estrutura de interesse neste trabalho. Cancelamento adaptativo de interferências, cancelamento de eco e formação adaptativa do diagrama de recepção de antenas (*adaptive beamforming*) são exemplos de aplicação desta configuração. Note-se que o algoritmo está dentro do bloco do filtro adaptativo nos diagramas da figura 3.2.

Alguns conceitos importantes para a análise de algoritmos para filtros adaptativos são relacionados a seguir [18]:

- i) taxa de convergência - é o número de iterações necessárias para que o algoritmo leve a resposta ao impulso do filtro adaptativo à solução ótima (de Wiener ou mínimos quadrados), quando em resposta à sinais estacionários, ou o mais perto possível dessa situação. Uma alta taxa de convergência significa uma velocidade de adaptação alta do algoritmo.
- ii) desajustamento (*misadjustment*) - este parâmetro provê uma estimativa da diferença entre o mínimo erro quadrado médio do filtro ótimo e o mínimo erro quadrado médio do conjunto de filtros adaptativos (em um ambiente estacionário).
- iii) rastreamento (*tracking*) - é a capacidade de o algoritmo seguir variações nas estatísticas dos sinais, influenciada pela taxa de convergência e pelo desajustamento.
- iv) robustez - é a capacidade de operar satisfatoriamente com dados de entrada mal condicionados, ou seja, que levam a uma matriz de correlação \mathbf{R} ou Φ mal condicionada.

- v) carga computacional - o número de operações (somas, produtos, etc) necessários para a realização de uma iteração do algoritmo; a memória usada; investimento para programar o algoritmo.
- vi) estrutura - como flui a informação pelo algoritmo.
- vii) propriedades numéricas - erros de quantização resultantes de conversão AD e representação interna; estabilidade numérica do algoritmo; precisão numérica do algoritmo.
- viii) estabilidade - se o algoritmo converge para a solução ótima.
- ix) curva de aprendizagem - o comportamento da função de custo no tempo, para uma família de filtros adaptativos com mesmas características, mesmo algoritmo e operando em ambiente semelhante.

3.2 Classificação

Devido ao elevado número de estruturas e algoritmos, é difícil a classificação de filtros adaptativos em geral. No entanto, os principais algoritmos podem ser vistos como pertencentes a uma das seguintes classes: mínimos quadrados ou *gradient descent* (descida pelo gradiente). Os algoritmos buscam atingir o ponto mínimo da superfície de performance ($J \times w_i$). Os algoritmos que fazem essa busca através do gradiente possuem como ponto de partida o método de *steepest descent* (descida mais íngreme), que busca atingir o ponto mínimo da superfície de performance através de sucessivas iterações do algoritmo, de acordo com um parâmetro de convergência que controla as correções aplicadas ao vetor de coeficientes. Essa classe de algoritmos, ao contrário do método de *steepest descent*, usa uma estimativa para o gradiente baseada nos dados de entrada. Dependendo de como é feita esta estimativa, há uma grande classe de algoritmos derivados. Note-se que, uma vez submetido à um ambiente estacionário, a solução ótima do filtro é a solução de Wiener. Note-se também que, nesse sentido, assume-se que os sinais são realizações de processos estocásticos. Em decorrência da estimativa do gradiente e do caráter recursivo (iterativo), estes algoritmos são os mais simples em termos de requerimentos computacionais e estrutura do algoritmo.

Os algoritmos que buscam o ponto mínimo através do método de mínimos quadrados tem como princípio a resolução direta das equações normais (capítulo 2). Nesse sentido, a solução a cada iteração é a ótima, a não ser pela qualidade das médias temporais (estimativas) usadas. No entanto, devido ao elevado número de amostras necessários para as estimativas, e como é necessário calcular uma nova solução a cada iteração com os novos dados, a carga computacional é bastante elevada e os algoritmos possuem uma estrutura complexa em relação à classe de algoritmos derivados do gradiente. Em contrapartida, possuem como vantagem a taxa de convergência mais elevada. Note-se que estes comentários não estão vinculados ao

modelo de sistema ou estrutura do filtro, mas somente à classe de origem dos algoritmos. Outras classes de algoritmos também podem ser citados, como *random search* [49] e *property restoral* [43].

Filtros adaptativos possuem limites de estabilidade nos seus algoritmos devido à alteração de parâmetros e também devido ao caráter recursivo. O uso de estruturas FIR contribui para facilitar a implementação de filtros adaptativos devido ao fato de que filtros FIR são intrinsecamente estáveis (no sentido que um sinal limitado na sua entrada gera um sinal limitado na sua saída - estabilidade BIBO - *bounded input bounded output*). Além disso, possuem uma função de custo que possui um único mínimo global para onde o algoritmo deve convergir, no caso de sinais de entrada WSS. Esta situação fez com que o uso de sistemas adaptativos FIR fosse disseminado em relação a filtros IIR, que não possuem estas características.

No contexto de estruturas para a implementação de filtros adaptativos, há maior interesse na implementação em forma direta e em forma *lattice* (grade). Para sistemas FIR, a forma direta apresenta os resultados mais simples. No entanto, um dos principais problemas sofridos por algoritmos para esta estrutura é a sua sensibilidade ao mau condicionamento da matriz de correlação \mathbf{R} (razão de $\lambda_{\max}/\lambda_{\min}$ alta), que afeta a velocidade de convergência. Este problema é resolvido em parte pelo uso da estrutura *lattice*, que é pouco sensível a esta perturbação [11] [9]. No entanto, o uso destas estruturas aumenta tremendamente a carga computacional e a complexidade dos algoritmos.

Para sistemas IIR, uma das vantagens da estrutura *lattice* é a possibilidade de garantir a estabilidade se os coeficientes k forem menores do que 1 (um) em módulo [36]. No entanto, o uso de esquemas para a forma direta é mais divulgado. Neste trabalho será tratada apenas a forma direta. Veja Fan [9] para uma análise de classes de algoritmos IIR e seus problemas de velocidade de convergência.

Quadro geral de classificação (principais classes):

modelo de sistema	estrutura	classe de algoritmo	representante
FIR	direta	<i>gradient descent</i>	LMS
		mínimos quadrados	RLS
			<i>fast RLS</i>
	<i>lattice</i>	<i>gradient descent</i>	GAL
		mínimos quadrados	LSL
			RLSL
IIR	direta	<i>gradient descent</i>	IIR LMS
		hiper-estabilidade	HARF, SHARF
	<i>lattice</i>		

3.3 Método de *steepest descent*

Devido ao fato de o método de *steepest descent* ser a origem da classe de algoritmos de busca de gradiente, seu estudo confere familiaridade com conceitos e comportamentos destes algoritmos. O método é recursivo no sentido que, partindo de um dado ponto na superfície de performance, conforme aumenta o número de aproximações, melhora a estimativa dos coeficientes em direção ao filtro de Wiener. Segundo Haykin [18], o método de *steepest descent* descreve um sistema de controle multivariável, realimentado e determinístico, que encontra o ponto mínimo da superfície de performance J sem o conhecimento da mesma.

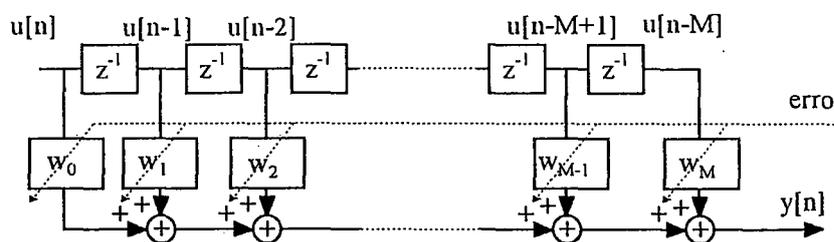


FIGURA 3.3: Filtro adaptativo FIR

Considere um filtro adaptativo FIR que usa estrutura direta e que os sinais são realizações de processos estocásticos WSS. A cada instante, o filtro adaptativo da figura 3.3 muda a sua resposta ao impulso $\mathbf{W}[n]$. A resposta ao impulso ótima é o filtro de Wiener:

$$\mathbf{W}_o^* = \mathbf{R}\mathbf{p}^{-1} \quad (3.1)$$

Onde

$$\mathbf{W}^* = [w_0^* \quad w_1^* \quad \dots \quad w_{M-1}^*]^T \quad (3.2)$$

e

$$\mathbf{u}[n] = [u[n] \quad u[n-1] \quad \dots \quad u[n-M+1]]^T \quad (3.3)$$

A saída do filtro é definida por:

$$y[n] = \mathbf{W}^T[n]\mathbf{u}[n] \quad (3.4)$$

O sinal de erro é dado por:

$$e[n] = d[n] - \mathbf{W}^T[n]\mathbf{u}[n] \quad (3.5)$$

e a função de custo é definida por

$$J[n] = \sigma_d^2 - \mathbf{W}^T[n]\mathbf{p} - \mathbf{p}^H \mathbf{W}^*[n] + \mathbf{W}^T[n]\mathbf{R}\mathbf{W}^*[n] \quad (3.6)$$

que indica a mudança do erro com a variação dos coeficientes, o que significa que o processo do erro de estimação $e[n]$ é não-estacionário. Apesar de $J[n]$ ser variável, o seu mínimo é

$$J_{\min} = \sigma_d^2 - \mathbf{p}^H \mathbf{W}_o^* \quad (3.7)$$

O algoritmo é o seguinte:

1. escolhe-se um vetor $\mathbf{W}[0]$ de modo arbitrário como valor inicial. Normalmente é zero.
2. usando a estimativa atual de $\mathbf{W}[n]$, calcula-se o gradiente de $J[n]$ com respeito a $\mathbf{W}[n]$.
3. calcula-se a próxima estimativa para $\mathbf{W}[n]$, $\mathbf{W}[n+1]$, fazendo uma mudança em $\mathbf{W}[n]$ na direção oposta ao vetor gradiente.
4. volta ao passo 2

O algoritmo é intuitivamente razoável se for pensado no caso de J bidimensional e com apenas um coeficiente. Denotando o vetor gradiente de $J[n]$ por $\nabla(J[n])$, a correção aplicada ao vetor de coeficientes pelo método é

$$\mathbf{W}^*[n+1] = \mathbf{W}^*[n] + \mu(-\nabla(J[n])) \quad (3.8)$$

onde μ é uma constante positiva e o gradiente de $J[n]$ é dado por (capítulo 2):

$$\nabla(J[n]) = \begin{bmatrix} \frac{\partial J[n]}{\partial a_0[n]} + j \frac{\partial J[n]}{\partial b_0[n]} \\ \frac{\partial J[n]}{\partial a_1[n]} + j \frac{\partial J[n]}{\partial b_1[n]} \\ \vdots \\ \frac{\partial J[n]}{\partial a_{M-1}[n]} + j \frac{\partial J[n]}{\partial b_{M-1}[n]} \end{bmatrix} = -2\mathbf{p} + 2\mathbf{R}\mathbf{W}^*[n] \quad (3.9)$$

Como \mathbf{R} e \mathbf{p} são assumidos conhecidos, pode-se calcular o gradiente de $J[n]$ exatamente, o que leva à forma final do algoritmo de *steepest descent*:

$$\mathbf{W}^*[n+1] = \mathbf{W}^*[n] + 2\mu(\mathbf{p} - \mathbf{R}\mathbf{W}^*[n]) \quad (3.10)$$

para $n=0,1,2,\dots$. Esta é a expressão matemática do método. Note-se que a constante μ controla o tamanho da correção incremental aplicada ao vetor de coeficientes $\mathbf{W}[n]$ a cada iteração, sendo chamada de parâmetro de *step-size* (tamanho de passo).

3.3.1 Análise de estabilidade

Como o algoritmo possui realimentação, há a possibilidade de se tornar instável. Os parâmetros que controlam a realimentação são o parâmetro μ e a matriz de correlação \mathbf{R} o que pode ser visto na figura 3.4.

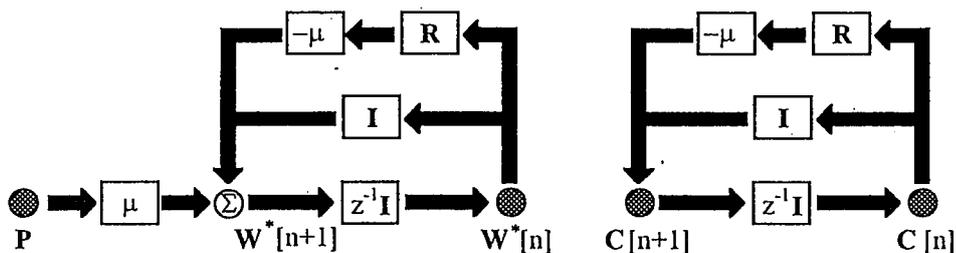


FIGURA 3.4: *Steepest descent* mostrando $\mathbf{W}[n]$ e $\mathbf{C}[n]$

Definindo o vetor de erro de coeficientes como

$$\mathbf{C}[n] = \mathbf{W}^*[n] - \mathbf{W}_o^*[n] \quad (3.11)$$

pode-se escrever

$$\mathbf{C}[n+1] = (\mathbf{I} - \mu\mathbf{R})\mathbf{C}[n] \quad (3.12)$$

onde \mathbf{I} é a matriz identidade. Definindo \mathbf{Q} como a matriz cujas colunas são os autovetores de \mathbf{R} (idem ao capítulo 2), e definindo o vetor

$$\mathbf{v}[n] = \mathbf{Q}^H \mathbf{C}[n] \quad (3.13)$$

chega-se ao seguinte resultado, que pode ser visto na figura 3.5 [18].

$$v_k[n+1] = (1 - \mu\lambda_k)^n v_k[n] \quad (3.14)$$

para $k=1,2,\dots,M$.

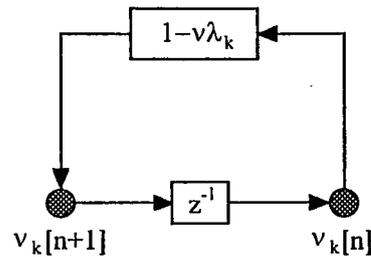


FIGURA 3.5: Realimentação de v_k

Como todos os autovalores de \mathbf{R} são reais e não-negativos, a resposta não oscila. Nota-se que esta resposta representa uma série geométrica com razão $(1-\mu\lambda_k)$. Para estabilidade e convergência esta razão deve ser menor em módulo que 1 para todo k . Isso significa que $\mathbf{W}[n]$ tende à solução de Wiener, $\mathbf{W}_o[n]$, quando $n \rightarrow \infty$.

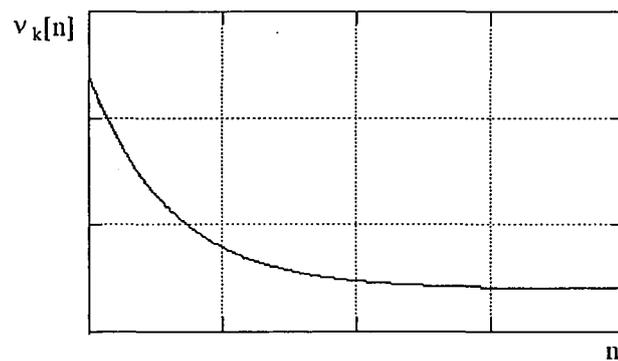


FIGURA 3.6: Curva de aprendizagem de v_k

Pode-se derivar desta expressão os limites para a constante μ

$$-1 < 1 - \mu\lambda_k < 1 \quad (3.15)$$

$$0 < \mu < \frac{2}{\lambda_k} \quad (3.16)$$

$$0 < \mu < \frac{2}{\lambda_{\max}} \quad (3.17)$$

Onde a expressão é válida para todo k e o autovalor máximo de \mathbf{R} é representado por λ_{\max} . A constante de tempo do k -ésimo modo é expressa por

$$1 - \mu\lambda_k = \exp(-1/\tau_k) \quad (3.18)$$

$$\tau_k = \frac{-1}{\ln(1 - \mu\lambda_k)} \approx \frac{-1}{\mu\lambda_k} \quad (3.19)$$

onde a aproximação é válida se μ for pequeno. Esses limites não indicam quanto tempo o algoritmo irá levar para convergir - isso dependerá do número de modos (coeficientes). No entanto, a constante de tempo global, τ_a , é limitada pelas constantes de tempo do maior e menor autovalores [18].

$$\frac{-1}{\ln(1 - \mu\lambda_{\max})} \leq \tau_a \leq \frac{-1}{\ln(1 - \mu\lambda_{\min})} \quad (3.20)$$

Pode-se expandir $\mathcal{J}[n]$ em [18]:

$$\mathcal{J}[n] = J_{\min} + \sum_{k=1}^M \lambda_k |v_k(n)|^2 = J_{\min} + \sum_{k=1}^M \lambda_k (1 - \mu\lambda_k)^{2n} |v_k(0)|^2 \quad (3.21)$$

de onde se conclui:

- i) se μ é escolhido de modo a deixar o algoritmo convergente, $\mathcal{J}[n] \rightarrow J_{\min}$ quando $n \rightarrow \infty$, independentemente das condições iniciais;
- ii) a curva de $(\mathcal{J}[n] \times n)$ é chamada de curva de aprendizagem (*learning curve*). Nesta situação ela consiste de uma soma de exponenciais, cada uma correspondendo a um modo natural do algoritmo (número de modos \equiv número de coeficientes, para filtro FIR de estrutura direta);
- iii) a constante de tempo de decaimento de $\mathcal{J}[0]$ até J_{\min} , para o k -ésimo modo natural, é:

$$\tau_{k,\text{mse}} \approx \frac{-1}{2 \ln(1 - \mu\lambda_k)} \quad (3.22)$$

Nota-se que, quanto menor μ , mais lento é o decaimento do modo.

3.4 Algoritmo LMS

O algoritmo LMS (*Least-Mean-Squares* - a menor média do quadrado) [46] [49] é um membro importante da família de algoritmos baseada no gradiente. Devido a sua simplicidade

matemática e de implementação aliadas a uma boa performance, ele é o algoritmo padrão usado em comparações com os demais algoritmos de filtros adaptativos.

O algoritmo LMS pode ser visto como sendo derivado do método de *steepest descent* através do uso de uma estimativa simplificada para o gradiente. Como o método de *steepest descent*, o algoritmo LMS também representa um sistema de controle realimentado, formado pelos processos de filtragem e adaptação dos coeficientes. A estimativa usada pelo algoritmo LMS para o gradiente é baseada nos dados de entrada do filtro. A simplificação consiste em usar o produto $u[n-k]e^*[k]$ como estimativa para $\varepsilon\{u[n-k]e^*[k]\}$, o k -ésimo componente do gradiente. Este fato ocasiona um erro na estimativa do gradiente a cada iteração. Diz-se então que o algoritmo LMS possui um gradiente ruidoso.

Esse gradiente ruidoso faz com que o comportamento do algoritmo LMS difira do método de *steepest descent* na busca do ponto mínimo da superfície de performance (solução de Wiener). O algoritmo LMS, em vez de terminar na solução ótima, executa um movimento aleatório na região em torno do ponto mínimo. Esse fato faz com que o erro mínimo obtido pelo algoritmo LMS em regime permanente ($J[\infty]$), seja maior que o erro mínimo do método de *steepest descent*. Segundo Haykin [18], "esta diferença representa o preço pago por usar um mecanismo adaptativo estocástico de controle dos coeficientes, em lugar de uma abordagem determinística como no método de *steepest descent*". A diferença $J_{\text{lms}}[\infty] - J_{\text{min}}$ é chamada de erro quadrado médio em excesso, $J_{\text{ex}}[\infty]$. A razão $J_{\text{ex}}[\infty]/J_{\text{min}}$ é chamada de desajustamento. Saliente-se que o fator de convergência μ , como parâmetro do projetista, pode ser escolhido de modo a controlar a taxa de convergência e o desajustamento, se estiver dentro dos limites de estabilidade.

3.4.1 Derivação

Do método de *steepest descent*, vem

$$\mathbf{W}^*[n+1] = \mathbf{W}^*[n] + \mu(-\nabla(J[n])) \quad (3.23)$$

$$\mathbf{W}^*[n+1] = \mathbf{W}^*[n] + 2\mu(\mathbf{p} - \mathbf{R}\mathbf{W}^*[n]) \quad (3.24)$$

Estima-se \mathbf{p} e \mathbf{R} do modo mais simples possível, isto é, através da estimativa instantânea:

$$\hat{\mathbf{R}}[n] = \mathbf{u}[n]\mathbf{u}^H[n] \quad (3.25)$$

$$\hat{\mathbf{p}}[n] = \mathbf{u}[n]d^*[n] \quad (3.26)$$

E a estimativa do gradiente é:

$$\hat{\nabla}(J[n]) = -2\mathbf{u}[n]d^*[n] + 2\mathbf{u}[n]\mathbf{u}^H[n]\hat{\mathbf{W}}^*[n] \quad (3.27)$$

que também pode ser vista como a aplicação do gradiente à função de custo instantânea, $J = |e[n]|^2$, sem o operador $\varepsilon\{\}$. Substituindo esta estimativa no método de *steepest descent* leva ao algoritmo LMS:

$$\hat{\mathbf{W}}^*[n+1] = \hat{\mathbf{W}}^*[n] + \mu\mathbf{u}[n](d^*[n] - \mathbf{u}^H[n]\hat{\mathbf{W}}^*[n]) = \hat{\mathbf{W}}^*[n] + \mu\mathbf{u}[n]e^*[n] \quad (3.28)$$

A figura 3.7 ilustra o funcionamento do algoritmo em um blocodiagrama. O resumo do algoritmo é dado na tabela 3.1.

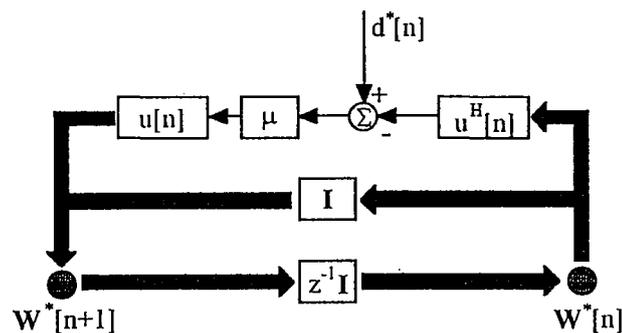


FIGURA 3.7: Estrutura de realimentação no algoritmo LMS

TABELA 3.1: Algoritmo LMS

DADOS:

M	n° de coeficientes
μ	parâmetro de adaptação
$u[n]$	dato de entrada no instante n
$d[n]$	resposta desejada no instante n
$y[n]$	saída do filtro no instante n

INICIALIZAÇÃO:

- 1) $\mathbf{W}[0] = \mathbf{0}$

LOOP:

- 2) $y[n] = \hat{\mathbf{W}}^T[n]\mathbf{u}[n]$
- 3) $e[n] = d[n] - y[n]$
- 4) $\hat{\mathbf{W}}^*[n+1] = \hat{\mathbf{W}}^*[n] + \mu\mathbf{u}[n]e^*[n]$

3.4.2 Análise de estabilidade

Devido à presença de não-linearidades (dependência de $\mathbf{u}^H[n]$), aleatoricidade (devido ao fato de $\mathbf{u}[n]$ e $\mathbf{d}[n]$ serem assumidos como realizações de processos estocásticos) e múltiplos parâmetros, a análise de estabilidade do algoritmo é difícil. Considera-se os aspectos de convergência na média (se a estimativa de $\mathbf{W}[n]$ tende para a solução de Wiener) e de convergência na média do quadrado (se $J[\infty] < \infty$). Para que seja possível a análise do algoritmo são feitas algumas hipóteses sobre a independência dos sinais e distribuição estatística dos vetores $\mathbf{u}[n]$ e $\mathbf{d}^*[n]$. A análise estatística resultante é chamada de "teoria da independência" [18]. A teoria da independência assume que a seqüência de vetores aleatórios que direcionam a busca do vetor de coeficientes ótimos sejam independentes estatisticamente. Embora longe da verdade, os resultados dessa teoria estão em concordância com experimentos e simulações [18].

Definindo o vetor de erro dos coeficientes como a diferença entre a solução atual e a solução ótima de Wiener:

$$\mathbf{E}[n] = \mathbf{W}^*[n] - \mathbf{W}^* \quad (3.29)$$

Substituindo esta equação na definição do algoritmo LMS e usando as hipóteses da teoria da independência, chega-se a

$$\varepsilon\{\mathbf{E}[n+1]\} = (\mathbf{I} - \mu\mathbf{R})\varepsilon\{\mathbf{E}[n]\} \quad (3.30)$$

Pela analogia com o método de *steepest descent*, chega-se à condição de convergência de $\mathbf{E}[n]$ à zero quando $n \rightarrow \infty$ ($\mathbf{W}[n]$ atinge a solução de Wiener na média):

$$0 < \mu < \frac{2}{\lambda_{\max}} \quad (3.31)$$

onde λ_{\max} é o maior autovalor de \mathbf{R} . Definindo a matriz $\mathbf{K}[n]$ como a matriz de correlação do vetor de erro dos coeficientes:

$$\mathbf{K}[n] = \varepsilon\{\mathbf{E}[n]\mathbf{E}^H[n]\} \quad (3.32)$$

Escrevendo o erro do algoritmo LMS como

$$e[n] = d[n] - \mathbf{W}^T[n]\mathbf{u}[n] \quad (3.33)$$

$$e[n] = e_o[n] - \mathbf{E}^H[n]\mathbf{u}[n] \quad (3.34)$$

onde $e_o[n]$ é o erro de estimação da solução ótima de Wiener. O desenvolvimento de $J_{lms}[n]$ leva a

$$J_{lms}[n] = J_{\min} + tr[\mathbf{R}[n]\mathbf{K}[n]] \quad (3.35)$$

onde o operador $tr[\mathbf{A}]$ representa o traço da matriz \mathbf{A} (o traço de uma matriz quadrada é definido como a soma dos elementos da diagonal principal). Esta equação salienta que a função de custo do algoritmo LMS possui dois componentes: J_{\min} , o erro mínimo que a solução ótima de Wiener consegue atingir e um componente que depende do comportamento transitório da matriz de correlação do vetor de erro dos coeficientes, $\mathbf{E}[n]$. Como esta matriz é positiva definida para todo n , o algoritmo produz um erro médio-quadrado a mais que o erro J_{\min} . Este erro é chamado de erro em excesso e denotado por J_{ex} .

$$J_{ex}[n] = J_{lms}[n] - J_{\min} = tr[\mathbf{R}[n]\mathbf{K}[n]] \quad (3.36)$$

A função de custo para o algoritmo LMS possui as seguintes propriedades [18]:

- i) as componentes transitórias de $J_{lms}[n]$ não exibem oscilação;
- ii) o erro médio-quadrado converge ao valor estacionário $J[\infty]$ se e somente se μ satisfaz

$$0 < \mu < \frac{2}{\lambda_{\max}} \quad (3.37)$$

$$\sum_{i=1}^M \frac{\mu\lambda_i}{2 - \mu\lambda_i} < 1 \quad (3.38)$$

para o algoritmo LMS real, a última condição é

$$\sum_{i=1}^M \frac{\mu\lambda_i}{2(1 - \mu\lambda_i)} < 1 \quad (3.39)$$

onde λ_i , $i=1,2,\dots,M$ são os autovalores da matriz de correlação \mathbf{R} e λ_{\max} é o maior deles;

iii) o erro médio-quadrado produzido pelo algoritmo LMS tem o valor final:

$$J_{lms}[\infty] = \frac{J_{\min}}{1 - \sum_{i=1}^M \frac{\mu\lambda_i}{2 - \mu\lambda_i}} \quad (3.40)$$

iv) o desajustamento é dado por:

$$M = \frac{J_{ex}[\infty]}{J_{\min}} = \frac{\sum_{i=1}^M \frac{\mu\lambda_i}{2 - \mu\lambda_i}}{1 - \sum_{i=1}^M \frac{\mu\lambda_i}{2 - \mu\lambda_i}} \quad (3.41)$$

Se $\mu \ll 2/\lambda_{\max}$ e se a curva de aprendizagem da média do algoritmo for assumida como exponencial, pode-se formar uma estimativa prática para o desajustamento

$$M \approx \frac{\mu N \lambda_{av}}{2} = \frac{N}{4\tau_{mse,av}} \quad (3.42)$$

$$\tau_{mse,av} = \frac{1}{2\mu\lambda_{av}} \quad (3.43)$$

onde N é o número de coeficientes do filtro e λ_{av} é a média aritmética dos autovalores de \mathbf{R} .

Dessa última equação se conclui:

- i) o desajustamento aumenta linearmente com o número de coeficientes (N);
- ii) o tempo de estabilização do algoritmo LMS é proporcional à constante de tempo médio, $\tau_{mse,av}$;
- iii) o desajustamento é inversamente proporcional à μ .

Outras observações podem ser feitas a respeito do algoritmo LMS em comparação com o método de *steepest descent*. A presença do termo J_{ex} é devido ao uso de uma estimativa para o gradiente, ao invés de usar o gradiente verdadeiro. A curva de aprendizado continua sendo uma soma de exponenciais, mas no caso do algoritmo LMS estas são exponenciais ruidosas. Para que seja construída a curva de aprendizado no algoritmo LMS, portanto, é necessário que se faça a média sobre um número grande de realizações do algoritmo, o que não acontece no método de *steepest descent* devido ao uso do gradiente verdadeiro.

Com relação à constante μ , a sua diminuição acarreta uma diminuição da velocidade de convergência do algoritmo, mas isso leva a um menor desajustamento. Este fato indica que a escolha de μ depende de quanto o projetista pode sacrificar de performance em regime permanente em favor de uma velocidade de convergência maior. Encontrar uma estimativa para o valor de μ em aplicações práticas é um problema que pode ser resolvido se é considerado que μ é pequeno suficiente ($\mu \ll 2/\lambda_{\max}$), situação que simplifica a condição de convergência na média do quadrado para

$$\frac{1}{\mu} > \frac{1}{2} \sum_{i=1}^N \lambda_i = \frac{1}{2} \left(\begin{array}{l} \text{potência total} \\ \text{de entrada} \end{array} \right) \quad (3.44)$$

$$\mu < 2 / \left(\begin{array}{l} \text{potência total} \\ \text{de entrada} \end{array} \right) \quad (3.45)$$

ou seja, μ deve ser escolhido menor que duas vezes o recíproco da potência do sinal de entrada.

Quando os autovalores de \mathbf{R} estão muito espalhados, ou seja, quando \mathbf{R} é mal condicionada, o erro em excesso em regime permanente, $J_{\text{ex}}[\infty]$, é determinado pelo maior autovalor e o tempo de convergência pelo menor. De um modo geral o algoritmo requer um número maior de iterações para convergir nesta condição.

3.4.3 Análise para ambiente não-estacionário

Quando um filtro adaptativo opera em um ambiente não-estacionário, o vetor de coeficientes ótimo assume uma forma variante com o tempo, onde o seu valor muda de iteração para iteração. Neste caso o algoritmo tem a necessidade de procurar o ponto de mínimo da superfície de performance, mas também precisa seguir a posição deste ponto, que muda continuamente com o tempo.

Se o vetor de erro dos coeficientes for escrito como

$$\mathbf{E}[n] = \mathbf{W}^*[n] - \mathbf{W}_o^*[n] \quad (3.46)$$

$$\mathbf{E}[n] = \left(\mathbf{W}^*[n] - \varepsilon \{ \mathbf{W}^*[n] \} \right) + \left(\varepsilon \{ \mathbf{W}^*[n] \} - \mathbf{W}_o^*[n] \right) \quad (3.47)$$

pode-se então identificar duas componentes [47] [18]:

- i) a diferença entre o vetor de coeficientes e sua média, devido a erros na estimativa do gradiente. Este termo é chamado "ruído no vetor de coeficientes", $\mathbf{E}_1 = \mathbf{W}^*[n] - \varepsilon \{ \mathbf{W}^*[n] \}$;
- ii) a diferença entre o vetor com o valor médio dos coeficientes e o valor ótimo, devido ao atraso no processo de adaptação, $\mathbf{E}_2 = \varepsilon \{ \mathbf{W}^*[n] \} - \mathbf{W}_o^*[n]$.

Desenvolvendo a fórmula para o erro em excesso (3.48) [18], nota-se que o resultado é uma soma de componentes devido ao erro \mathbf{E}_1 e \mathbf{E}_2 . Como a dependência com o fator de convergência μ é diferente para cada um destes erros ($\mathbf{E}_1 \propto \mu$ e $\mathbf{E}_2 \propto 1/\mu$), pode-se traçar o gráfico da figura 3.8 para o desajustamento em termos de μ . O valor ótimo é μ_o , que minimiza o desajustamento total.

$$J_{ex}[n] = J[n] - J_{min} = tr[\mathbf{R}[n]\mathbf{K}_1[n]] + tr[\mathbf{R}[n]\mathbf{K}_2[n]] \quad (3.48)$$

Segundo Haykin [18], esta análise que segue o trabalho de Widrow [47] só é válida se μ for pequeno.

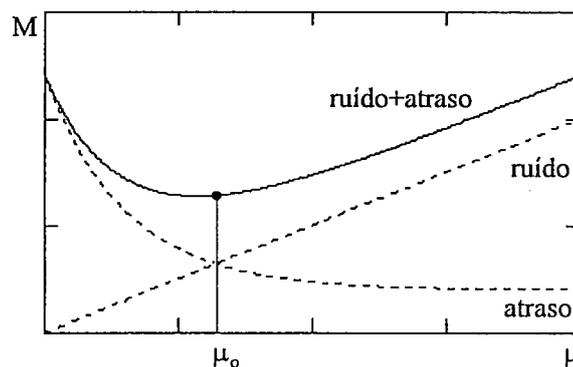


FIGURA 3.8: Desajustamento M e seus componentes devido a \mathbf{E}_1 (ruído) e \mathbf{E}_2 (atraso)

3.4.4 Algoritmos relacionados

Diversos algoritmos são derivados do algoritmo LMS. Alguns são extensões e outros simplificações suas. O algoritmo LMS aplica uma correção proporcional ao termo $\mathbf{u}[n]\mathbf{e}^*[n]$. Quando um destes termos for grande, há uma correção grande no gradiente. Este fato pode ser encarado como uma amplificação no ruído do gradiente. Uma tentativa de resolver este problema é feita através do desenvolvimento do algoritmo LMS normalizado [18], onde a correção aplicada é normalizada em relação à norma euclidiana de $\mathbf{u}[n]$. Essa normalização é feita pela introdução de um termo que pode ser visto como modificando o valor da constante de proporcionalidade μ . Logo, este algoritmo pode ser visto como tendo uma constante de convergência variável. Neste caso os limites para a estabilidade do fator de convergência são $0 < \mu < 2$ [18]. O fator de correção é proporcional à potência do sinal de entrada [43]. Em

casos onde esta for muito pequena, torna-se necessário introduzir um termo para prevenir a amplificação excessiva da correção (divisão por um número muito pequeno). A fórmula básica para o cálculo recursivo é

$$\hat{\mathbf{W}}^*[n+1] = \hat{\mathbf{W}}^*[n] + \frac{\mu}{\alpha + \|\mathbf{u}[n]\|^2} \mathbf{u}[n]e^*[n] \quad (3.49)$$

Segundo Treichler [43], pode-se estimar a norma de $\mathbf{u}[n]$ (potência) recursivamente.

Alguns algoritmos que visam reduzir ainda mais a carga computacional do algoritmo LMS utilizam apenas o sinal algébrico dos termos $\mathbf{u}[n]$, $e[n]$ ou de ambos, sendo chamados de algoritmo *sign-error*, *sign-data* ou *sign-sign* LMS [43]. A idéia básica é que se exclua o uso de multiplicações a um conjunto de somas e deslocamentos binários. No entanto, estas simplificações afetam muito o comportamento do algoritmo LMS normal, principalmente devido ao uso de uma estimativa pior para o gradiente, sofrendo então uma redução de performance e eventuais problemas de convergência. Apesar disso, segundo Treichler [43], o algoritmo *sign-sign* LMS é um algoritmo popular, sendo padrão da CCITT para ADPCM.

Outra forma de simplificação utiliza o algoritmo LMS mas somente um grupo selecionado de coeficientes é adaptado a cada iteração [31]. Deste modo reduz-se a carga computacional por iteração do algoritmo, mantendo o número de coeficientes. Maiores detalhes sobre este método serão tratados no capítulo 4. Outros problemas do algoritmo LMS, como desvios no período de convergência, *drift*, efeitos numéricos e *leakage* de coeficientes são discutidos em diversas referências [18] [43] [49].

3.5 Algoritmo Recursive Least Squares ou RLS

O método dos mínimos quadrados pode ser usado para calcular os coeficientes do filtro ótimo através da resolução das equações normais. Do ponto de vista adaptativo, é interessante que se tenha esta resposta a cada instante em que uma nova amostra de dados está disponível, ou seja, conhecendo $\mathbf{W}[n]$, $\mathbf{u}[n]$ e $d[n]$, deseja-se obter $\mathbf{W}[n+1]$, a nova solução ótima para os coeficientes.

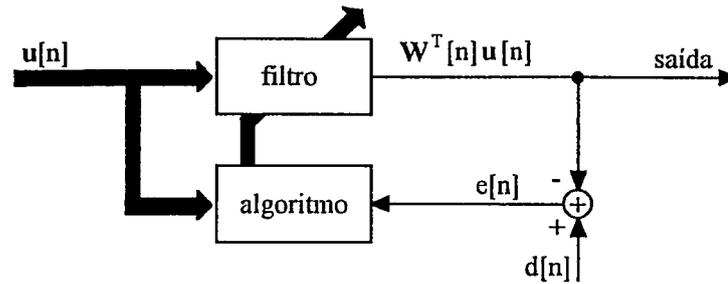


FIGURA 3.9: Sinais correspondentes ao algoritmo RLS

A forma mais simples consiste em utilizar o seguinte esquema [43]:

- 1) calcular a matriz de autocorrelação no instante $n+1$:

$$\Phi[n+1] = \Phi[n] + \mathbf{u}[n]\mathbf{u}^T[n] \quad (3.50)$$

- 2) calcular o vetor de correlação cruzada:

$$\Theta[n+1] = \Theta[n] + d[n]\mathbf{u}[n] \quad (3.51)$$

- 3) inverter a matriz de autocorrelação $\Phi[n+1]$;
- 4) calcular a solução diretamente:

$$\mathbf{W}^*[n+1] = \Theta[n+1]\Phi^{-1}[n+1] \quad (3.52)$$

Apesar de simples, este procedimento possui uma carga computacional de $O(N^3)$, onde N é o número de coeficientes do filtro. No algoritmo RLS, como a resolução das equações normais implica na inversão de $\Phi[n]$, é usado um resultado da álgebra matricial conhecido como lema ABCD: "Sejam \mathbf{A} e \mathbf{B} matrizes positivas definidas $M \times M$, \mathbf{C} matriz $M \times N$ e \mathbf{D} matriz positiva definida $N \times N$. Se $\mathbf{A} = \mathbf{B}^{-1} + \mathbf{C}\mathbf{D}^{-1}\mathbf{C}^H$ então $\mathbf{A}^{-1} = \mathbf{B} - \mathbf{B}\mathbf{C}(\mathbf{D} + \mathbf{C}^H\mathbf{B}\mathbf{C})^{-1}\mathbf{C}^H\mathbf{B}$ " [18]. A aplicação deste resultado leva à seguinte equação para o cálculo recursivo de Φ^{-1} [43] [18]:

$$\Phi^{-1}[n+1] = \Phi^{-1}[n] - \frac{\Phi^{-1}[n]\mathbf{u}[n]\mathbf{u}^H[n]\Phi^{-1}[n]}{1 + \mathbf{u}^H[n]\Phi^{-1}[n]\mathbf{u}[n]} \quad (3.53)$$

Desse modo, $\Phi[n]$ nunca é calculado e nem invertido diretamente. Como resultado, a carga computacional do algoritmo reduz-se a $O(N^2)$. A recursão para o cálculo do vetor de coeficientes é dada por [43] [18]:

$$\mathbf{W}^*[n+1] = \Phi^{-1}[n+1]\Theta[n+1] = \mathbf{W}^*[n] + \frac{(d[n] - y[n])^* \mathbf{Z}[n]}{1+q} \quad (3.54)$$

onde podem ser destacados os seguintes termos:

$$y[n] = \mathbf{u}^T[n]\mathbf{W}[n] \quad (3.55)$$

$$q = \mathbf{u}^H[n]\mathbf{Z}[n] \quad (3.56)$$

$$\mathbf{Z}[n] = \Phi^{-1}[n]\mathbf{u}[n] \quad (3.57)$$

saída *a priori*, potência de entrada normalizada e vetor de informação filtrado, respectivamente. A razão para o uso dos termos *a priori* e *a posteriori* refere-se ao instante de cálculo dos termos no algoritmo. Quando um novo dado entra para formar o vetor $\mathbf{u}[n]$, a saída é calculada usando o filtro ótimo calculado no instante anterior, $\mathbf{W}[n]$. A saída $y[n]$ é calculada, então, antes que $\mathbf{u}[n]$ seja usada para atualizar o vetor de coeficientes $\mathbf{W}[n]$ no instante atual.

Com relação à inversão direta, o algoritmo RLS possui como vantagens:

- i) o melhor comportamento numérico;
- ii) o algoritmo RLS produz um vetor de coeficientes ótimo a cada instante, não apenas ao fim da seqüência de dados;
- iii) menor ordem de carga computacional, $O(N^2)$, frente a $O(N^3)$ do método direto.

3.5.1 Algoritmo RLS com peso de dados exponencial

Como o método de mínimos quadrados utiliza informação desde o momento que o algoritmo inicia, é feita uma modificação na função de custo de forma a possibilitar que, se desejado, o algoritmo "esqueça" as amostras mais antigas, possibilitando sua operação em um ambiente não-estacionário. A esta modificação do algoritmo é dado o nome de algoritmo RLS com peso de dados exponencial (janelamento exponencial).

A função de custo é definida de modo levemente diferente do definido no método dos mínimos quadrados, com a inclusão de um "fator de esquecimento" $\beta[n,i]$.

$$J_{ls}[n] = \sum_{i=1}^N \beta[n,i] |e[i]|^2 \quad (3.58)$$

onde $0 < \beta[n,i] \leq 1$, para $i=1,2,\dots,n$. O fator exponencial é definido por

$$\beta[n,i] = \lambda^{n-i} \quad (3.59)$$

onde λ é uma constante pouco menor que 1. Para $\lambda=1$, tem-se o método de mínimos quadrados convencional. Caso contrário, refere-se o método como mínimos quadrados com peso exponencial. Este peso pode ser pensado como uma janela exponencial no tempo.

O valor ótimo dos coeficientes é definido pelas equações normais

$$\Phi[n] \mathbf{W}^*[n] = \Theta[n] \quad (3.60)$$

onde a matriz de correlação é definida por

$$\Phi[n] = \sum_{i=1}^n \lambda^{n-i} \mathbf{u}[i] \mathbf{u}^H[i] \quad (3.61)$$

e o vetor de correlação cruzada é definido por

$$\Theta[n] = \sum_{i=1}^n \lambda^{n-i} \mathbf{u}[i] d^*[i] \quad (3.62)$$

Em contraste ao capítulo 2, é usado o método do pré-janelamento [18]. Pode-se escrever as equações recursivas para a matriz de correlação e o vetor de correlação cruzada como

$$\Phi[n] = \lambda \Phi[n-1] + \mathbf{u}[n] \mathbf{u}^H[n] \quad (3.63)$$

$$\Theta[n] = \lambda \Theta[n-1] + \mathbf{u}[n] d^*[n] \quad (3.64)$$

Segundo Treichler [43], o peso exponencial possui uma constante de tempo

$$\tau = \frac{1}{1-\lambda} \quad (3.65)$$

Se $\lambda=1$, o intervalo de tempo é infinito, o que significa que todas as amostras possuem o mesmo peso. As equações necessitam de pequenas modificações para a incorporação deste fator [43] [18]:

$$\mathbf{W}^*[n+1] = \mathbf{W}^*[n] + \frac{(d[n] - y[n])^* \mathbf{Z}[n]}{\lambda + q} \quad (3.66)$$

$$\Phi^{-1}[n+1] = \frac{\Phi^{-1}[n]}{\lambda} - \frac{\mathbf{Z}[n]\mathbf{Z}^H[n]}{\lambda(\lambda + q)} \quad (3.67)$$

O esquema de adaptação dos coeficientes neste algoritmo pode ser visto na figura 3.10.

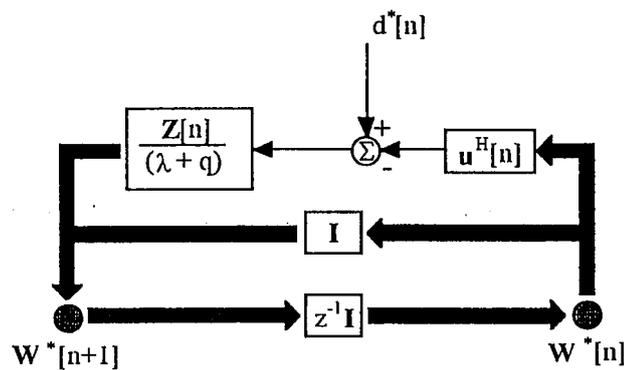


FIGURA 3.10: Diagrama de fluxo de sinais no algoritmo RLS com peso exponencial

TABELA 3.2: Algoritmo *Recursive Least Squares* com peso de dados exponencial

INICIALIZAÇÃO:

- 1) $\Phi^{-1}[0] = \delta^{-1} \mathbf{I}$; $\mathbf{W}[0] = \mathbf{0}$; $\mathbf{u}[0] = \mathbf{0}$

LOOP:

- 2) $\mathbf{u}[n]$ é formado com $u[n]$
- 3) $y[n] = \mathbf{u}^T[n]\mathbf{W}[n]$
- 4) $e[n] = d[n] - y[n]$
- 5) $\mathbf{Z}[n] = \Phi^{-1}[n]\mathbf{u}[n]$
- 6) $q = \mathbf{u}^H[n]\mathbf{Z}[n]$
- 7) $v = \frac{1}{\lambda + q}$
- 8) $\tilde{\mathbf{Z}}[n] = v\mathbf{Z}[n]$
- 9) $\mathbf{W}^*[n+1] = \mathbf{W}^*[n] + e^*[n]\tilde{\mathbf{Z}}[n]$
- 10) $\Phi^{-1}[n+1] = \frac{\Phi^{-1}[n]}{\lambda} - \frac{\mathbf{Z}[n]\mathbf{Z}^H[n]}{\lambda(\lambda + q)}$

Na inicialização, δ deve ser pequeno em comparação com $0.01\sigma_{u_x}^2$, fração da variância de $u[n]$.

Para uma análise detalhada da convergência do algoritmo RLS, veja Haykin [18]. O algoritmo RLS não produz erro em excesso para operação em ambiente estacionário, com $\lambda=1$ e para $n \rightarrow \infty$. Isso implica que, teoricamente, não há desajustamento neste caso. A taxa de convergência é tipicamente uma ordem de magnitude maior que o algoritmo LMS. Esta rapidez faz com que esta classe de algoritmos seja empregada em situações que necessitam de rápida convergência, como equalizadores de *modem* de início rápido (*fast startup*) e *encoders* ADPCM.

3.5.2 Operação em ambiente não-estacionário

Segundo Haykin [18], o uso do coeficiente exponencial menor que 1, $\lambda < 1$, muda drasticamente o comportamento do algoritmo RLS, passando a existir desajustamento. O desajustamento produzido pelo ruído no vetor de coeficientes, E_1 , quando λ está bem próximo de 1, pode ser aproximado por

$$M \approx \frac{1-\lambda}{1+\lambda} N \quad (3.68)$$

De modo geral, uma rápida adaptação leva a um processo mais ruidoso.

O vetor de erro de atraso, E_2 , possui uma constante de tempo para cada termo dada por

$$\tau_i = \frac{1}{1-\lambda} \quad (3.69)$$

Esse resultado mostra que a constante de tempo é independente do espalhamento dos autovalores, ao contrário do LMS.

A performance do algoritmo em um ambiente não-estacionário deve ser estudada levando-se em conta que a capacidade de rastreamento é influenciada pela taxa de convergência e pela flutuação da performance em regime permanente. Segundo Haykin [18], quando ambos algoritmos operam sob mesmas condições (minimizar o desajustamento M através da escolha adequada das constantes μ e λ), o algoritmo LMS tem uma performance de rastreamento superior ao RLS. Para uma comparação do algoritmo RLS com filtros de Kalman, veja Haykin [18].

3.5.3 Algoritmos *fast* RLS

Esforços no sentido de diminuir a carga computacional dos algoritmos de mínimos quadrados geraram a família de algoritmos rápidos. Um algoritmo é considerado *fast* se sua complexidade computacional aumenta linearmente com o número de coeficientes adaptativos e, portanto, de modo similar ao algoritmo LMS. De modo geral, tem-se a situação expressa pela tabela 3.3 para o grau de complexidade de algoritmos originários do método de mínimos quadrados. No entanto, a classe *fast* ainda é mais custosa computacionalmente do que algoritmos de busca de gradiente similares.

TABELA 3.3: Comparação da ordem da carga computacional de algoritmos de mínimos quadrados

inversão direta	RLS normal	<i>fast</i> RLS
$O(N^3)$	$O(N^2)$	$O(N)$

Haykin [18] cita como algoritmos RLS rápidos para filtros transversais os algoritmos *fast* RLS (Kalman), *fast transversal filters* e FAEST. Uma visão mais aprofundada desta classe de algoritmos pode ser vista em diversos trabalhos [5] [4] [8]. Em geral, estas implementações de algoritmos procuram encontrar e se utilizar de propriedades inexploradas pelo algoritmo RLS normal (como a propriedade de deslocamento do vetor de sinal de entrada, *shifting property*, que é modificado pela adição de apenas uma amostra de iteração a iteração). O emprego de definições inteligentes para reduzir a carga computacional de versões anteriores do algoritmo também é evidenciado [4].

3.6 Algoritmos para filtros *lattice*

O uso da estrutura *lattice* em processamento adaptativo de sinais é muito popular, sendo que é possível a distinção de duas classes de algoritmos [9]: *gradient adaptive lattice* (GAL) ou *lattice* LMS e *least squares lattice* (LSL). Em comparação aos algoritmos implementados em estrutura direta, esta estrutura apresenta como vantagens a maior velocidade de convergência, relativa imunidade ao condicionamento dos dados (dispersão dos autovalores da matriz de correlação \mathbf{R}), melhores propriedades numéricas em implementações de precisão finita, entre outras [18] [29] [38]. Esta estrutura tem sido empregada em problemas de equalização rápida de canais de comunicação [29] [38]. Griffiths [17] apresenta estruturas *lattice* de múltiplos canais que podem ser usadas em cancelamento adaptativo de ruído com várias entradas de referência.

3.6.1 Derivação do algoritmo *lattice* LMS ou GAL para o caso real

O objetivo desta seção é prover uma derivação simplificada do algoritmo para o caso real [38] [1]. O algoritmo se baseia no ajuste dos coeficientes da parte *lattice* e da parte *ladder* com duas funções de custo distintas. Na verdade, são dois algoritmos adaptativos trabalhando na mesma estrutura. Para este caso, os coeficientes de reflexão do *lattice* são iguais, dados por $k_m^l[n]$. Griffiths [16] sugere dois algoritmos para o ajuste dos coeficientes de um filtro de predição *lattice*, baseados na minimização da média do quadrado do erro, definido de modo análogo ao feito no desenvolvimento do algoritmo LMS. Um dos métodos considera que os coeficientes k do *lattice* para frente e para trás são diferentes entre si, e o outro método considera que os mesmos são iguais ($k^f=k^b=k^l$) (consideração seguida aqui).

De acordo com o método de *steepest descent*, tem-se que:

$$k_m^l[n+1] = k_m^l[n] + \mu(-\nabla(J[n])) \quad (3.70)$$

O ajuste dos coeficientes k^l da parte *lattice* do filtro é feito de modo a minimizar a função de custo na seção (estágio) i , definida como a soma das variâncias dos erros de predição para a frente e para trás na saída do estágio, conforme esquematizado na figura 3.11:

$$J[n] = \varepsilon\left\{\left(e_{m+1}^f(n)\right)^2\right\} + \varepsilon\left\{\left(e_{m+1}^b(n)\right)^2\right\} \quad (3.71)$$

Onde os erros são definidos por

$$e_{m+1}^f[n] = e_m^f[n] - k_m^l[n]e_m^b[n-1] \quad (3.72)$$

$$e_{m+1}^b[n] = e_m^b[n-1] - k_m^l[n]e_m^f[n] \quad (3.73)$$

$$e_1^f[n] = e_1^b[n] = u[n] \quad (3.74)$$

A minimização desta função é feita em relação aos coeficientes de reflexão do estágio

$$\frac{\partial J[n]}{\partial k_m^l[n]} = \varepsilon\left\{2e_{m+1}^f[n]\left(-e_m^b[n-1]\right)\right\} + \varepsilon\left\{2e_{m+1}^b[n]\left(-e_m^f[n]\right)\right\} \quad (3.75)$$

$$\frac{\partial J[n]}{\partial k_m^l[n]} = -2 \left(\varepsilon \{ e_{m+1}^f[n] e_m^b[n-1] \} + \varepsilon \{ e_{m+1}^b[n] e_m^f[n] \} \right) \quad (3.76)$$

A substituição em (3.70) permite escrever:

$$k_m^l[n+1] = k_m^l[n] + 2\mu \left(\varepsilon \{ e_{m+1}^f[n] e_m^b[n-1] \} + \varepsilon \{ e_{m+1}^b[n] e_m^f[n] \} \right) \quad (3.77)$$

O algoritmo LMS toma a estimativa mais simples para o gradiente, ou seja, é feita a estimativa instantânea das esperanças estatísticas (ignora-se o operador $\varepsilon\{\}$):

$$k_m^l[n+1] = k_m^l[n] + 2\mu \left(e_{m+1}^f[n] e_m^b[n-1] + e_{m+1}^b[n] e_m^f[n] \right) \quad (3.78)$$

Esta forma do algoritmo deve ser modificada, segundo Griffiths [17], para que seja possível seguir as variações de potência dos sinais em cada estágio. Esta modificação é feita por um parâmetro calculado recursivamente com base nos sinais que entram no estágio.

Segundo Ahmed [1], substitui-se o termo 2μ por

$$2\mu \rightarrow \frac{\alpha}{\delta_m[n]} \quad (3.79)$$

onde

$$\delta_m[n] = \beta \delta_m[n-1] + \alpha \left((e_m^b[n-1])^2 + (e_m^f[n])^2 \right) \quad (3.80)$$

e onde $\alpha=1-\beta$. Segundo Satorius [38] e Makhoul [29], $\alpha=1$ e $\beta \neq 1-\alpha$.

A aplicação do método de *steepest descent* para os coeficientes da parte *ladder* leva a

$$k_m[n+1] = k_m[n] + \mu \left(-\nabla(J[n]) \right) \quad (3.81)$$

Onde a função de custo é definida como sendo a variância (média do quadrado) do erro do preditor *lattice* de m-ésima ordem,

$$J[n] = \varepsilon \left\{ (e_m[n])^2 \right\} \quad (3.82)$$

Os erros da parte *ladder* são definidos como

$$e_m[n] = e_{m-1}[n] - k_m[n]e_m^b[n] \quad (3.83)$$

$$e_0[n] = d[n] \quad (3.84)$$

A minimização em relação ao m-ésimo coeficiente *ladder* resulta em:

$$\frac{\partial J[n]}{\partial k_m[n]} = 2\varepsilon\{e_m[n](-e_m^b[n])\} \quad (3.85)$$

De modo análogo, a substituição em 3.81 e a simplificação usada pelo algoritmo LMS levam a

$$k_m[n+1] = k_m[n] + 2\mu e_m[n]e_m^b[n] \quad (3.86)$$

A mesma normalização com respeito à potência dos sinais é feita, levando à substituição do fator de convergência fixo por um variável, dado por

$$2\mu \rightarrow \frac{\alpha}{\gamma_m[n]} \quad (3.87)$$

onde

$$\gamma_m[n] = \beta\gamma_m[n-1] + \alpha(e_m^b[n])^2 \quad (3.88)$$

e onde $\alpha=1-\beta$. Segundo Satorius [38] e Makhoul [29], $\alpha=1$ e $\beta \neq 1-\alpha$.

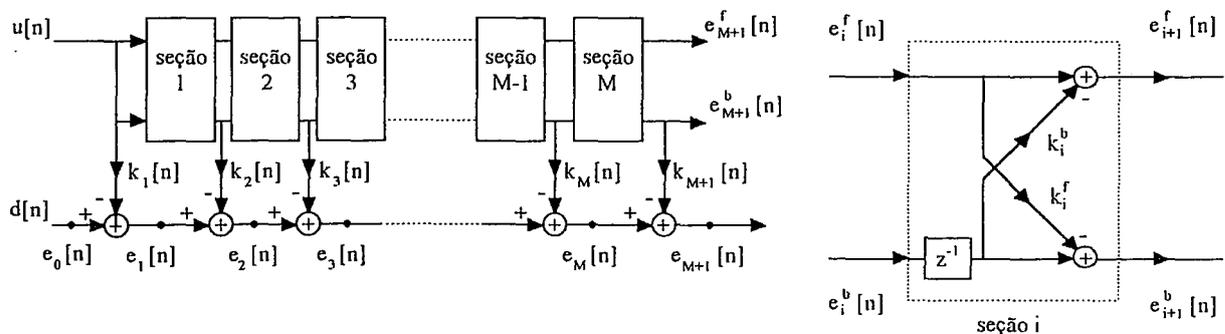


FIGURA 3.11: Estrutura do filtro *lattice* para estimação de processo conjunto

Os sinais correspondentes podem ser vistos na figura 3.11. Esta figura apresenta o preditor *lattice* para processo conjuntamente estacionário e os sinais correspondentes da

estrutura. O número de estágios *lattice* é M , e o número de coeficientes da parte *ladder* é igual a $M+1$. O algoritmo pode ser visto na tabela 3.4.

TABELA 3.4: Algoritmo *Gradient Adaptive Lattice* (GAL) ou *Lattice LMS* - caso real

INICIALIZAÇÃO: $m=1,2,\dots,M+1$

1) $e_m^b[n] = e_m^b[n-1] = k_m[n] = k_m^l[n] = \gamma_m[n] = \delta_m[n] = e_m^f[n] = e_m[n] = 0$

LOOP:

2) $e_1^f[n] = e_1^b[n] = x[n]$; $e_0[n] = d[n]$

PARTE LATTICE: $m=1,2,\dots,M$

3) $e_{m+1}^f[n] = e_m^f[n] - k_m^l[n]e_m^b[n-1]$

4) $e_{m+1}^b[n] = e_m^b[n-1] - k_m^l[n]e_m^f[n]$

5) $\delta_m[n] = \beta\delta_m[n-1] + \alpha\left((e_m^b[n-1])^2 + (e_m^f[n])^2\right)$

6) $k_m^l[n+1] = k_m^l[n] + \frac{\alpha}{\delta_m[n]}(e_{m+1}^f[n]e_m^b[n-1] + e_{m+1}^b[n]e_m^f[n])$

PARTE LADDER: $m=1,2,\dots,M+1$

7) $e_m[n] = e_{m-1}[n] - k_m[n]e_m^b[n]$

8) $\gamma_m[n] = \beta\gamma_m[n-1] + \alpha(e_m^b[n])^2$

9) $k_m[n+1] = k_m[n] + \alpha e_m[n]e_m^b[n] / \gamma_m[n]$

Segundo Satorius [38] e Makhoul [29], o parâmetro α é igual a 1 (um).

3.7 Algoritmo *fast RLS* para filtros *lattice*

Um dos motivos pelo interesse em estruturas *lattice* para implementação de algoritmos RLS está no fato de que estes algoritmos possuem melhores propriedades numéricas que suas implementações em forma direta [26]. Do mesmo modo que no caso do algoritmo GAL, ambos os coeficientes do filtro, *lattice* e *ladder*, são variáveis. No caso de algoritmos RLS, os coeficientes de reflexão são assumidos diferentes entre si, ou seja, $k^f \neq k^b$. O algoritmo aqui descrito é uma forma de algoritmo recursivo LSL usando erros de estimação *a priori* e realimentação de erro [26] [18], sendo sua principal vantagem a robustez numérica. Este algoritmo pertence à classe de algoritmos *fast RLS*.

TABELA 3.5: Algoritmo *Recursive Least Squares Lattice using a priori errors with error feedback*

INICIALIZAÇÃO: $m=1,2,\dots,M+1$

- 1) $\alpha_m^f[n-1] = \alpha_m^b[n-1] = \alpha_m^b[n-2] = \sigma$
- 2) $e_m^b[n-1] = e_m^b[n-2] = e_m^f[n-1] = e_m[n-1] = 0$
- 3) $\alpha_m^*[n-2] = 1$; $\alpha_m^f[n-2] = \sigma$; $k_m[n-1] = k_m^f[n-1] = k_m^b[n-1] = 0$

LOOP:

- 4) $e_1^f[n] = e_1^b[n] = x[n]$; $e_0[n] = d[n]$; $\alpha_1^*[n-1] = 1$
- 5) $\alpha_1^f[n] = \alpha_1^b[n] = \lambda \alpha_1^f[n-1] + (x[n])^2$

PARTE LATTICE: $m=1,2,\dots,M$

- 6) $k_m^f[n] = k_m^f[n-1] + e_m^f[n-1]e_{m+1}^b[n-1]\alpha_m^*[n-2] / \alpha_m^f[n-1]$
- 7) $k_m^b[n] = k_m^b[n-1] + e_m^b[n-2]e_{m+1}^f[n-1]\alpha_m^*[n-2] / \alpha_m^b[n-2]$
- 8) $e_{m+1}^f[n] = e_m^f[n] - k_m^b[n]e_m^b[n-1]$
- 9) $e_{m+1}^b[n] = e_m^b[n-1] - k_m^f[n]e_m^f[n]$
- 10) $\alpha_{m+1}^f[n-1] = \lambda \alpha_{m+1}^f[n-2] + (e_{m+1}^f[n-1])^2 \alpha_{m+1}^*[n-2]$
- 11) $\alpha_{m+1}^*[n-1] = \alpha_m^*[n-1] - (e_m^b[n-1])^2 (\alpha_m^*[n-1])^2 / \alpha_m^b[n-1]$
- 12) $\alpha_{m+1}^b[n-1] = \lambda \alpha_{m+1}^b[n-2] + (e_{m+1}^b[n-1])^2 \alpha_{m+1}^*[n-1]$

PARTE LADDER: $m=1,2,\dots,M+1$

- 13) $k_m[n] = k_m[n-1] + e_m^b[n-1]e_m[n-1]\alpha_m^*[n-1] / \alpha_m^b[n-1]$
- 14) $e_m[n] = e_{m-1}[n] - k_m[n]e_m^b[n]$

Os termos de erro *a priori* e *a posteriori* são definidos como o erro calculado usando a estimativa para os coeficientes obtida em um instante anterior, $(n-1)$, e o erro calculado com a estimativa dos coeficientes obtida no instante atual. Os termos "erro de predição para frente" e "erro de predição para trás" referem-se ao equivalente filtro FIR de erro de predição. Os erros de predição para frente e para trás são os sinais que se propagam na estrutura *lattice*. A parte *lattice* se refere à estrutura do filtro de predição e a parte *ladder* se refere ao estimador do processo conjunto (filtro FIR de estrutura direta associado).

Outras formas de algoritmos possuem necessidades de cálculos maiores (como o algoritmo QRD LSL - *QR Decomposition based Least Squares Lattice algorithm* - que requer o cálculo de senos e cossenos) [18]. Neste trabalho o interesse é maior na comparação das principais famílias de algoritmos e suas características. Para uma comparação de performance das estruturas *lattice* GAL e LSL veja Friedlander [14]. O resumo do algoritmo LSL,

conforme Ling [26], pode ser visto na tabela 3.5, seguindo a mesma notação que o algoritmo GAL e figura 3.11.

3.8 Algoritmos para filtros adaptativos IIR

O uso de filtros adaptativos que modelam sistemas IIR é interessante devido à possibilidade de redução na carga computacional frente a filtros FIR equivalentes. Para a aproximação de uma resposta ao impulso de maior duração, filtros FIR tendem a ser longos (com um grande número de coeficientes). O processo de adaptação de coeficientes tende a elevar a carga computacional de acordo. Filtros IIR, em contrapartida, por sintetizarem o modelo de sistema IIR diretamente, podem realizar a mesma resposta ao impulso com um número muito menor de coeficientes. Os problemas principais são o controle da estabilidade do filtro, controle da estabilidade do algoritmo e a forma da função de custo, que pode apresentar mínimos locais. Outro problema é a velocidade de convergência dos algoritmos.

Esses problemas inibiram o desenvolvimento dos primeiros algoritmos de filtros adaptativos IIR, sendo o primeiro trabalho publicado em 1975, segundo Treichler [43]. Diversas estruturas existem para a implementação de filtros IIR, como direta, *lattice* e outros [9]. Neste trabalho são analisados filtros adaptativos IIR de forma direta e alguns algoritmos. Algoritmos *lattice* IIR podem ser vistos em diversos trabalhos [9] [36] [37].

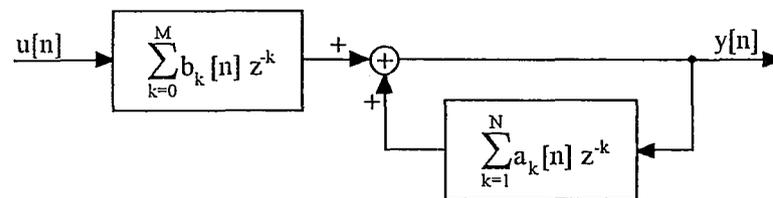


FIGURA 3.12: Filtro IIR de estrutura direta

3.8.1 Algoritmo IIR LMS

Pode ser desenvolvido um algoritmo LMS para filtros IIR [49]. No caso da figura 3.12, tem-se as seguintes definições (sinais reais):

$$y[n] = \sum_{k=0}^M b_k[n] u[n-k] + \sum_{k=1}^N a_k[n] y[n-k] \quad (3.89)$$

$$W[n] = [b_0[n] \quad \dots \quad b_M[n] \quad a_1[n] \quad \dots \quad a_N[n]]^T \quad (3.90)$$

$$\mathbf{U}[n] = [u[n] \ \dots \ u[n-M] \ y[n-1] \ \dots \ y[n-N]]^T \quad (3.91)$$

$$e[n] = d[n] - y[n] = d[n] - \mathbf{W}^T[n]\mathbf{U}[n] \quad (3.92)$$

A aplicação do método de *steepest descent* leva à equação de adaptação dos coeficientes

$$\mathbf{W}[n+1] = \mathbf{W}[n] - \mathbf{M}\nabla(J[n]) \quad (3.93)$$

onde

$$\mathbf{M} = \text{diag}[\mu_0[n] \ \dots \ \mu_M[n] \ \nu_1[n] \ \dots \ \nu_N[n]] \quad (3.94)$$

e $J[n]$ é a função "média do quadrado do erro". A utilização do algoritmo LMS implica em usar a estimativa simplificada (instantânea), em vez da média, para o cálculo do gradiente.

$$\hat{\nabla}(J[n]) = \frac{\partial e^2[n]}{\partial \mathbf{W}} = 2e[n] \frac{\partial e[n]}{\partial \mathbf{W}} \quad (3.95)$$

$$\hat{\nabla}(J[n]) = -2e[n] \left[\frac{\partial y[n]}{\partial b_0[n]} \ \dots \ \frac{\partial y[n]}{\partial b_M[n]} \ \frac{\partial y[n]}{\partial a_1[n]} \ \dots \ \frac{\partial y[n]}{\partial a_1[n]} \right]^T \quad (3.96)$$

Devido ao caráter recursivo do filtro e também devido ao fato de os coeficientes serem dependentes do tempo, a saída $y[n]$ no instante de tempo n reflete a história da variação temporal dos coeficientes. Por essas razões, a avaliação das derivadas em 3.96 é uma tarefa extremamente difícil [19]. No entanto, se é desprezada a dependência temporal dos coeficientes, isto é, se os coeficientes são tratados como constantes, as derivadas podem ser escritas como

$$\alpha_i[n] = \frac{\partial y[n]}{\partial a_i} = 0 + \frac{\partial}{\partial a_i} \left(\sum_{k=1}^N a_k y[n-k] \right) = \sum_{k=1}^N \left(y[n-k] \frac{\partial a_k}{\partial a_i} + a_k \frac{\partial y[n-k]}{\partial a_i} \right) \quad (3.97)$$

$$\alpha_i[n] = y[n-i] + \sum_{k=1}^N a_k \frac{\partial y[n-k]}{\partial a_i} = y[n-i] + \sum_{k=1}^N a_k \alpha_i[n-k] \quad (3.98)$$

$$\beta_i[n] = \frac{\partial y[n]}{\partial b_i} = x[n-i] + \sum_{k=1}^N a_k \frac{\partial y[n-k]}{\partial a_i} = x[n-i] + \sum_{k=1}^N a_k \beta_i[n-k] \quad (3.99)$$

O que permite calcular os termos recursivamente. O gradiente fica, então

$$\hat{\nabla}(J[n]) = -2e[n][\beta_0[n] \ \dots \ \beta_M[n] \ \alpha_1[n] \ \dots \ \alpha_N[n]]^T \quad (3.100)$$

Note-se que este resultado não é válido para o caso de coeficientes variáveis. No algoritmo IIR LMS, entretanto, são desprezados os efeitos de se considerar os coeficientes do filtro fixos, ao invés de variáveis, na estimativa do gradiente de $J[n]$.

TABELA 3.6: Algoritmo IIR LMS

<i>INICIALIZAÇÃO</i>	
1)	$W=0, \alpha_i=0, \beta_i=0$
<i>LOOP</i>	
2)	$U[n]$ é formado com $u[n]$ e $y[n-1]$
3)	$y[n] = W^T[n]U[n]$
<i>PARA $i=1$ ATÉ N</i>	
4)	$\alpha_i[n] = y[n-i] + \sum_{k=1}^N a_k \alpha_i[n-k]$
<i>PARA $i=0$ ATÉ M</i>	
5)	$\beta_i[n] = u[n-i] + \sum_{k=1}^N a_k \beta_i[n-k]$
6)	$\hat{\nabla}(J[n]) = -2e[n][\beta_0[n] \ \dots \ \beta_M[n] \ \alpha_1[n] \ \dots \ \alpha_N[n]]^T$
7)	$W[n+1] = W[n] - M\hat{\nabla}(J[n])$

3.8.1 Algoritmo *Recursive* LMS

Se as derivadas de $y[n]$ anteriores forem truncadas no primeiro termo

$$\alpha_i[n] = y[n-i] \quad (3.101)$$

$$\beta_i[n] = u[n-i] \quad (3.102)$$

o gradiente de $J[n]$ é simplificado para

$$\hat{\nabla}(J[n]) = -2e[n][\beta_0[n] \ \dots \ \beta_M[n] \ \alpha_1[n] \ \dots \ \alpha_N[n]]^T \quad (3.103)$$

$$\hat{\nabla}(J[n]) = -2e[n][u[n] \ \dots \ u[n-M] \ y[n-1] \ \dots \ y[n-N]]^T \quad (3.104)$$

$$\hat{\nabla}(J[n]) = -2e[n]\mathbf{U}^T[n] \quad (3.105)$$

O algoritmo que usa esta estimativa para o gradiente é conhecido como algoritmo *Recursive LMS* [13], e seu comportamento é discutido em diversas comunicações [48] [22]. Um de seus problemas é não minimizar efetivamente $J[n]$, a média do erro quadrado. No entanto, é um algoritmo simples do ponto de vista computacional e possui aplicação, segundo Widrow [48] nos campos de detecção de sinal, estimativa instantânea da frequência e análise espectral.

TABELA 3.7: Algoritmo *Recursive LMS* - caso real

<i>INICIALIZAÇÃO</i>
1) $\mathbf{W}=0, \alpha_i=0, \beta_i=0$
<i>LOOP</i>
2) $\mathbf{U}[n]$ é formado com $u[n]$ e $y[n-1]$
3) $y[n] = \mathbf{W}^T[n]\mathbf{U}[n]$
4) $\hat{\nabla}(J[n]) = -2e[n]\mathbf{U}^T[n]$
5) $\mathbf{W}[n+1] = \mathbf{W}[n] - M\hat{\nabla}(J[n])$

3.9 Algoritmo para filtro adaptativo IIR baseado em hiper-estabilidade

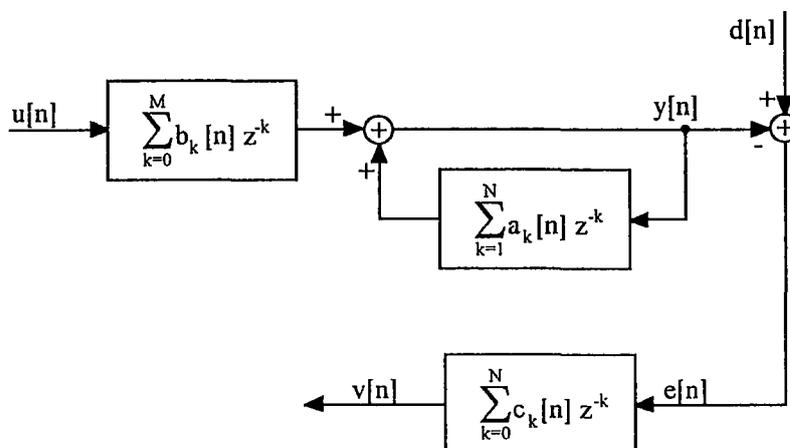


FIGURA 3.13: Filtro IIR com algoritmo SHARF

Trabalhos na área de identificação recursiva de sistemas e o estudo de conceitos da teoria de estabilidade não-linear, derivados da área de Controle, permitiram o desenvolvimento do algoritmo HARF (*Hyperstable Adaptive Recursive Filter*) [24] [21]. O algoritmo HARF é a primeira técnica proposta para filtragem adaptativa IIR com propriedades de convergência prováveis. Embora tenha uma complexidade computacional moderada, esta pode ser reduzida pelo uso de simplificações que preservam as características mais desejáveis do algoritmo original, entretanto reduzindo a taxa de convergência. Estas simplificações levam ao algoritmo *Simplified HARF* ou SHARF.

Um dos problemas dessa classe de algoritmos é a escolha dos coeficientes de *error smoothing*, c_i , conforme a figura 3.13. Esta escolha de coeficientes afeta a convergência e o comportamento do algoritmo. Note-se que este algoritmo utiliza os coeficientes c_i para prover uma versão filtrada do erro [49]. Em geral [21], o algoritmo não segue uma estratégia descendente pelo gradiente no espaço de erro dos parâmetros. A derivação do algoritmos e um estudo de suas propriedades de convergência podem ser vistas nos trabalhos de Treichler e Fan [24] [21] [9].

TABELA 3.8: Algoritmo SHARF

INICIALIZAÇÃO

1) $\mathbf{W}[n]=\mathbf{0}$, $v=0$, $c_0 \stackrel{\Delta}{=} 1$

LOOP

2) $\mathbf{U}[n]$ é formado com $u[n]$ e $y[n-1]$

3) $y[n] = \mathbf{W}^T[n]\mathbf{U}[n]$

4) $v[n] = \sum_{k=0}^P c_k e[n-k]$

5) $\mathbf{W}[n+1] = \mathbf{W}[n] + v[n]\mathbf{M}\mathbf{U}[n]$

3.10 Problemas de filtros IIR

Apesar de potencialmente menos custosos computacionalmente na síntese de funções de transferência de maior duração do que similares FIR, filtros adaptativos IIR possuem diversos problemas, sendo os principais a estabilidade do filtro e do algoritmo, a presença de mínimos locais na função de custo (superfície JxW) e a velocidade de convergência dos algoritmos.

Sistemas com resposta ao impulso infinita não adaptativos podem se tornar instáveis se alguns de seus parâmetros a_k ou b_k tiver valores inadequados (não obedecem o critério de estabilidade BIBO). Isso leva à necessidade de teste da estabilidade do filtro em si e de seu algoritmo após o cálculo dos novos coeficientes.

Outro problema destas estruturas é o fato de que a superfície de performance pode apresentar mínimos locais [40]. O problema de superfície de erro (função de custo) multimodal é sério no sentido que algoritmos de busca de gradiente não funcionam corretamente neste caso [49]. No entanto, há indicação [40] que para casos de suficiência do número de parâmetros do filtro (onde a ordem do filtro adaptativo é maior ou igual à necessária), a superfície de performance JxW possui um único mínimo global. Segundo Stearns [40], os vetores gradiente da superfície de erro de filtros adaptativos suficientes podem variar muito em magnitude nestas superfícies, e podem apontar para bem longe do mínimo (global ou local). Esta característica é outro problema para algoritmos de busca de gradiente. Um método para garantir que um filtro adaptativo IIR tenha uma função de performance quadrática é usar pólos fixos, descobertos *a priori*, e adaptar apenas os zeros do filtro [50].

A baixa taxa de convergência de algoritmos de filtros adaptativos IIR é analisado por FAN [9], no contexto de identificação de sistemas. Verifica-se que filtros adaptativos IIR

convergem de modo extremamente lento ou até mesmo não convergem para pólos pouco amortecidos e de baixa frequência (*low damped low frequency poles*). Os algoritmos vistos aqui estão sujeitos a este problema, ou seja, são ideais para identificar pólos que estejam bem dentro do círculo de raio unitário.

3.11 Comparação dos algoritmos através de um exemplo

Com o objetivo de comparar algumas características básicas de desempenho dos algoritmos vistos foi realizada uma simulação, contando com um sinal de voz digitalizado a uma taxa de 10kHz e ruído branco. A figura 3.14 ilustra a geração dos sinais utilizados e o primeiro segundo dos arquivos. Estes sinais serão utilizados na configuração de cancelador adaptativo de interferências. Conforme visto no capítulo 2, esta configuração melhora a relação sinal-ruído do sinal presente na sua entrada primária através da minimização da potência do sinal de saída do sistema (sinal de erro). O processo de minimização é desempenhado pelo filtro adaptativo.

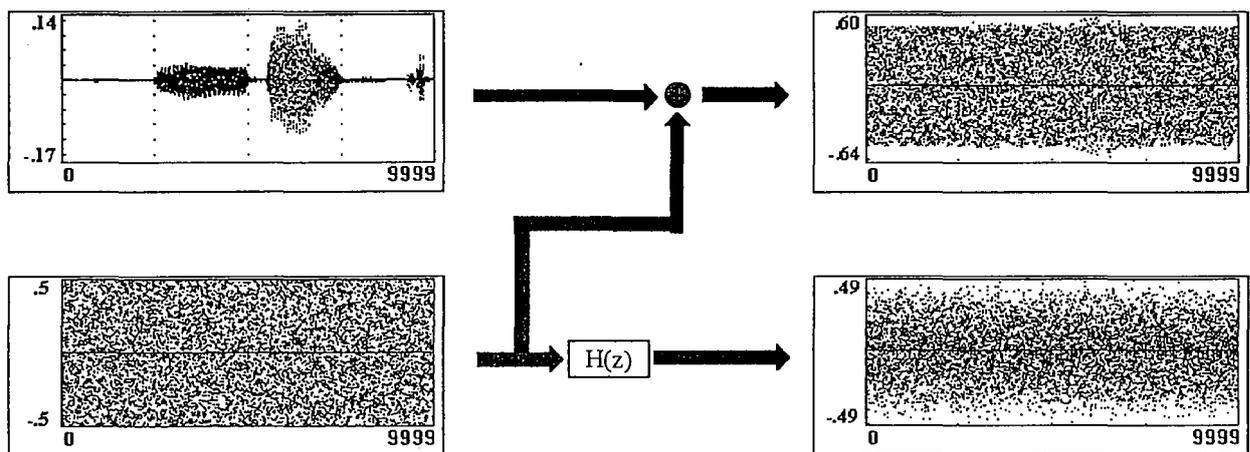


FIGURA 3.14: Trechos dos arquivos de sinais utilizados para a simulação e sua formação

A função $H(z)$ utilizada na geração do sinal de referência do cancelador é um filtro passa-altas de 2ª ordem, de ganho máximo unitário, com dois zeros em $0.7 \pm \angle 20^\circ$. Conforme visto no capítulo 2, o sistema de cancelamento atinge seu melhor desempenho neste caso quando a função de transferência do filtro adaptativo for igual à função inversa de $H(z)$, $H_{fa}(z)=1/H(z)$.

$$H(z) = 0.3564 \frac{z^2 - 1.3156z + .49}{z^2} \quad (3.106)$$

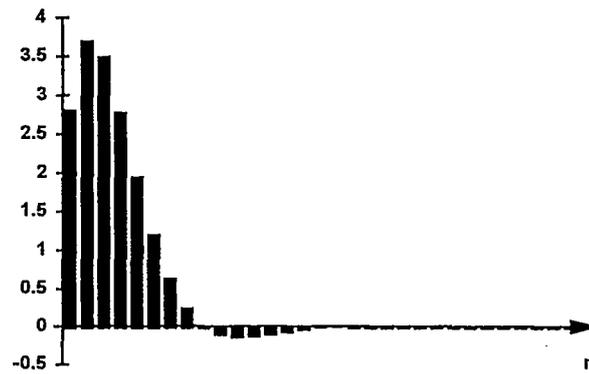


FIGURA 3.15: Resposta ao impulso de $1/H(z)$

A resposta ao impulso da função inversa de $H(z)$ pode ser vista na figura 3.15, de onde se conclui que a parte mais significativa desta resposta possui duração de cerca de 16 amostras. Este valor será usado como o comprimento da linha de atraso dos filtros adaptativos FIR. Para os filtros IIR, serão usados dois pólos e dois zeros.

A seguir são apresentados os resultados das simulações. O sinal primário (entrada para o cancelador) não foi atrasado e as curvas de aprendizagem mostram o resultado de apenas uma simulação.

- 1) algoritmo LMS, com 16 coeficientes e fator de convergência igual a 0.2. Carga computacional por iteração de 32 somas, 32 produtos e 15 deslocamentos.

A figura 3.16 apresenta as curvas de aprendizagem instantâneas para o caso de não haver sinal de voz e para o caso em que existe este sinal. O índice das abscissas refere-se ao número da amostra, digitalizada a uma taxa de 10kHz. Portanto, são mostrados 5 segundos de simulação.

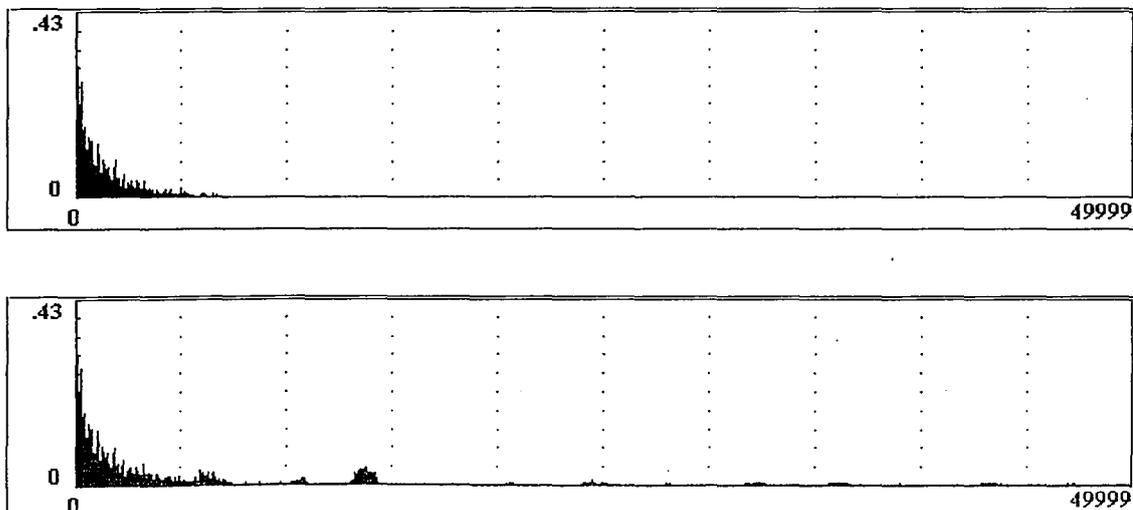


FIGURA 3.16: Curvas de aprendizagem instantâneas para o algoritmo LMS

- 2) algoritmo GAL, com 15 estágios e fator de convergência igual a .999999. Carga computacional por iteração igual a 138 somas, 231 produtos, 31 divisões e 16 deslocamentos.

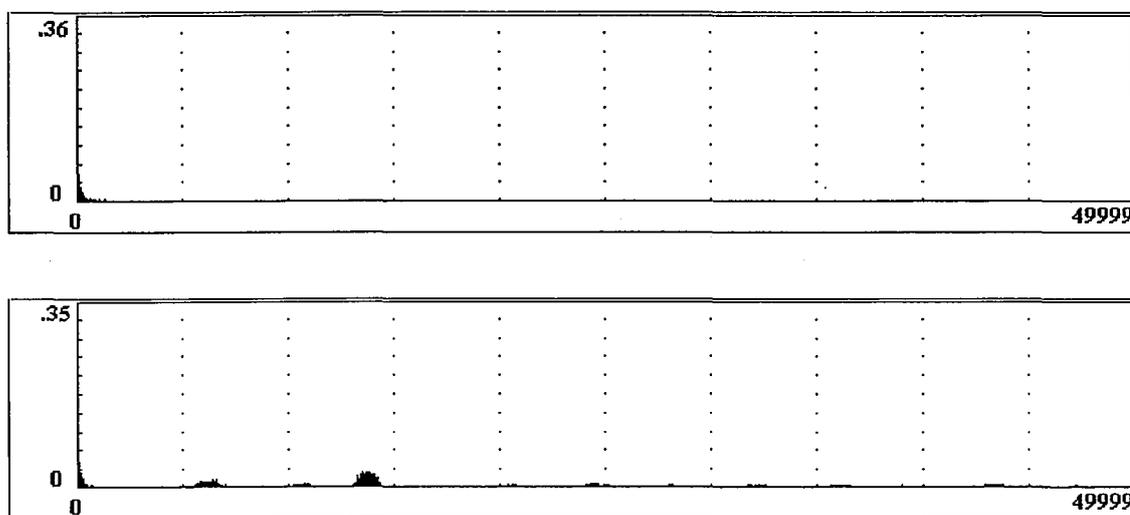


FIGURA 3.17: Curvas de aprendizagem instantâneas para o algoritmo *lattice* LMS ou GAL

- 3) algoritmo RLS, com 16 coeficientes e fator λ igual a .999. Carga computacional por iteração igual a 1632 somas, 1622 produtos, 1 divisão e 239 deslocamentos.

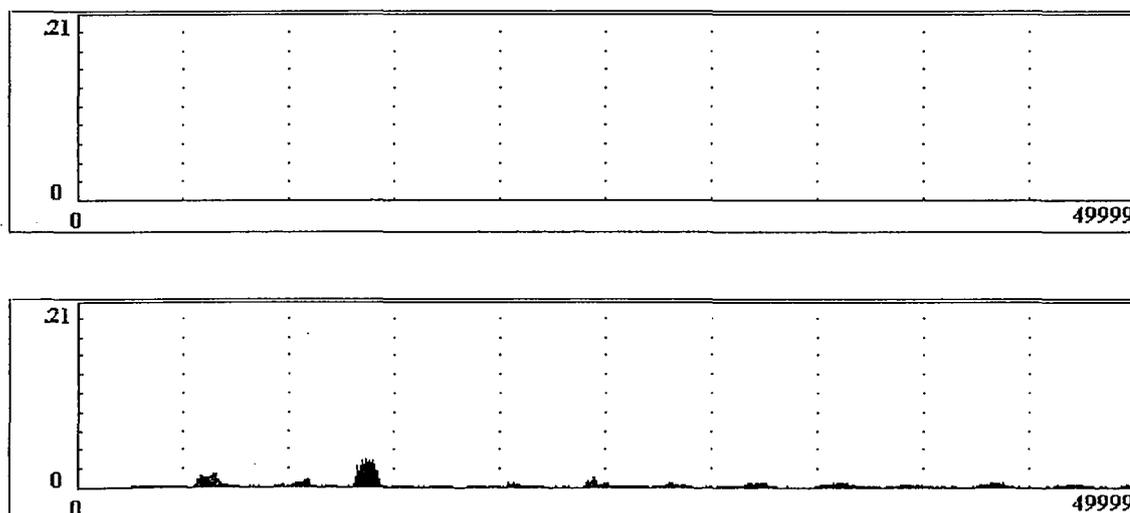


FIGURA 3.18: Curvas de aprendizagem instantâneas para o algoritmo RLS

- 4) algoritmo LSL, com 15 estágios e fator λ igual a .999. Carga computacional por iteração igual a 138 somas, 275 produtos, 61 divisões e 172 deslocamentos.

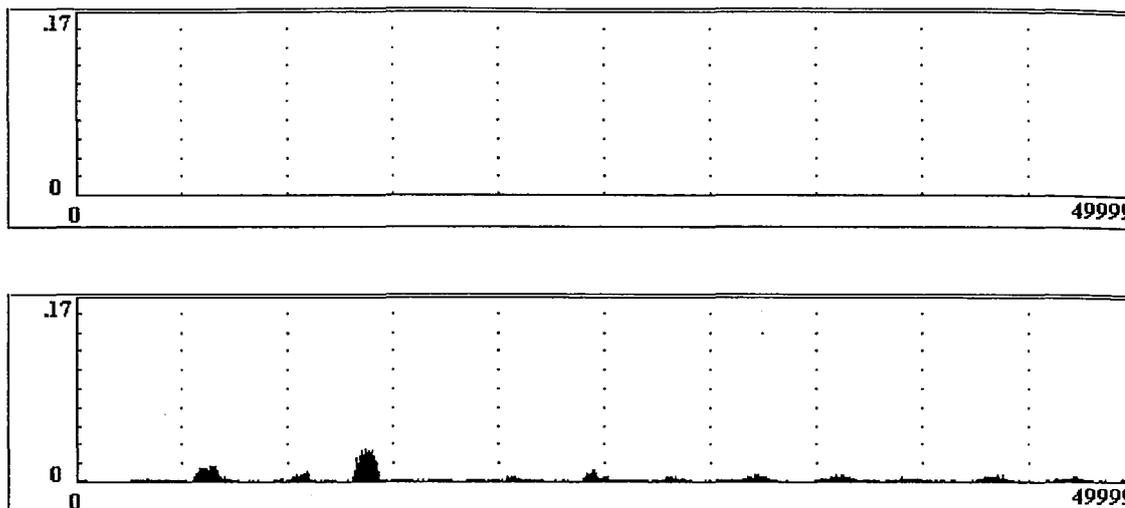


FIGURA 3.19: Curvas de aprendizagem instantâneas para o algoritmo LSL

- 5) algoritmo IIR LMS, com 2 pólos e dois zeros, fator de convergência dos pólos e zeros iguais entre si e iguais a .01. Carga computacional por iteração igual a 25 somas, 20 produtos, 0 divisões e 13 deslocamentos.

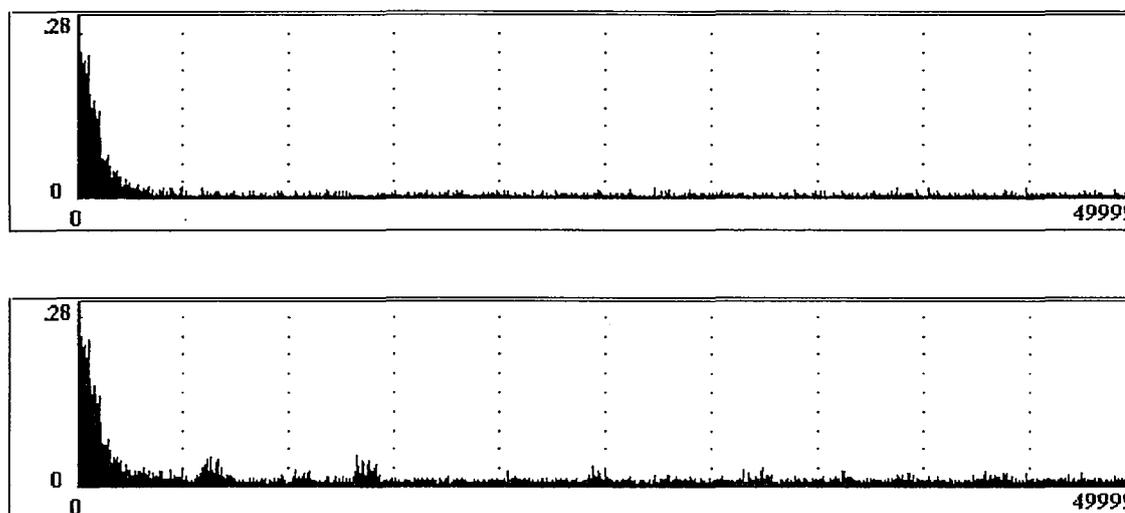


FIGURA 3.20: Curvas de aprendizagem instantâneas para o algoritmo IIR LMS

- 6) algoritmo *Recursive* LMS, com 2 pólos e dois zeros, fator de convergência dos pólos e zeros iguais entre si e iguais a .01. Carga computacional por iteração igual a 10 somas, 6 produtos, 0 divisões e 3 deslocamentos.

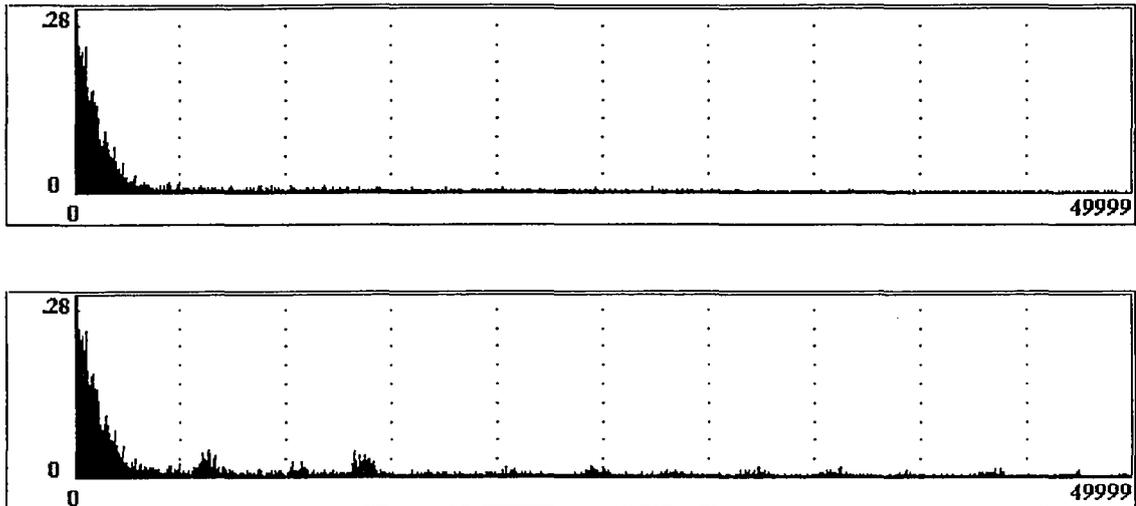


FIGURA 3.21: Curvas de aprendizagem instantâneas para o algoritmo *Recursive LMS*

- 7) algoritmo SHARF, com 2 pólos e dois zeros, fator de convergência dos pólos e zeros iguais entre si e iguais a .01. Fator $c_1 = -1$. Carga computacional por iteração igual a 11 somas, 7 produtos, 0 divisões e 3 deslocamentos

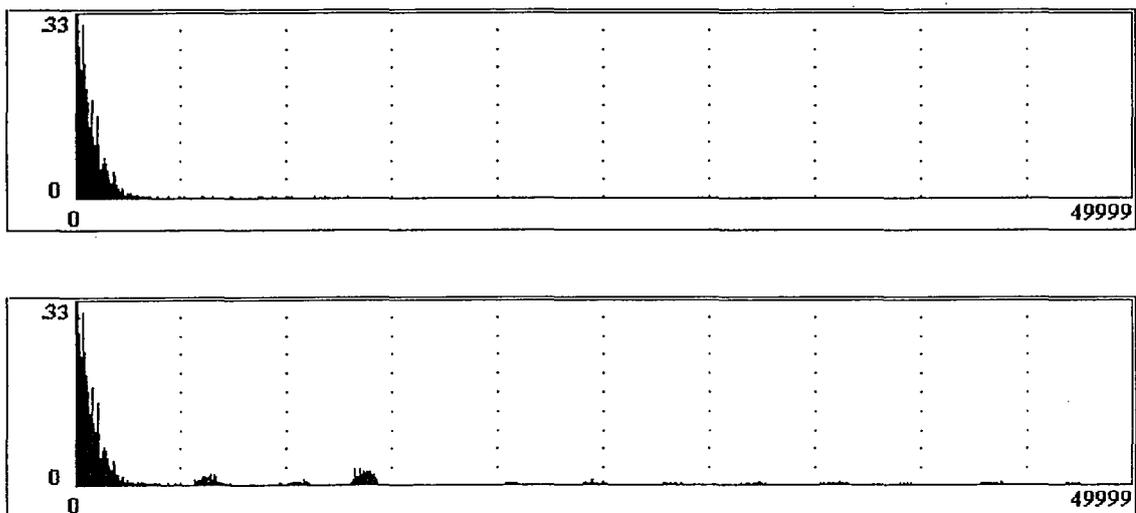


FIGURA 3.22: Curvas de aprendizagem instantâneas para o algoritmo SHARF

O lugar das raízes (2 pólos, x , e 2 zeros, o) durante o processo de adaptação (10 segundos) é mostrado na figura 3.23, com um *zoom* do primeiro segundo mostrado na figura 3.24.

Nota-se neste exemplo o caráter de maior velocidade de convergência dos algoritmos RLS. Também é notado este caráter no uso do algoritmo GAL frente ao LMS. Os algoritmos IIR apresentaram desempenho semelhante ao LMS. O processo de migração dos pólos e zeros mostra sua tendência em direção à solução ideal. De modo geral, este exemplo mostra que é possível aplicar diversos filtros adaptativos ao problema do cancelador adaptativo de ruído. O

desempenho do sistema, em especial o desempenho dinâmico, vai depender das características próprias do filtro usado.

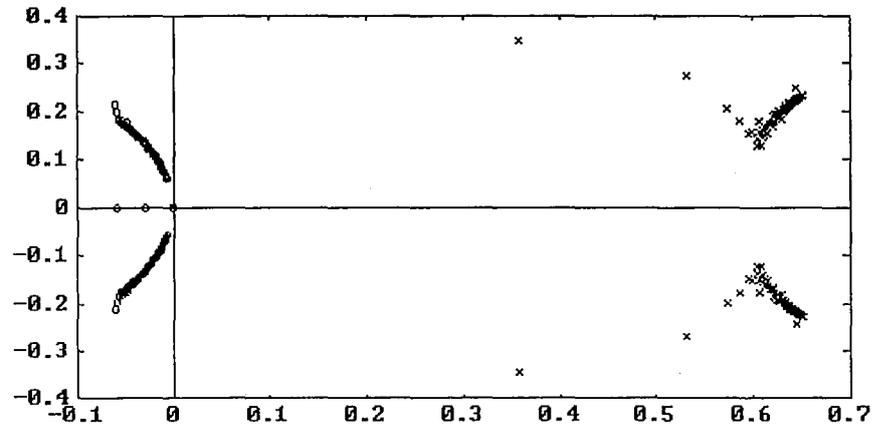


FIGURA 3.23: *Root locus* para o algoritmo Sharf (10 segundos)

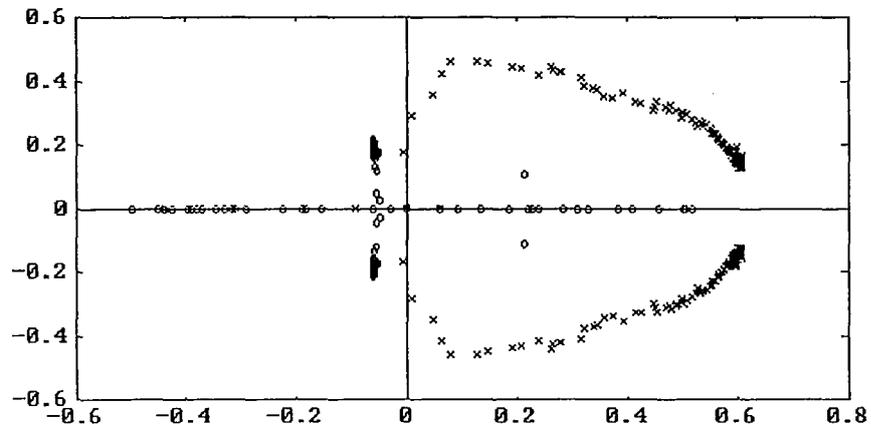


FIGURA 3.24: *Root locus* para algoritmo Sharf - *zoom* do primeiro segundo

4 APLICAÇÃO AO PROBLEMA DE CANCELAMENTO DE ECO

4.1 Introdução

Este capítulo apresenta uma aplicação da técnica de separação de sinais com base em filtros adaptativos. Esta aplicação utiliza a configuração de cancelador adaptativo de ruído para cancelar ecos presentes em telefonia. É feita uma exposição do problema de ecos em telefonia em geral e é apresentado o problema específico enfrentado.

A abordagem inicial tomada é justificada e são apresentados alguns resultados de implementação.

É feita uma avaliação da performance do sistema através da experimentação e da simulação de alguns problemas intrínsecos ao sistema usado. Também são apresentados testes subjetivos de cancelamento.

Outros tipos de estrutura para o filtro adaptativo são analisadas.

4.2 O problema de eco em telefonia

4.2.1 Noções de telefonia

O funcionamento básico de um telefone consiste no circuito mostrado na figura 4.1 [49]. O microfone de carvão necessita de tensão contínua para sua operação, fornecida pela central telefônica. O sinal de voz modula a resistência elétrica do microfone de carvão, gerando então um sinal que é captado pelo fone de ouvido dos dois aparelhos. Esta realimentação do sinal do microfone no fone de ouvido de cada usuário (*sidetone*) torna o som do telefone mais agradável e real, além de permitir o controle do volume da fala.

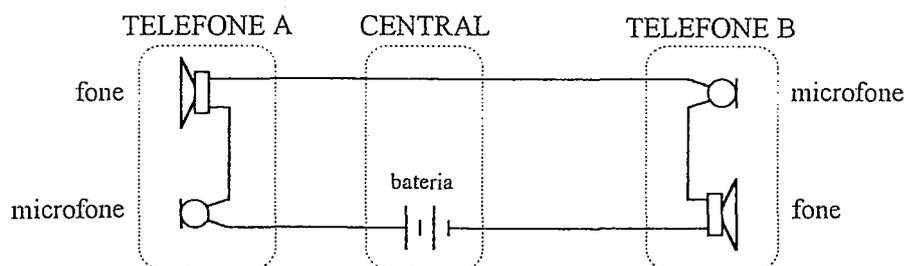


FIGURA 4.1: Sistema telefônico local a dois fios

A rede 4 fios / 2 fios mostrada na figura 4.2 consiste num acoplador magnético com a função de impedir o fluxo de corrente DC pelo alto-falante, mas permitir seu fluxo pelo microfone e ainda permitir a passagem de parte do sinal do microfone para o alto-falante. Esta rede recebe o nome de anti-local e é empregada nos aparelhos telefônicos.

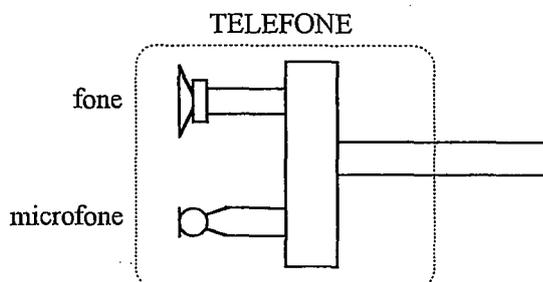


FIGURA 4.2: A anti-local

Para o caso de ligações a longa distância existe a necessidade de se colocar repetidores ao longo da linha. No entanto, estes repetidores exigem o uso de canais separados para o sinais, ou seja, há a necessidade de usar um sistema a 4 fios: 2 para o sinal de transmissão e mais 2 para o sinal de recepção.

Na figura 4.3 é representado o sistema telefônico usado em ligações de longa distância, onde **AL** representa a rede anti-local, **H** representa a híbrida e **R** indica o repetidor. O dispositivo usado nas centrais para transformar o sinal de 2 fios que vêm do assinante (telefone A) para os quatro fios que interligam as centrais é chamado de "híbrida".

A híbrida "ideal" passa o sinal de voz do assinante do telefone A para os dois fios do sinal de transmissão da central A (caminho t), e recebe o sinal da linha recepção da central A e o passa para a linha do assinante do telefone A (caminho r).

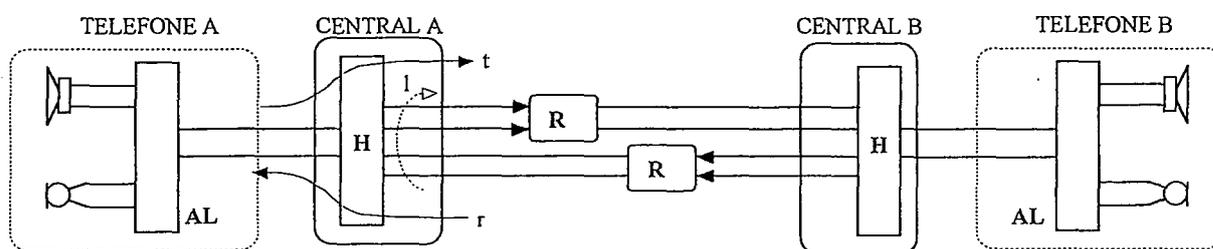


FIGURA 4.3: Sistema telefônico para longa distância

4.2.2 Ecos e sistemas para cancelamento

Na prática, o comportamento da híbrida é diferente do ideal, ocorrendo passagem de parte do sinal que chega na central destino (p.ex. A) de volta para a central de origem (p. ex. B), como mostra o caminho tracejado l. Este fato origina o problema de ecos em ligações

telefônicas [49]. Na figura 4.4 podem ser vistos os tipos mais comuns de eco que decorrem desta situação: *talker echo* (eco do locutor), *listener echo* (eco do ouvinte) e *singing*, situação onde não há atenuação suficiente na linha e o mesmo sinal percorre a linha muitas vezes em ciclo.

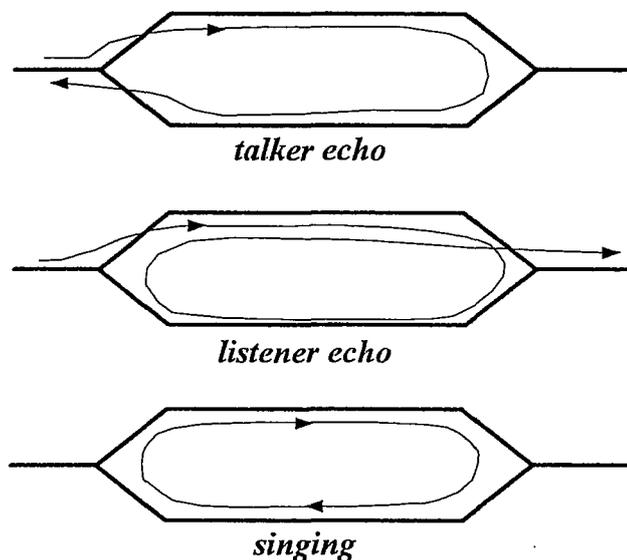


FIGURA 4.4: Tipos de eco

O efeito desse sinal de retorno (eco) depende de 3 fatores básicos: nível de amplitude do sinal de retorno, atraso do sinal de retorno e tolerância do locutor à presença deste sinal. Seu efeito é tolerável se a atenuação for razoável (6 dB ou mais) [31] e se o atraso for pequeno, porque é praticamente indistinguível da realimentação (*sidetone*) do telefone, sendo sentido como um aumento de volume. Para médias distâncias, o locutor pode ter a sensação de estar "falando num barril". Quanto maior a distância, mais atrasado chega o sinal de retorno e começa a ser percebido pelo locutor como sua própria voz (eco). Ele pode pensar que a outra pessoa está querendo interromper, e isto interfere no processo normal da fala. Quando temos uma ligação via satélite, o atraso em cada linha fica em torno de 270 ms em virtude da altitude dos mesmos. Isso significa que teremos um eco após cerca de 540 ms, o que torna muito difícil uma conversação.

Para a minimização dos problemas gerados pelo eco tem sido adotadas duas abordagens. A primeira consiste em interromper o caminho do sinal de retorno através de uma chave controlada, e a segunda consiste em subtrair do sinal de retorno uma estimativa para o eco. Estas abordagens geraram, respectivamente, supressores e canceladores de eco.

Um supressor de eco consiste em uma chave controlada pelo nível de intensidade dos sinais de voz, conforme a figura 4.5, sendo usado em circuitos de longa distância com atrasos de até 100 ms. Nesta figura, parte da central B é mostrada em um blocodiagrama e a conexão entre as duas centrais é simplificada. Enquanto uma pessoa estiver falando e não houver resposta do ouvinte, o supressor de eco do lado do ouvinte abre o caminho do sinal de retorno.

No entanto, quando houver resposta, a chave é fechada e o caminho de retorno é restabelecido novamente. Se as duas pessoas estiverem falando ao mesmo tempo não haverá interrupção do circuito, permanecendo o problema de eco. Quando o interlocutor faz uma pausa, o supressor libera o tráfego no sentido inverso após 10 ms.

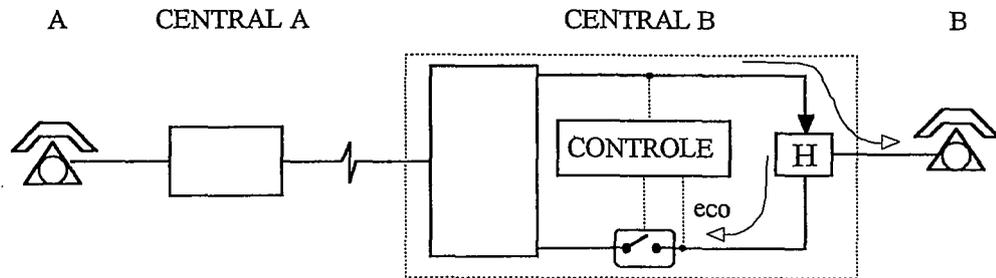


FIGURA 4.5: Supressor de eco

Este chaveamento dos canais introduz uma característica desagradável ao ruído de fundo e à voz de quem responde, já que o chaveamento não é imediato, mas tem alguns milissegundos de atraso. Esta característica piora quando o atraso da linha aumenta, como em circuitos via satélite. Nestes casos devem ser usados canceladores de eco.

O cancelador de eco consiste num filtro adaptativo ligado conforme ilustrado na figura 4.6. Nesta configuração o filtro (H_{fa}) tende a se ajustar de modo a produzir uma réplica do eco com base no sinal que chega à híbrida. Esta réplica de sinal é subtraída do sinal que retorna da híbrida para a linha. Idealmente, o filtro adaptativo ajusta a sua função de transferência de modo que esta seja o mais próxima possível da função de transferência do conjunto "híbrida+telefone" para o sinal de eco. Se estas forem idênticas, todo o sinal de eco será eliminado. Note que o caminho não é interrompido.

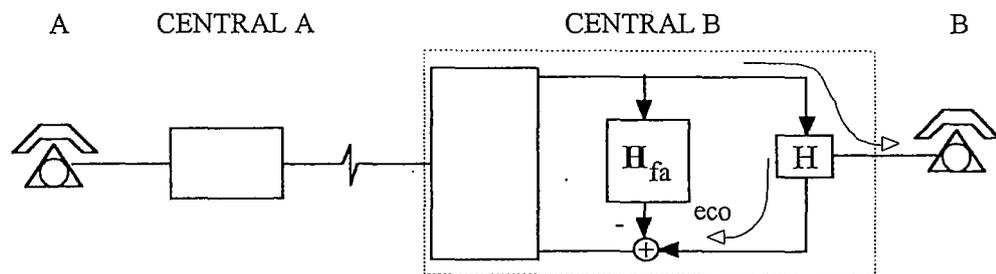


FIGURA 4.6: Cancelador de eco

O modelo digital pode ser visto na figura 4.7, onde são mostradas as funções de transferência para a análise do problema no domínio Z [33]. A influência do conjunto "híbrida+telefone" sobre o sinal de voz que gerará o eco é representada por $H_t(z)$. Este sinal é somado à voz da pessoa no telefone B quando esta estiver falando.

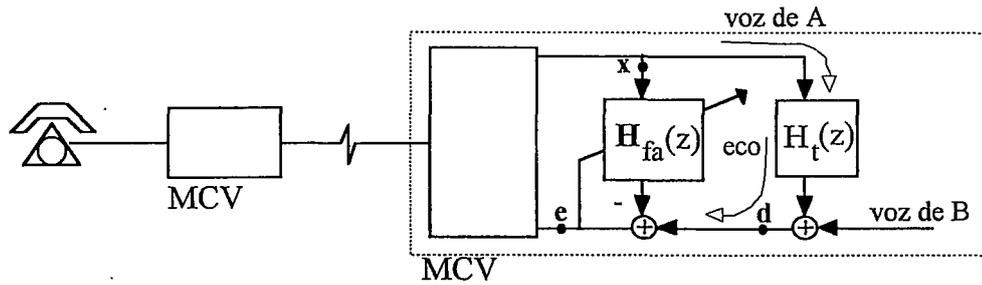


FIGURA 4.7: O modelo digital para o problema de cancelamento de eco

A semelhança deste problema com a configuração de cancelador adaptativo de interferência é evidenciada na figura 4.8, onde d é o sinal primário (voz de B e sinal de eco), x é o sinal de referência (voz de A que gera o sinal de eco) e e é a saída do cancelador (sinal que efetivamente retorna para A). Neste caso, o sinal de interesse é o sinal de voz de B. A interferência é o sinal de eco. A estimativa para a interferência é o sinal de voz de A.

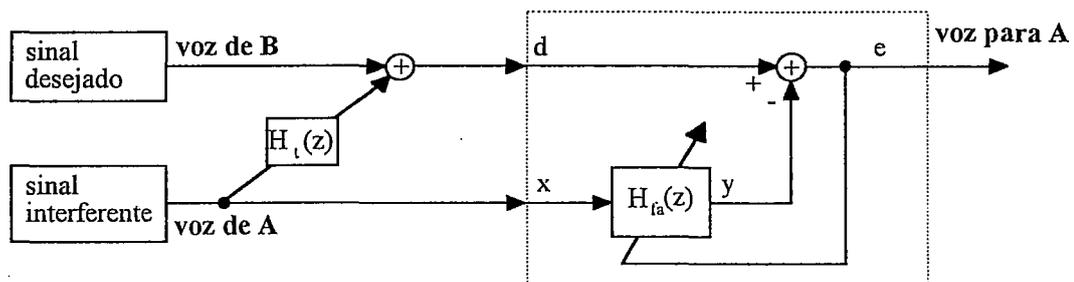


FIGURA 4.8: O cancelador adaptativo de ruído como um cancelador de eco

Características como não-linearidades na função de transferência da híbrida e características implícitas de filtros adaptativos, como desajustamento e ruído no vetor de coeficientes limitam a performance desse tipo de sistema. Em sistemas de cancelamento em uso atualmente é utilizado o supressor de eco depois do estágio cancelador, para eliminar o eco residual após o cancelamento.

4.3 O Módulo Compressor de Voz -MCV

4.3.1 O projeto Digitel-LaPSi

Com o intuito de otimizar a utilização de um canal telefônico para transmissão simultânea de dados e de sinais de voz foi desenvolvido pelo LaPSi, com o apoio da empresa Digitel, o MCV - Módulo Compressor de Voz, como pode ser visto na figura 4.9. O objetivo é utilizar um mesmo canal de voz, que esteja sendo utilizado para transmissão de dados (p.ex., a

uma taxa de 19200 bps), para transmitir sinais de voz também (mas com a metade da taxa de transmissão de dados, i.e., a uma taxa de 9600 bps).

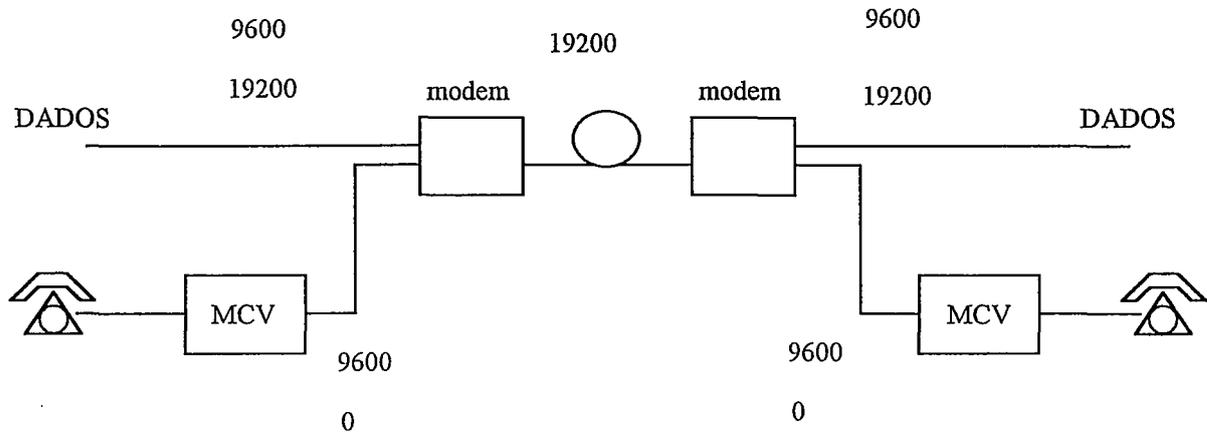


FIGURA 4.9: Utilização do módulo compressor de voz

Um canal de voz típico de telefone possui uma banda que vai de 300 Hz a cerca de 3300 Hz. A uma taxa de amostragem de 7400 Hz e utilizando conversores analógico-digitais de 14 bits, a taxa de transmissão deveria ser de $7400 \times 14 = 103600$ bps. A função do módulo compressor de voz é, então, compactar os dados do sinal de voz digitalizado e transmitir estes dados compactados a uma taxa bem inferior à que seria necessária sem o uso de compactação. Métodos para compactação de sinais de voz são discutidos por Weihmann [45].

4.3.2 O problema de eco específico do MCV

A ligação entre os dois aparelhos telefônicos feita pelos módulos MCV pode ser vista conforme mostra a figura 4.10. O comportamento do MCV neste caso é similar ao de uma central telefônica, pois possui um circuito de transformação de 2 para 4 fios (híbrida). O uso de 4 fios é necessário para que seja feita a digitalização, processamento, transmissão e recepção do sinal de voz.

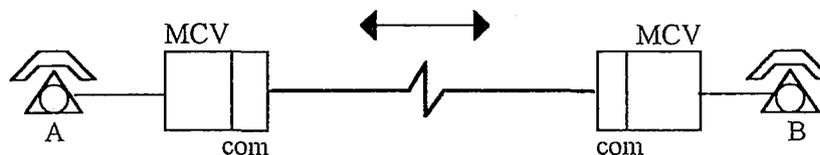


FIGURA 4.10: Ligação entre aparelhos telefônicos pelos módulos MCV

A presença da híbrida, neste caso um circuito ativo, ocasiona o mesmo problema de eco discutido anteriormente no contexto de centrais telefônicas. O diagrama em blocos do sistema MCV pode ser visto na figura 4.11, onde pode ser percebida uma etapa analógica,

formada pelos circuitos de interface com o aparelho telefônico propriamente dito, incluindo a híbrida, e por uma etapa digital, encarregada da comunicação e processamento dos sinais.

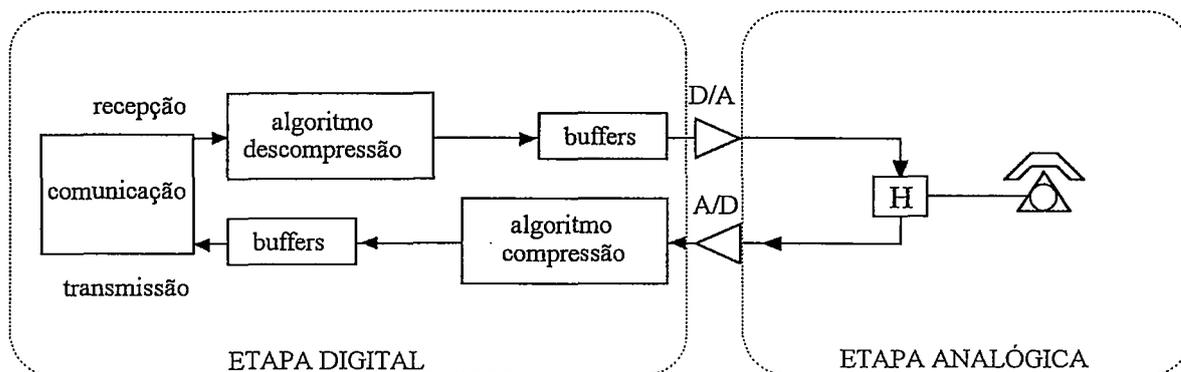


FIGURA 4.11: Diagrama em blocos do Módulo Compressor de Voz

A etapa digital do módulo é implementada através de um microprocessador específico para processamento digital de sinais, o DSP TMS320C25, e o conversor utilizado é o TLC320C44, de 14 bits e comunicação serial com o processador, que possui conversores AD e DA com filtros *anti-aliasing* embutidos no mesmo componente.

O funcionamento dos algoritmos de codificação e compressão de sinais ocupam a maior parte do tempo da CPU do processador e memória interna. Este detalhe é importante porque a implementação do cancelador de eco neste sistema terá como primeira restrição o uso de pouco tempo do processador, ou seja, o número de instruções a serem usadas será limitado.

O segundo detalhe prático destes módulos é que a comunicação entre os mesmos é assíncrona, o que ocasiona a perda de uma amostra de sinal, periodicamente, a cada 15 s em média. Apesar de não ter influência significativa na restauração dos sinais em si, este fato pode influenciar no funcionamento do sistema cancelador de eco.

O sinal de eco que retorna para o locutor possui uma intensidade subjetiva alta, tornando bastante difícil uma conversação, mesmo que as pessoas tenham uma boa tolerância ao problema. O atraso ocasionado pelo processamento e transmissão do sinal de voz faz com que o sinal de eco chegue com um atraso total de cerca de 100 ms após o sinal original de voz.

4.4 Abordagem do problema

4.4.1 Abordagem inicial

Segundo a CCITT [20], canceladores de eco possuem as seguintes requisições fundamentais:

- i) rápida convergência;
- ii) baixo nível subjetivo de eco durante a fala do locutor (apenas uma pessoa falando);
- iii) pouca divergência durante períodos de fala simultânea (duas pessoas falando);

De acordo com as restrições sofridas pelo sistema vistas no item anterior, verifica-se que o filtro adaptativo a ser implementado deve ter uma performance aceitável, conforme as requisições acima, e deve ter uma pequena carga computacional.

Conforme visto no capítulo 3, estruturas IIR possuem menor carga computacional em comparação com filtros FIR semelhantes, mas possuem problemas que podem tornar sua aplicação inviável. Em contrapartida, filtros FIR com um pequeno número de coeficientes apresentam uma carga computacional pequena. O uso de algoritmos derivados de métodos de busca pelo gradiente se faz necessário devido à menor carga computacional e simplicidade de implementação.

Deve ser salientado que a influência do sistema onde será implementado o cancelador é bastante grande, pois o processador usado não possui representação numérica em ponto flutuante, o que faz necessário tomar certas precauções sobre resultados numéricos e efeitos de precisão limitada enfrentados. No caso do TMS320C25, números são representados por uma palavra que possui 16 bits e a unidade lógica e aritmética armazena os resultados de multiplicações em 32 bits [41].

O sistema teve como abordagem inicial a implementação de um cancelador de eco com base num filtro adaptativo de estrutura direta e algoritmo LMS. Nesta abordagem os coeficientes do filtro foram representados com dupla precisão (32 bits) para a adaptação e simples precisão (16 bits) para o cálculo do filtro, sendo a constante de adaptação com valor fixo, implementada como um deslocamento. Esta implementação usou o algoritmo LMS para a adaptação dos coeficientes e código linear. A performance desse sistema deixou bastante a desejar, ou seja, o nível subjetivo do eco de retorno era alto, já que apenas 16 taps podiam ser calculados sem causar problemas ao sistema (*crash*).

Para que fosse diminuída a carga computacional desta implementação foi necessário então buscar o uso de coeficientes em simples precisão (16 bits) sempre e utilizar instruções em modo de repetição quando possível. No entanto, como grande parte da carga computacional está no algoritmo de adaptação dos coeficientes, buscou-se um algoritmo que tivesse uma performance comparável ao algoritmo LMS e menor carga. O uso dos algoritmos LMS do tipo *sign* foi descartado devido a sua performance ser teoricamente inferior à do algoritmo LMS, embora pudessem ser tentados em caso de não haver outra solução.

Foi usado então um algoritmo baseado no LMS mas que adapta os coeficientes do filtro em blocos, de modo a poupar ciclos de máquina e permitir a implementação de filtros mais longos (com maior número de coeficientes). Este algoritmo, *block update* LMS, possui comportamento semelhante ao algoritmo LMS [31], ao contrário dos algoritmos do tipo *sign*.

O algoritmo *block update* LMS é apresentado no anexo D. Esta implementação permitiu um aumento do número de coeficientes do filtro de 16 para 64, com uma acentuada melhora da atenuação subjetiva de eco.

Os resultados obtidos para níveis de atenuação absolutos para sinais de voz são mostrados na figura 4.12, conforme o procedimento descrito no anexo C. Os dados mostrados são para o algoritmo *block update* LMS com fator de convergência variável em função da potência do sinal de entrada, que mostrou-se como o melhor dentre os algoritmos testados do ponto de vista de atenuação absoluta e velocidade de convergência.

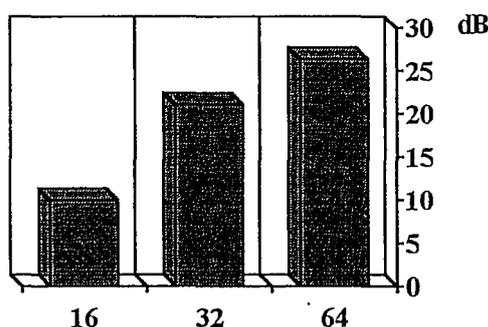


FIGURA 4.12: Atenuação do eco de retorno em função do número de coeficientes do filtro

Nota-se que a atenuação de um filtro com 16 coeficientes é bastante inferior à atenuação de um filtro com 64 coeficientes. Esse comportamento era esperado, pois o aumento do número de coeficientes possibilita ao filtro adaptativo formar uma estimativa melhor para a função de transferência do conjunto "híbrida+telefone".

O algoritmo divide os coeficientes do filtro em 8 blocos. Em termos de carga computacional, como pode ser visto na tabela 4.13, não há um aumento linear com o aumento do número de coeficientes. Esta é uma das maiores vantagens deste algoritmo, enquanto seu principal problema é a diminuição da velocidade de convergência.

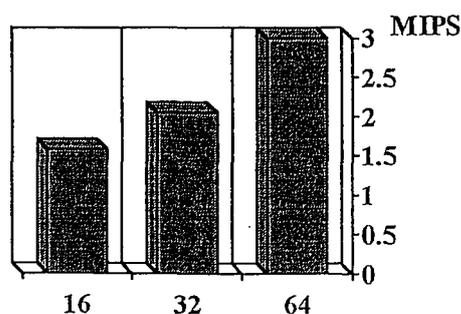


FIGURA 4.13: Carga computacional do cancelador de eco

4.4.2 Porque a atenuação não alcança níveis maiores ?

Apesar de apresentar uma performance subjetiva melhor que a do cancelador de 16 coeficientes, o eco residual presente ainda podia ser percebido e, dependendo do grau de tolerância do locutor, poderia prejudicar a conversação ou inviabilizar esta implementação, indicando a necessidade de uso de outros processadores e algoritmos para a solução do problema. Este nível de eco residual impede o uso de supressores de eco no estágio de saída do cancelador devido ao efeito de chaveamento ser perceptível.

Com o intuito de identificar as causas deste aparente limite de atenuação do sinal de eco, foram então estudadas as seguintes hipóteses:

- 1) Não existe um número de coeficientes suficiente para a síntese da função de transferência do conjunto "híbrida+telefone". Em consequência, a atenuação do eco será limitada por esse fato.
- 2) A perda de amostra que ocorre devido à falta de sincronismo entre os módulos causa instabilidades que impedem que o cancelador de eco alcance a atenuação máxima.

4.4.2.1 Identificação da duração da resposta ao impulso do conjunto "híbrida+telefone"

Para que o cancelador de eco atinja sua performance ótima, é necessário que a função de transferência do filtro adaptativo seja igual à função de transferência do conjunto "híbrida+telefone". Esta função pode conter pólos. Neste caso, a função de transferência obtida pelo filtro adaptativo gerará a melhor aproximação possível e o cancelamento não será perfeito.

Foi feito o levantamento da função de transferência do circuito da híbrida com o telefone fora do gancho e com o microfone substituído por um resistor escolhido adequadamente, a fim de não introduzir ruídos do ambiente nas medidas. Este levantamento foi baseado na análise da resposta ao salto aplicado na entrada do circuito da híbrida. O procedimento seguido e os resultados obtidos estão colocados no anexo A, onde pode ser notada a presença de pólos pouco amortecidos e de baixa frequência na função de transferência $H_i(z)$.

Do ensaio de resposta ao salto pode-se notar que a parte mais significativa da função de transferência é coberta por um filtro de 64 taps com $f_s=7.4$ kHz, (8.6 ms) embora esta resposta se estenda por um tempo maior (de cerca de 30 ms), conforme visto na resposta ao salto mostrada na figura 4.14. Para gerar uma função de transferência dessa ordem de atraso seriam necessários cerca de 220 coeficientes. A resposta ao impulso de um sistema FIR possui duração finita, dada pelo número de coeficientes do sistema. A resposta ao salto desses sistemas tende a um valor constante, e a duração do período transitório é dada pelo número de coeficientes do sistema.

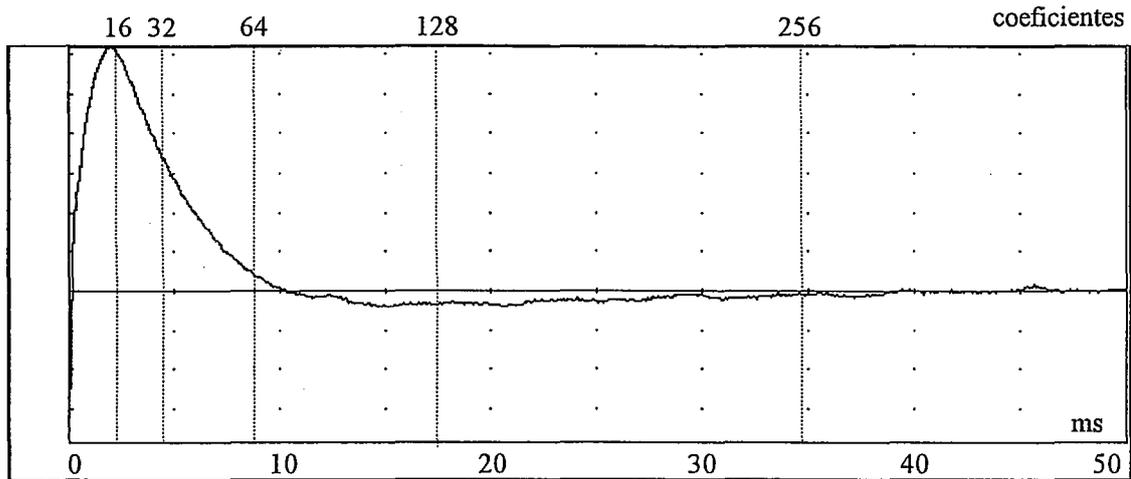


FIGURA 4.14: Resposta ao salto do conjunto "híbrida+telefone"

Para que fosse confirmada a hipótese de que o número de coeficientes do filtro era insuficiente, seria necessário passar para uma fase de experimentação. No entanto, esta implementação não poderia ser feita porque requereria maior carga computacional que a disponível. Para resolver este problema foram feitas modificações nos módulos a nível de *software* e *hardware*, que consistiram na retirada da parte de codificação e compressão de sinal e na sincronização dos módulos. Estas modificações alteraram levemente a taxa de amostragem do sinal (de 7400 para 7200 Hz) e aumentaram sensivelmente a taxa de transmissão da comunicação serial (de 9600 bps para 115200 bps). Estas mudanças permitiram também isolar os problemas de perda de amostra e de falta de coeficientes, já que o problema de perda de amostra era uma decorrência da falta de sincronismo entre os módulos.

Com este novo sistema foi feita uma extensa bateria de testes sobre a influência de fatores como número de coeficientes, número de blocos e fator de convergência, na atenuação do sinal de eco e na velocidade de convergência do algoritmo. O método empregado para a obtenção das medidas e o quadro completo são apresentados com mais detalhes no anexo C. Também foram feitas medidas com os mesmos parâmetros usando o sistema com codificação.

A figura 4.15 apresenta os resultados de atenuação absoluta para sinais de voz obtidos pelo algoritmo para a implementação sem o uso de compressão e com os módulos sincronizados. O algoritmo é o mesmo usado na implementação com o uso de compressão e comunicação assíncrona entre os módulos, cujos resultados obtidos foram apresentados na figura 4.12.

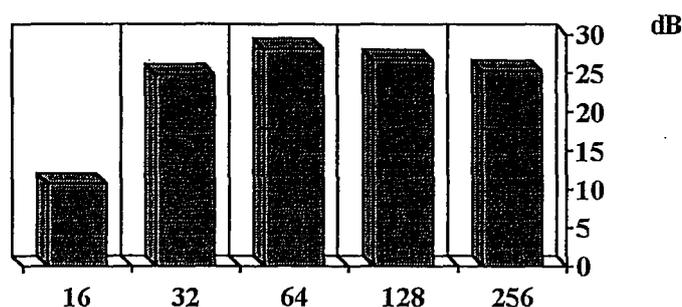


FIGURA 4.15: Atenuação do eco de retorno em função do número de coeficientes do filtro

A carga computacional desta implementação é apresentada na figura 4.16. As diferenças em relação à outra implementação são devidas ao uso de memória interna do processador (mais rápida) e de instruções de deslocamento específica para esta memória, que antes estava em uso pelos algoritmos de compressão e descompressão de sinal.

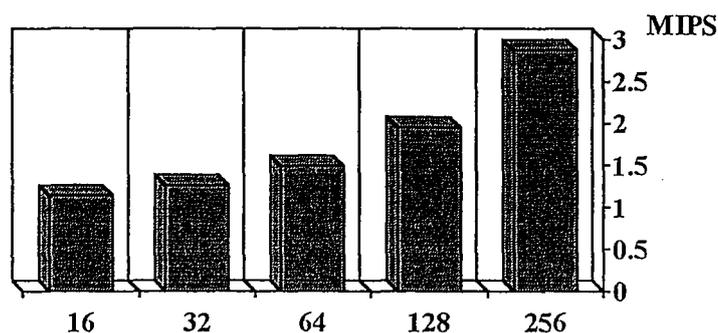


FIGURA 4.16: Carga computacional do cancelador de eco

4.4.2.2 Análise das medidas de atenuação

A análise das medidas de atenuação pode ser feita através da figura 4.17, que mostra um quadro comparativo das atenuações obtidas para os casos dos módulos sincronizados (coluna da esquerda) e não sincronizados (coluna da direita).

Nota-se que a maior atenuação é atingida quando o filtro possui 64 coeficientes. Do ponto de vista dessa implementação, esse efeito só pode ser justificado por problemas numéricos e erros decorrentes do caráter adaptativo do filtro, como desajustamento. Esses efeitos são proporcionais ao aumento do número de coeficientes, tendendo a introduzir erros. O resultado final é a curva de atenuações obtida, que cresce até atingir o máximo com 64 coeficientes, onde o efeito de elevar o número de coeficientes melhora a aproximação da

função de transferência. Após este valor o aumento do número de coeficientes causa a queda da atenuação máxima, devido à introdução de erros numéricos e de desajustamento.

No entanto, esses efeitos não explicam a diferença entre as atenuações máximas para um mesmo número de coeficientes, presentes na implementação normal, que são de .66, 3.72 e 1.7 dB. Em escala linear, estas diferenças correspondem a 1.079, 1.424 e 1.216 vezes, o que pode ser melhor observado na figura 4.18.

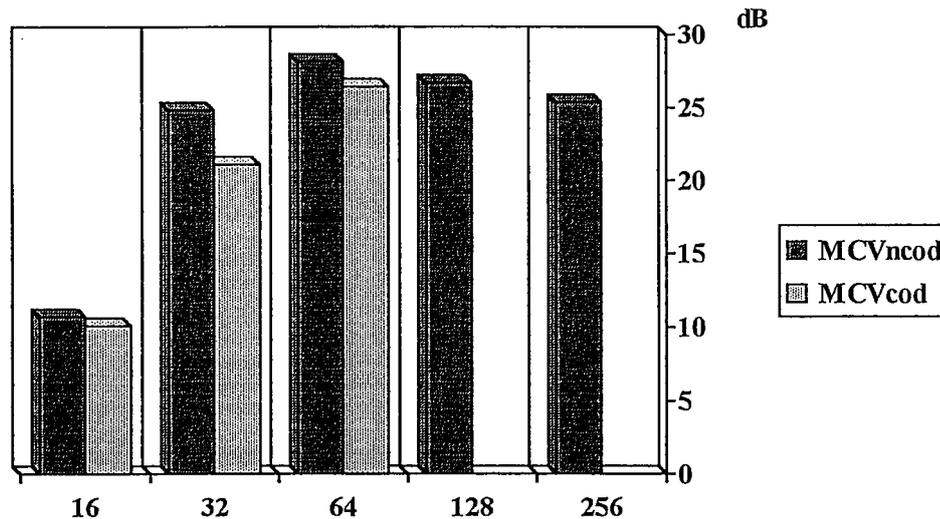


FIGURA 4.17: Quadro comparativo de atenuações

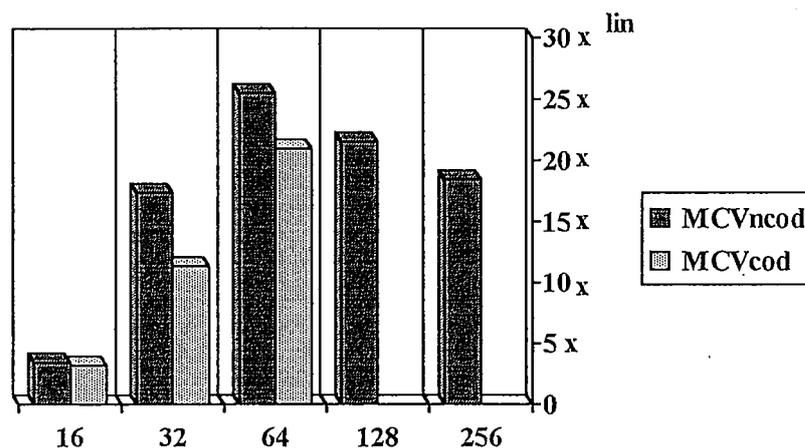


FIGURA 4.18: Quadro comparativo de atenuações em escala linear

Como resultados parciais da análise observando apenas os níveis de atenuação máximos que foram obtidos com o filtro de 64 coeficientes, pode-se dizer que este seria o melhor comprimento de filtro do ponto de vista desta implementação. Isso decorre do fato de apresentar a melhor relação atenuação / carga computacional de todos os casos estudados.

4.4.2.3 Análise do efeito de perda de amostra

Da análise anterior nota-se que existe uma diferença significativa entre os níveis de atenuação de eco para as duas implementações. As diferenças principais destas implementações estão na mudança da taxa de amostragem do sinal de voz (de 7400 Hz para 7200 Hz) e na eliminação do efeito de perda de amostra através da sincronização dos módulos. A última implementação mostrou-se superior no nível de atenuação máximo atingido.

Foram feitas simulações sobre a possibilidade de o efeito de perda de amostra influir sobre a performance do sistema de cancelamento de eco. Como as medidas de atenuação foram tomadas após um intervalo de tempo grande, de 1 a 3 minutos, o efeito de perda de amostra a cada 15 segundos possui frequência considerável dentro do intervalo de medida.

Foi então simulado o efeito de perda de amostra nos arquivos utilizados para as simulações apresentadas no capítulo 3. Neste caso, foi simulada uma perda de amostra a cada 2 segundos (20000 amostras) no arquivo utilizado para referência do cancelador. O resultado da simulação, utilizando os mesmos parâmetros para o filtro adaptativo e algoritmo LMS, é apresentado na figura 4.19, juntamente com a simulação sem o efeito de perda de amostra.

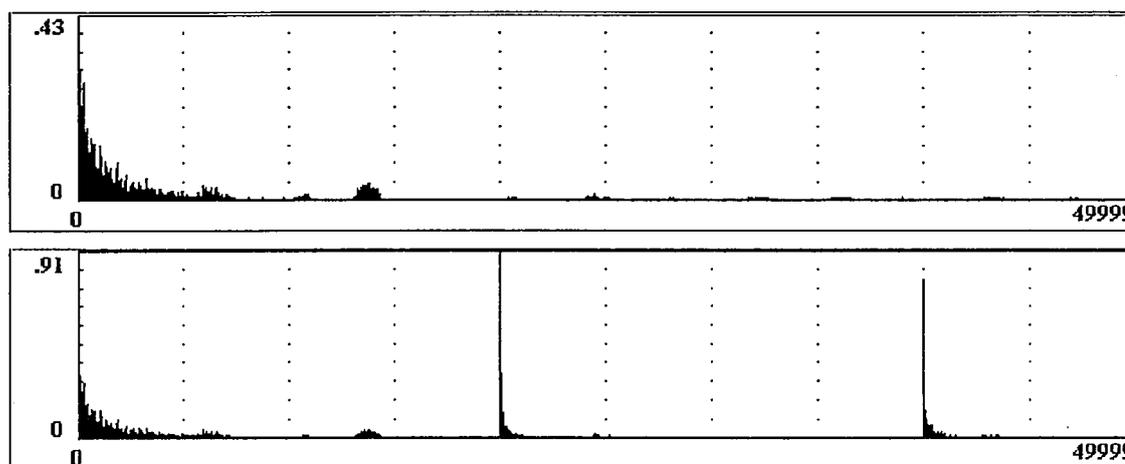


FIGURA 4.19: Curvas de aprendizagem instantâneas sem e com perda de amostra (simulação)

Nota-se neste caso que a perda de amostra faz com que o erro instantâneo fique em níveis elevados. Esse efeito introduz, então, uma perturbação no sinal de erro e, conseqüentemente, no próprio filtro adaptativo. No caso simulado o efeito introduzido pela perda de amostra é compensado pelo próprio mecanismo de adaptação em pouco tempo (cerca de 1/4 de segundo, ou 2500 amostras).

No caso do algoritmo implementado em tempo real no MCV, foram feitos alguns ensaios para verificar o tempo de estabilização do algoritmo nas suas condições de operação normais (com codificação de voz e módulos assíncronos - MCVcod) e nas condições de operação modificadas (sem codificação de voz e módulos sincronizados - MCVncod). Alguns

destes resultados podem ser vistos na figura 4.20. Estes resultados apresentam o tempo que o cancelador de eco leva para atingir uma atenuação de cerca de 70% da máxima atenuação no caso em questão [ANEXO C].

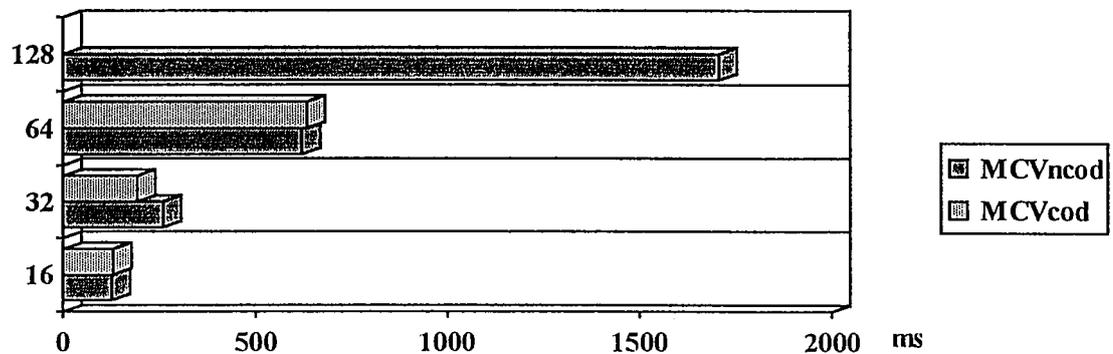


FIGURA 4.20: Comparação do tempo para atingir cerca de 70% da atenuação máxima

No entanto, devido ao caráter exponencial da curva de aprendizagem, o tempo para atingir a atenuação máxima é bem maior. Como exemplo, é mostrada na figura 4.21 a curva de aprendizagem para o filtro com 64 coeficientes, adquirida pela placa de DSP durante 5s, com o módulo assíncrono e com compressão de voz. A primeira marca mostra o início do processo de adaptação, e a segunda mostra a retirada do sinal de entrada do cancelador (ruído), restando apenas o ruído ambiente [ANEXO C]. Percebe-se que a tendência da curva era de queda no instante em que foi interrompido o ensaio.

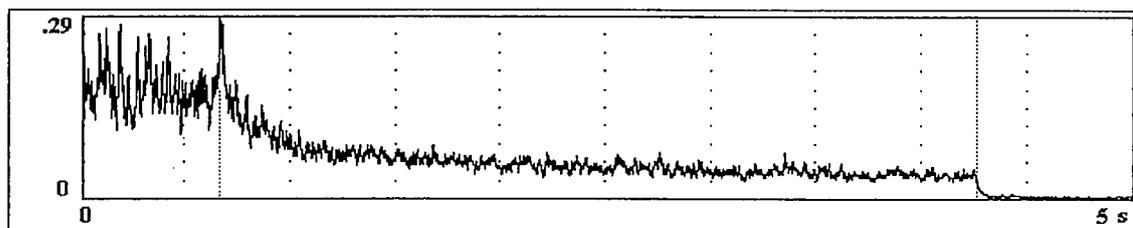


FIGURA 4.21: Curva de aprendizagem instantânea do módulo MCV digitalizada

Dessas observações pode-se então dizer que o efeito de perda de amostra influi na atenuação máxima do modo em que esta foi medida, pois o longo tempo possibilita a ocorrência de muitas situações que produzem efeitos que tendem a diminuir a atenuação até então obtida. Mesmo que o cancelador recupere a atenuação até então alcançada, esta não será igual àquela que poderia ter sido atingida se não houvesse a perturbação causada pela perda de amostra.

4.4.2.4 Conclusões

Da análise feita nas seções anteriores pode-se então concluir que o número de coeficientes para o filtro adaptativo escolhido, 64, embora incapaz de igualar toda a duração da resposta ao impulso do conjunto "híbrida+telefone", apresentou na implementação a melhor atenuação máxima. Este fato se deve ao aparecimento de erros proporcionais ao número de coeficientes, que começam a se tornar significativos nos filtros maiores (128 e 256 coeficientes).

A diferença de atenuação máxima observada entre as implementações permite deduzir que o efeito de perda de amostra introduz perturbações que provocam a diminuição da atenuação máxima, principalmente devido ao modo como as medidas de atenuação foram feitas. O fato de esperar alguns minutos para o filtro atingir a máxima atenuação permite que o filtro seja submetido a um grande número de perturbações que pioram sua performance final face ao filtro não submetido a estas mesmas perturbações.

O efeito de distorções introduzidas no sinal reconstruído pelo algoritmo de codificação não foi considerado, mas pode ser responsável por parte desse comportamento para sinais de voz. Para ruído, as distorções do algoritmo causaram uma drástica diminuição da atenuação.

4.4.3 Qual a impressão do usuário sobre o funcionamento do cancelador implementado ?

Com o intuito de avaliar o fator humano de tolerância ao problema de eco foram realizadas algumas medidas subjetivas e informais sobre o desempenho dos canceladores de eco implementados. As medidas foram realizadas conforme descrito no anexo B, onde também encontram-se os resultados obtidos. Foi testado o desempenho de canceladores de 16, 32, 64 e 128 coeficientes, com os algoritmos que apresentaram as maiores atenuações para estes casos.

Com o objetivo de avaliar o nível de eco residual em relação ao cancelador desligado (nível absoluto), foi questionado como a pessoa classificaria o eco residual numa escala entre muito alto, alto, baixo, muito baixo e baixíssimo. O nível "muito alto" estaria associado à ausência de cancelador de eco. Os resultados podem ser vistos na tabela 4.1, com os níveis correspondentes de atenuação do sinal (aproximados).

TABELA 4.1: Níveis subjetivos de amplitude do eco residual e atenuação do sinal

16	32	64	128
médio (entre alto e baixo)	muito baixo	muito baixo	muito baixo
10.7 dB	24.8 dB	30.5 dB	27.7 dB

Foi feita a pergunta se o nível de eco percebido atrapalharia uma conversa telefônica, ou seja, se a pessoa julga o cancelador de eco adequado ou não para o uso normal. O cancelador de 16 coeficientes não obteve consenso entre as respostas do grupo, ao passo que os demais foram julgados adequados ao uso.

Também foram feitas medidas no sentido de avaliar o desempenho relativo dos canceladores de eco, um em relação ao outro. Assim, foram apresentados os canceladores de 16, 32, 64 e 128 coeficientes e, em cada apresentação, pedia-se que a pessoa classificasse o desempenho do cancelador em relação ao cancelador anterior, como muito melhor, pouco melhor, igual, pouco pior e muito pior. Os resultados podem ser vistos na tabela 4.2.

TABELA 4.2: Desempenho subjetivo do cancelador em relação ao cancelador anterior (com menos coeficientes)

16	32	64	128
muito melhor que sem o cancelador	muito melhor que o anterior (16)	igual ao anterior (32)	igual ao anterior (64)

Esta análise subjetiva indica que o grupo analisado de 19 pessoas não notou diferença significativa entre o desempenho de um cancelador de eco com 32 coeficientes ou mais. Outro fator percebido é que os canceladores foram tidos como muito lentos, ou seja, seu tempo de convergência foi considerado muito elevado. Com base nestes dados, então, parece ser mais interessante para a implementação um cancelador de eco que atinja a atenuação do cancelador de 32 coeficientes usado nas medidas. A carga computacional disponível deve ser usada para melhorar a velocidade de convergência do cancelador nestas condições.

4.4.4 Resultados de implementações alternativas

4.4.4.1 Algoritmos IIR - SHARF e *Recursive LMS*

Com o objetivo de avaliar o desempenho de um filtro IIR em substituição a um filtro FIR usual, foram realizadas algumas implementações com estrutura IIR e algoritmos SHARF e *Recursive LMS*. Os testes mediram a atenuação máxima obtida segundo o mesmo critério usado para os filtros FIR anteriores. Os resultados para as implementações dos algoritmos *Recursive LMS* e SHARF com 3 pólos e 3 zeros e também com 8 pólos e 8 zeros podem ser vistos na figura 4.22.

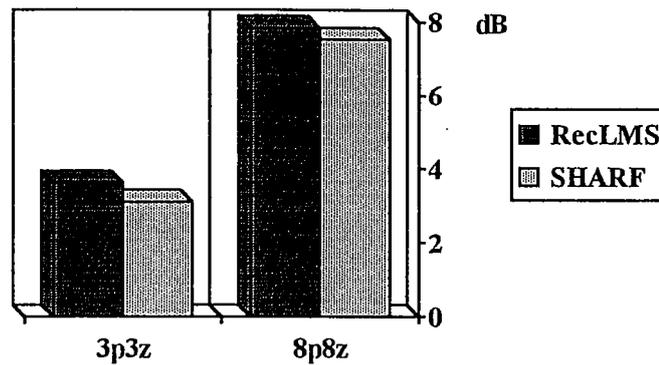


FIGURA 4.22: Níveis de atenuação absolutos para algoritmos IIR implementados

A carga computacional destas implementações pode ser vista na figura 4.23. O fato de ser relativamente elevada é que esta implementação utilizou-se de memória externa. Em relação à figura 4.13, nota-se que a carga computacional dos algoritmos em relação ao nível de atenuação obtido é inferior ao filtro FIR de 16 coeficientes. Portanto, esta classe de filtros não se mostrou adequada para esta aplicação.

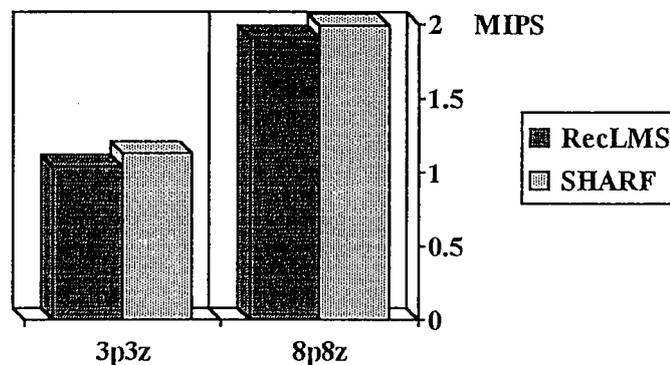


FIGURA 4.23: Carga computacional dos algoritmos IIR

Segundo FAN [10], filtros IIR podem ser superiores em termos de atenuação máxima a filtros FIR após a convergência. No entanto, fatores como ordem insuficiente do filtro e presença de erros e ruído degradam consideravelmente o desempenho atingido. Conforme levantamento da função de transferência do conjunto "híbrida+telefone", foi observado que uma boa aproximação seria usar uma filtro com 3 pólos e 3 zeros. No entanto, esta função apresenta pólos de baixa frequência e pouco amortecidos, que podem prejudicar bastante a velocidade de convergência destes algoritmos, até mesmo impedindo a convergência [10] [9].

4.4.4.2 Algoritmo *lattice* simplificado

Devido aos bons resultados da estrutura *lattice*, foi feita uma implementação simplificada utilizando fatores de adaptação constantes, ao invés de variáveis conforme a potência, conforme descrito no capítulo 3. Os resultados obtidos para os níveis de atenuação e carga computacional são apresentados no anexo D. Para o caso de um filtro de 15 estágios e fatores de convergência constante, foi medida uma atenuação de 5.37 dB para sinais de voz e com uma correspondente carga computacional de 4.84 MIPS. Nota-se que, embora otimizado, o algoritmo apresentou uma carga computacional muito elevada para a pequena atenuação máxima obtida. Para uma implementação similar do algoritmo sem simplificações, a carga computacional seria de cerca de 13.8 MIPS.

Note-se que este nível de carga computacional está vinculado ao DSP usado, que não possui instruções dedicadas à implementação deste tipo de algoritmos, como a instrução de repetição de blocos. Neste caso, o uso de algoritmos *lattice* não é uma alternativa viável do ponto de vista computacional.

4.5 Outros resultados experimentais

4.5.1 Efeitos de variação do fator de convergência e do número de blocos

Foram feitas medidas de atenuação e tempo de convergência visando observar a influência de fatores como o número de blocos e o valor do fator de convergência. Estas medidas foram tomadas para os diferentes números de coeficientes do filtro e para a implementação síncrona, sem codificação.

A influência do fator de convergência na atenuação pode ser observada através da análise da figura 4.24, onde são apresentadas medidas de atenuação para sinais de voz e para os casos de fator de convergência $\mu/2$, μ , 2μ e, finalmente, μ variável de acordo com a potência do sinal, podendo atingir um máximo de 5μ .

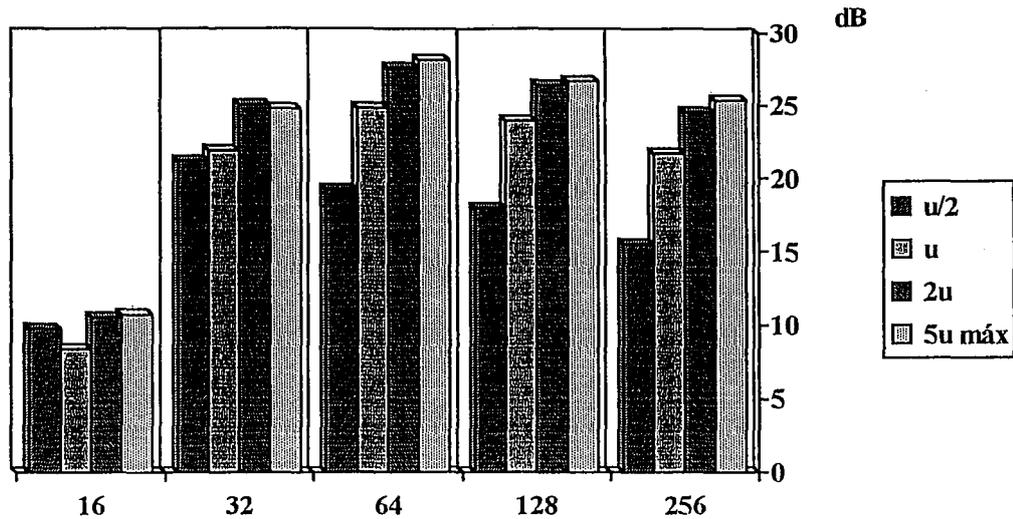


FIGURA 4.24: Atenuação em função do número de coeficientes e do fator de convergência

As implementações para um mesmo número de coeficientes possuem carga computacional semelhante. Percebe-se a tendência de aumento da atenuação com o aumento do fator de convergência, se escolhido dentro de uma faixa estável. Esse aumento de atenuação indica que os erros dentro da implementação com mesmo número de coeficientes, como desajustamento, são menores que o ganho de atenuação conseguido devido ao efeito de amplificação do sinal de erro produzido por um fator de convergência maior. A tendência de limitação da atenuação máxima com o aumento de μ , percebida melhor nos 3 últimos grupos, indica a faixa onde os erros proporcionais a μ começam a se tornar mais acentuados.

A avaliação da influência do número de blocos do algoritmo *block update* LMS na atenuação pode ser visto na figuras 4.26. A figura 4.25 mostra o tempo de convergência para que o filtro atinja cerca de 70% da sua atenuação máxima.

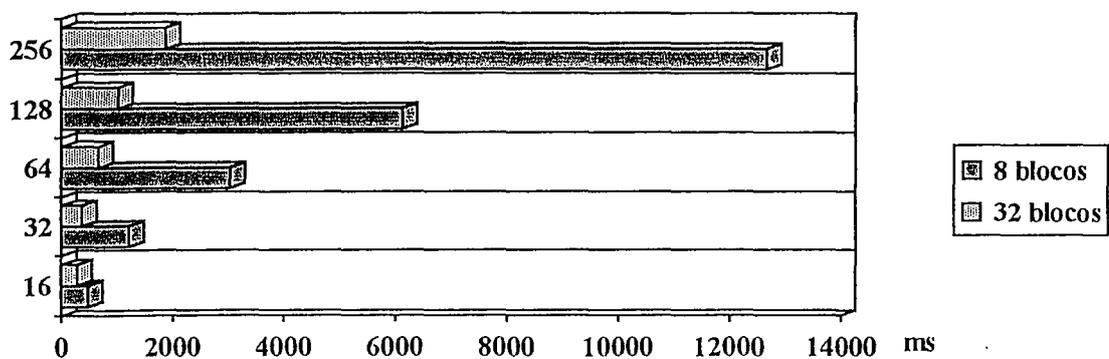


FIGURA 4.25: Tempo para o cancelador atingir cerca de 70% da sua atenuação máxima

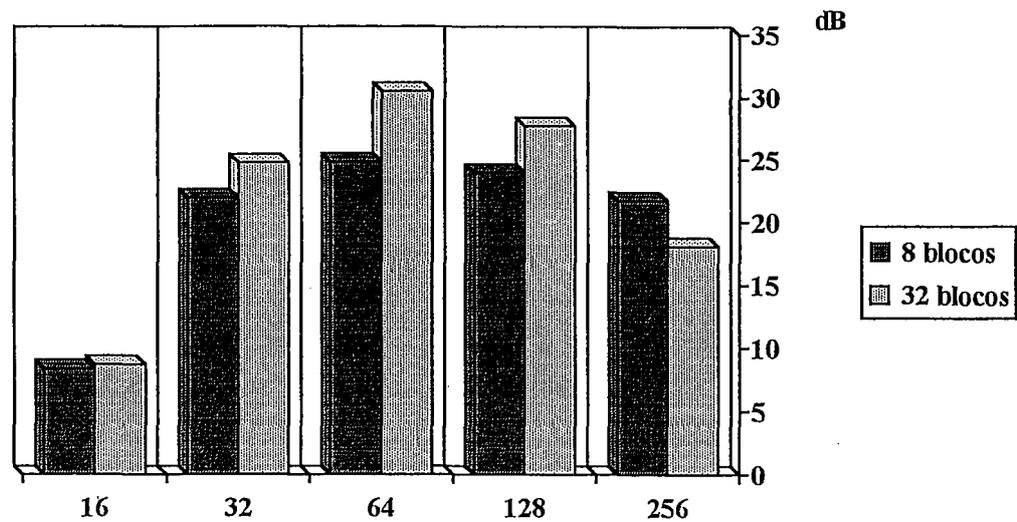


FIGURA 4.26: Atenuação máxima obtida em função do número de blocos

Nestas figuras, o filtro de 16 coeficientes possui 16 blocos ao invés de 32. Na realidade, este número de blocos indica que todos os coeficientes destes dois filtros serão adaptados a cada instante, ou seja, tem-se o algoritmo LMS normal. Nota-se, analisando a figura, que a performance do cancelador fica prejudicada pela adoção de uma estratégia de adaptação em blocos, tanto em termos de atenuação máxima como em termos de velocidade de convergência. A grande vantagem desta estratégia consiste, no entanto, na redução da carga computacional, conforme pode ser notado na figura 4.27.

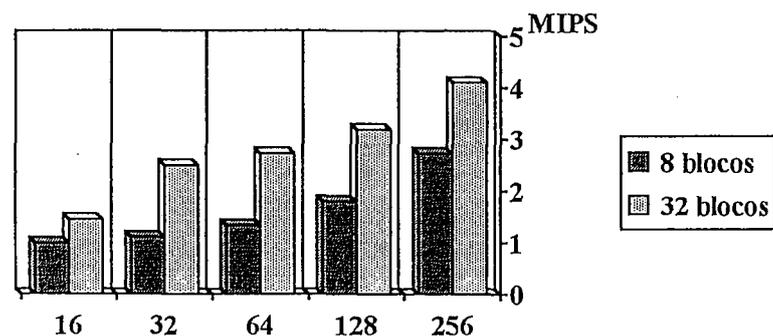


FIGURA 4.27: Carga computacional versus número de blocos

Essas limitações aparecem devido ao uso de um fator de convergência igual e constante, usado para que seja possível fazer comparações. Na realidade, a adoção de um fator de convergência variável com a potência do sinal de entrada do filtro adaptativo aumenta sensivelmente a atenuação absoluta e a velocidade de convergência, com um pequeno

acréscimo na carga computacional, fazendo com que a performance do cancelador com um menor número de blocos se aproxime bastante da performance do cancelador LMS normal.

4.5.2 Diferença de atenuação para sinais de voz e ruído

Foi constatado pelas medidas feitas que os níveis de atenuação obtidos são dependentes do tipo de sinal envolvido, podendo haver uma grande diferença entre esses níveis para um filtro com o mesmo número de coeficientes. Com o objetivo de minimizar estas diferenças foram tomadas as médias aritméticas das atenuações máximas dos filtros de mesmo número de coeficientes, tanto para o caso síncrono como assíncrono. Os resultados podem ser vistos nas figuras 4.28 e 4.29.

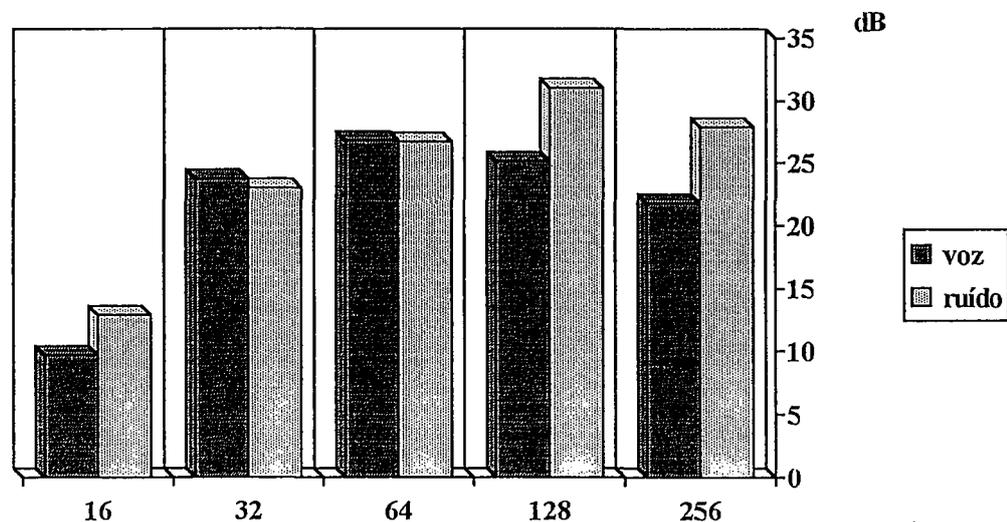


FIGURA 4.28: Médias das atenuações dos sinais de voz e ruído para módulos síncronos e sem compressão

Nota-se uma diferença marcante no desempenho dos canceladores de 16, 128 e 256 coeficientes. O uso de médias diminui a influência dos erros de medição, embora estes devam ser considerados. No entanto, a marcante diferença de atenuação parece indicar que não existe equivalência confiável no grau de atenuação medido com sinais de voz e com ruído. O comportamento de saturação da atenuação máxima para ruído, no entanto, acontece para o filtro de 128 coeficientes, ao invés de 64, como no caso de sinais de voz. A origem dessa diferença parece estar ligada ao próprio tipo de sinal. Esta diferença pode influir na medição do grau de desempenho do cancelador. Conforme visto no capítulo 2, a base teórica para o desenvolvimento de filtros adaptativos faz considerações com base em sinais estacionários. A aplicação de sinais não-estacionários, como sinais de voz, torna o desempenho destas estruturas inferior.

No presente caso, se tivesse sido levada em conta somente a atenuação para ruído, teria sido escolhida a performance de um cancelador de 128 coeficientes como a melhor. Além disso, a performance real, com sinais de voz, seria bem inferior à medida. Cabe salientar neste momento que testes de performance de canceladores de eco, segundo a CCITT [20], é feita utilizando-se de ruído branco limitado em frequência (300-3400 Hz). O uso de sinais mais representativos de sinal de voz real está em estudo [20].

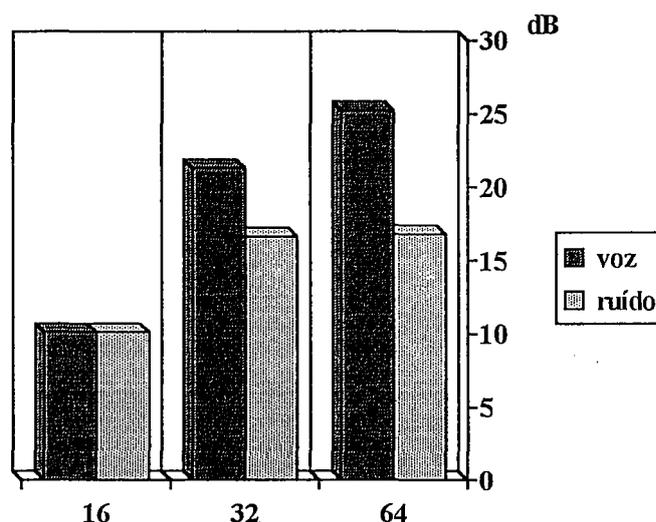


FIGURA 4.29: Médias das atenuações máximas para módulos assíncronos e com compressão

Para o caso de módulos assíncronos e com compressão, figura 4.29, percebe-se que a atenuação do sinal de voz é superior ao ruído. Na verdade, o que ocorre é que a atenuação máxima obtida para ruído foi diminuída sensivelmente. A origem deste comportamento está no algoritmo de compressão de voz, que interliga os módulos, já que o mesmo é construído para funcionar com sinais de voz e não com ruído. A aplicação de ruído parece introduzir perturbações no sinal regenerado que impedem o melhor funcionamento do sistema, estabelecendo o limite de cerca de 16 dB para a atenuação máxima.

Neste caso, novamente, a aplicação de ruído como teste de desempenho do sistema de cancelamento de eco produziria resultados errôneos, inferiores aos realmente alcançados com sinais de voz.

4.6 Conclusões

Neste capítulo foi estudado o problema de cancelamento de eco em linha telefônica através de um caso similar, que surgiu no desenvolvimento de um projeto que se utilizava de um módulo de compressão de voz. O problema específico e sua relação com o problema mais geral em telefonia foram situados.

O estudo de canceladores de eco sujeitos às restrições impostas, quais sejam, de o cancelador de eco utilizar pequena carga computacional e obter performance adequada mesmo sujeito ao problema de perda de amostras, mostrou que a atenuação do módulo é limitada mais pelo efeito de perda de amostra do que propriamente pela limitação no número de coeficientes. O uso de um esquema de adaptação em blocos também pode influir, apesar do uso de um esquema de variação de μ com a potência do sinal de entrada minimizar seu efeito.

O fator tolerância do usuário ao eco, dependente de fatores subjetivos, foi avaliado através de medidas realizadas com um grupo de 19 pessoas. Foi notada que uma atenuação de eco de cerca de 23 a 25 dB parece ser satisfatória para o uso. Outro resultado é que atenuações maiores não foram percebidas diferentes (até cerca de 30 dB). Dependendo deste nível de tolerância poderia ser usado o cancelador de eco sem o supressor [20].

Com base nestas medidas, pode-se concluir que neste projeto, sem mudança do DSP, o melhor cancelador de eco deve prover uma atenuação de 25 dB para o sinal de voz e possuir uma estratégia de adaptação em blocos com fator de convergência variável de acordo com a potência do sinal. Estes dados levam a uma implementação de um cancelador de 64 coeficientes no MCV com perda de amostra, utilizando o algoritmo *block update* LMS para que seja possível sua implementação.

A implementação de outras estruturas de filtros foi analisada do ponto de vista experimental. Implementações de filtros IIR, embora com possibilidades de funcionar melhor, não funcionaram. As causas podem ser diversas, como sensibilidades numéricas, mas principalmente a presença de pólos pouco amortecidos e de baixa frequência na função de transferência do conjunto "híbrida+telefone", que impedem a convergência do algoritmo. Esses problemas têm sido apontados em alguns trabalhos [9] [10]. Filtros *lattice* não puderam ser implementados devido à elevada carga computacional. Esquemas de simplificação prejudicam demasiadamente o desempenho do cancelador. Algoritmos *fast* RLS não puderam ser implementados pela maior carga computacional que o próprio LMS. Neste caso não foram tentadas simplificações.

Foi observado o problema de pouca velocidade de convergência nas medidas subjetivas. Este problema só pode ser melhorado com o uso de um algoritmo mais rápido que o algoritmo usado. Poderia valer a pena tentar uma estratégia de adaptação em blocos para implementar um cancelador de mínimos quadrados, embora não se tenha conhecimento de tal implementação em publicações.

Se fosse possível obter uma maior atenuação seria possível então continuar a implantação de um cancelador com supressor de eco residual, e finalmente submeter o cancelador aos testes da CCITT de performance. No entanto, existem problemas com medidas de performance usando ruído como sinal de teste, tanto para canceladores normais como para o MCV com compressão. Esse efeito foi descrito e suas prováveis causas foram indicadas.

Para um cancelador melhor, seria interessante avaliar a possibilidade econômica de troca do DSP por outro que possibilite a implementação de algoritmos de maior carga computacional (com conjunto de instruções melhores (*repeat block*), mais rápido, etc).

Também foi possível avaliar na prática os efeitos de variação de μ e do número de blocos no desempenho do cancelador. Observou-se que o aumento do número de coeficientes do filtro usado no cancelador tende a aumentar o nível de atenuação máxima, embora este aumento não seja linear devido ao aumento de outros erros. O uso de uma estratégia de adaptação em blocos mostrou prejudicar a performance face ao algoritmo LMS normal, tanto em termos de atenuação máxima como de tempo de convergência. O fato de usar um fator de convergência variável com a potência diminui drasticamente estes efeitos.

5 CONCLUSÃO

O interesse em processamento de sinais é uma decorrência natural da Instrumentação, pois a característica de transformação de uma grandeza física em um sinal que tenha utilidade está presente em cada instrumento de medição. Com o desenvolvimento da tecnologia de microeletrônica, a possibilidade de uso de técnicas de processamento digital de sinais tornou-se uma realidade. Essa tecnologia permite a obtenção de soluções para problemas que, de forma analógica tradicional, seriam dificilmente resolvidos de forma eficiente.

Uma técnica deste tipo foi apresentada neste trabalho, baseada principalmente em conceitos de processamento digital de sinais. O chamado "cancelador adaptativo de ruído" possibilita eliminar uma interferência conhecida de um sinal qualquer, desde que seja possível fornecer ao sistema uma estimativa (correlacionada) para esta interferência.

Foi salientada a importância deste sistema na abordagem do problema de entradas interferentes de um sistema de medição. Seu funcionamento geral foi descrito. Percebe-se que este tipo de sistema faz uso de um dispositivo chamado "filtro adaptativo". Um filtro adaptativo possui uma resposta ao impulso que é modificada no tempo para minimizar uma função de custo. A minimização desta função de custo é alcançada através de uma mudança dos parâmetros internos do filtro (coeficientes) a cada iteração de um algoritmo. O algoritmo define de que maneira a minimização da função de custo é feita em função da estrutura particular do filtro digital.

Como o tópico de filtros adaptativos deve ser estudado para que se tenha um bom entendimento do funcionamento do "cancelador adaptativo de ruído", decidiu-se atacar o problema de análise teórica destes sistemas. No entanto, devido ao número elevado de pré-requisitos, foi necessário estudar cada tópico com certo cuidado. O interesse maior consistiu em fornecer uma referência para o estudo inicial dos tópicos envolvidos.

Deste modo, foi analisado o problema de filtragem ótima linear e conceitos de probabilidade e estatística, que são úteis na análise e modelagem de sinais. O problema de filtragem ótima vai apresentar como solução a mesma solução que é esperada que o filtro adaptativo atinja após o período de adaptação. A solução ótima pode ser encontrada por dois caminhos equivalentes, seja através de métodos de descida pelo gradiente (*gradient descent*) ou por mínimos quadrados (*least-squares*). Estes dois métodos originam as principais famílias de algoritmos para filtros adaptativos de resposta ao impulso de duração finita (FIR).

Com base nestes conceitos é feito o estudo mais profundo do cancelador adaptativo de ruído, seguindo a mesma abordagem de Widrow [46].

Filtros adaptativos são estudados com maior profundidade. São apresentados os algoritmos principais para filtros de resposta ao impulso de duração finita (FIR), tanto de estrutura direta quanto de estrutura *lattice*. Algoritmos para filtros adaptativos de resposta ao impulso de duração infinita (IIR), bem como algumas vantagens e problemas, são mostrados.

A possibilidade de emprego de praticamente qualquer destas estruturas em um cancelador adaptativo de ruído torna o sistema muito maleável. O possível emprego de um filtro adaptativo em especial vai depender das características do problema a ser enfrentado e restrições impostas na sua solução. A tarefa do projetista, neste caso, é saber escolher qual o melhor caminho a ser tomado entre os tantos existentes, ou seja, é saber escolher qual a estrutura e o algoritmo ideais para a presente situação.

Foi abordado o problema de cancelamento de eco em linha telefônica, que surgiu em decorrência de um projeto entre a empresa Digital e o Laboratório de Processamento de Sinais do Depto. de Eng^a. Elétrica da UFRGS. A situação específica do problema foi mostrada e sua semelhança com o problema mais geral em telefonia foi salientada.

Este problema consistiu na implementação de um sistema de cancelamento de eco em um equipamento que já possuía implementados sistemas para compressão e comunicação de sinais de voz. Todo este sistema foi baseado no microprocessador TMS320C25, projetado para aplicações de processamento digital de sinais. Devido a ser esta uma implementação em conjunto, o número de instruções disponíveis para a implementação do cancelador de eco era limitado. Outro problema foi o uso de memória externa ao processador pelo algoritmo que implementou o cancelador de eco, fato que impede o uso de algumas instruções especiais e mais rápidas.

Outra característica que foi salientada foi o problema de perda de amostras entre os dois módulos, devido a problemas de falta de sincronismo. Esta perda de amostra não representa um problema para a reconstrução do sinal de voz, mas foi mostrado através de simulações que ela perturba o funcionamento do sistema de cancelamento de eco.

Estas limitações levaram ao uso de filtros adaptativos de pequena carga computacional, como filtros FIR de poucos coeficientes e algoritmos de busca pelo gradiente. O algoritmo LMS, no entanto, mostrou-se muito dispendioso para a implementação proposta e não pode ser possível implementar filtros com mais de 16 coeficientes. Assim, foi adotada uma estratégia de adaptação dos coeficientes em blocos, ao invés de adaptar todos os coeficientes do filtro de uma só vez como ocorre normalmente. Esta abordagem permitiu a implementação de filtros de até 64 coeficientes.

Com o objetivo de estudar o desempenho destes sistemas de cancelamento e isolar os problemas de carga computacional e de falta de sincronismo, foi feita a modificação temporária da implementação dos módulos, tornando-os síncronos e eliminando o código de compressão

de voz. Dessa maneira pode-se fazer medidas de atenuação máxima para filtros adaptativos de maior comprimento e também de outras estruturas.

As medidas de atenuação máxima foram feitas para os módulos na situação anterior e na situação de *hardware* modificado (sem compressão, módulos sincronizados). Foi observado por estas medidas que a atenuação máxima atingida para sinais de voz era inferior na situação de funcionamento normal dos módulos, para os mesmos sistemas de cancelamento. Foi notado que o efeito de perda de amostra, ao introduzir perturbações no sistema, pode afetar o desempenho de atenuação máxima do sinal no modo com que esta foi medida. Distorções introduzidas pelo algoritmo de codificação também podem ter influenciado, mas não foi feito um estudo mais detalhado.

O uso de filtros maiores mostrou aumentar o nível de atenuação, de modo não equivalente, mas que atinge um limiar em função dos erros numéricos e efeitos advindos do número excessivo de coeficientes. Foi notado que de 64 a 128 coeficientes está o número que melhor aproxima a parte mais significativa da resposta ao impulso do conjunto "híbrida+telefone" para a taxa de amostragem utilizada (7400 Hz).

A característica de atenuação máxima foi mostrada ser dependente do tipo de sinal, voz ou ruído. Esta característica pode afetar a medida de desempenho de sistemas de cancelamento, já que os testes atualmente empregados pela recomendação da CCITT [20] sobre canceladores de eco emprega ruído branco limitado em frequência como sinal padrão. Particularmente, no sistema implementado a atenuação de ruído mostrou ser bastante inferior à atenuação obtida por sinais de voz. Este efeito é atribuído ao algoritmo de compressão de voz, que não codifica bem sinais aleatórios. As medidas feitas usando o equipamento modificado indicam que, em geral, a atenuação de ruído é maior que a atenuação para sinais de voz.

Foram realizadas mais medidas para avaliar o efeito de mudanças de diversos parâmetros do algoritmo *block update* LMS. Em geral, o aumento do fator de convergência do algoritmo, na faixa de estabilidade, mostrou aumentar a atenuação máxima atingida. O efeito de adaptar os coeficientes do filtro em blocos prejudica bastante a velocidade de convergência. O aumento do número de coeficientes, conforme já dito, aumenta a atenuação máxima até um certo limiar, a partir do qual esta tende a diminuir. O uso de um fator de convergência variável com a potência do sinal de entrada do filtro mostrou aumentar bastante a velocidade de convergência do algoritmo, sendo útil no caso de adaptação dos coeficientes em blocos.

O uso de outras estruturas de filtros adaptativos para o problema foi tentado. Foi implementada uma versão simplificada do algoritmo LMS para filtros *lattice*. No entanto, a elevada carga computacional desta estrutura impediu um desempenho satisfatório. A simplificação usada, qual seja, de manter um coeficiente de adaptação igual e constante para os estágios *lattice* e *ladder*, não levou a bons resultados e mesmo assim, só permitiu a implementação de um filtro com 15 estágios. Os níveis de atenuação atingidos foram inferiores aos de um filtro FIR equivalente (16 coeficientes).

Também foi feita uma implementação de filtros IIR. Esta implementação também não mostrou resultados satisfatórios, apesar de o sistema ser capaz de implementar uma função de transferência de ordem superior à mínima, descoberta através da identificação da função de transferência do conjunto "híbrida + telefone". Este fato decorre de problemas que sofrem os filtros adaptativos IIR, inclusive de velocidade de convergência prejudicada na situação em que o filtro deve formar pólos de baixa frequência e pouco amortecidos, como neste caso. Outro problema decorrentes de implementação, como o uso de precisão finita na representação dos coeficientes do filtro, também pode ter contribuído para o fraco desempenho, inferior ao filtro FIR de 16 coeficientes.

Foi estudado o desempenho subjetivo dos sistemas de cancelamento com filtros FIR e algoritmos de adaptação em blocos. Para tanto foi feita uma pesquisa com um grupo de 19 pessoas, que mostrou que o nível de atenuação atingido pelos canceladores de eco com 32 coeficientes ou mais (equipamento sem compressão, sincronizado) ou de cerca de 25 dB, parece ser bom para uma conversa telefônica. Não foi possível a distinção de desempenho entre os canceladores de 32 ou mais coeficientes. O fato de ser difícil manter uma conversa sem o sistema de cancelamento foi notado. Outro resultado é que o sistema de cancelamento demora para atingir uma boa atenuação, e deve ser melhorado.

Neste trabalho foi feito o estudo de uma técnica para separação de sinais utilizando filtros adaptativos. Mostrou-se que essa técnica tem um grande número de aplicações e que o problema de cancelamento de eco é um destes casos. Para o problema específico enfrentado, a solução proposta permitiu o uso do sistema que, de outro modo, estaria inviabilizado pela presença de eco. No entanto, o desempenho do sistema de cancelamento ainda precisa ser melhorado, especialmente por problemas de velocidade de convergência. O uso de processadores mais rápidos e de outros algoritmos, como os baseados no método de mínimos quadrados, são propostas para o aperfeiçoamento do sistema que devem ser melhor pesquisadas.

Sistemas de cancelamento de eco estão presentes no mundo atual em cada ligação a longa distância que fazemos. Os requisitos destes sistemas são justamente fazer com que não percebamos a sua existência. Técnicas de processamento digital de sinais é que tornam possível esta realidade.

ANEXO A - IDENTIFICAÇÃO DA FUNÇÃO DE TRANSFERÊNCIA DO CONJUNTO "HÍBRIDA+TELEFONE"

A.1 Introdução

Este anexo busca apresentar com mais detalhes o trabalho desenvolvido com o objetivo de identificar a função de transferência do conjunto "híbrida+telefone". Conforme visto no capítulo 4, a função de transferência do filtro adaptativo deve ser o mais próximo possível da função de transferência do conjunto "híbrida+telefone". Fala-se em "conjunto" porque a função que age sobre o sinal de eco depende tanto do circuito da híbrida quanto dos circuitos internos do aparelho telefônico conectado. A figura A.1, reproduzida do capítulo 4, ilustra o modelo no domínio digital.

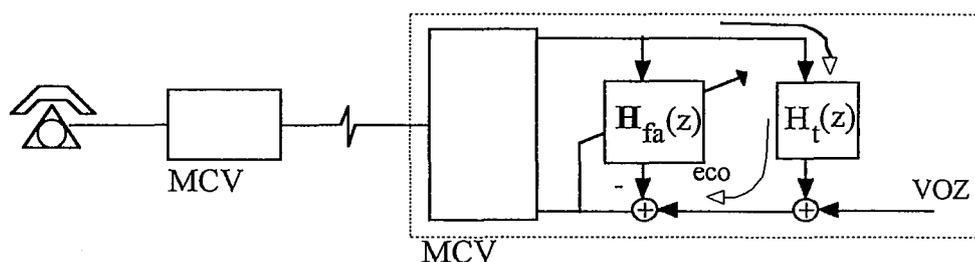


FIGURA A.1: Funções de transferência em Z do filtro adaptativo e do conjunto "híbrida+telefone"

O levantamento da função de transferência foi baseado em um procedimento computacional, através da análise da resposta ao salto [28]. Este procedimento busca adaptar um suposto modelo com determinado número de pólos e zeros à resposta obtida, dentro de uma faixa de frequências determinada. Para tanto é feito o cálculo numérico da transformada de Fourier para uma faixa de frequências e delta f escolhidos (programa IDENT1.EXE). Após, é ajustada a resposta do modelo sugerido pelo usuário a esta resposta através do método dos mínimos quadrados (programa IDENT2.EXE). Estes programas foram desenvolvidos durante o curso de graduação em Eng^a. Elétrica da UFRGS, na disciplina de Controle.

Este levantamento não foi muito importante no caso de filtros FIR, porque o importante é saber a duração da resposta ao impulso ou a duração do transiente na resposta ao salto. No entanto, a digitalização da resposta ao impulso foi necessária para que este atraso pudesse ser estimado. Para o caso de filtros IIR esse levantamento foi essencial, pois permitiu estimar a ordem mínima dos filtros.

O procedimento seguido para a identificação consistiu nas seguintes etapas:

- i) preparação do módulo para o ensaio;
- ii) ensaio e aquisição sob diversas taxas de amostragem;
- iii) aplicação do algoritmo para obter a função de transferência em S ;
- iv) teste do modelo através de simulação;
- v) transformação do modelo em S para o domínio Z .

A.2 Preparação do módulo

Para que fosse possível submeter o circuito do conjunto "híbrida+telefone" ao ensaio, seria necessário identificar os melhores pontos para a aplicação do sinal. Decidiu-se que o melhor método seria aplicar um salto de tensão diretamente na saída do conversor D/A, com o mesmo retirado do circuito, e digitalizar o sinal recebido na entrada do conversor A/D.

Para que fosse eliminado o efeito provocado por som ambiente captado pelo microfone do aparelho telefônico, este foi substituído por um resistor de valor adequado (330Ω), que mantivesse a resposta ao salto do conjunto semelhante e livre da influência de ruídos sonoros.

A.3 Ensaio

O ensaio foi realizado com o auxílio de um gerador de sinais, com uma onda quadrada de 1 Vpp e frequência de .8 Hz. Para a digitalização dos sinais foi utilizada uma placa de processamento digital de sinais e o módulo ADMUX [27], além de um amplificador para o sinal proveniente da entrada do A/D (saída do sistema "híbrida+telefone", cujos níveis de tensão estavam abaixo do ideal para a entrada no módulo ADMUX). A figura A.2 apresenta a aquisição de 3 segundos do ensaio, para uma taxa de 10 kHz. Note que as escalas dos dois sinais estão diferentes devido à amplificação necessária no sinal de saída de H_1 .

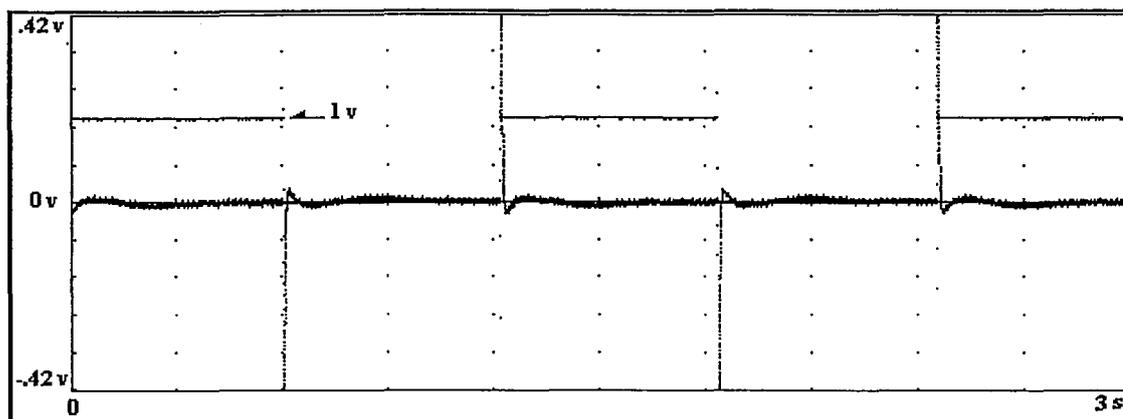


FIGURA A.2: Ensaio ao salto digitalizado

A.4 Aplicação do algoritmo e simulação dos modelos

Para a aplicação do algoritmo, foi feita a edição do arquivo de dados para que fosse registrada apenas uma das respostas. Foram feitas então diversas aplicações do algoritmo para tentar identificar a resposta que tivesse o comportamento o mais próximo possível da resposta observada. O número de pólos e zeros do modelo foi aumentado seqüencialmente.

Para os arquivos adquiridos a uma taxa de 10 kHz, foi possível aumentar o grau de aproximação do modelo até 6. No entanto, a partir da 3ª ordem começam a se cancelar os pares de pólos e zeros. Dois pólos e dois zeros nunca se cancelam. No entanto, a aproximação do segundo grau não é boa e foi descartada.

A função escolhida possui 3 pólos e 3 zeros e utilizou o arquivo do ensaio amostrado a uma taxa de 40 kHz, com a resposta truncada em 100 ms de duração.

Foi utilizado o programa IDENT1.EXE com os parâmetros $w_i = .1/40k = 2.5E-6$, $w_f = 30000/40k = 750E-3$, $dw = 30/40k = 750E-6$, $t_0 = 0$ e $dt = 1$, gerando os arquivos hr40k100.dat, hi40k100.dat e w40k100.dat. O programa IDENT2.EXE, utilizando estes arquivos e com ordem de aproximação 3 (grau 3), obteve como resultado (arquivo o40k1003.dat), a seguinte função de transferência em S:

$$H(s) = \frac{-.125306s^3 + .006883s^2 + .000639s}{s^3 + .152771s^2 + .002597s + .000021} \quad (A.1)$$

Esta função de transferência apresenta pólos e zeros que devem ser desnormalizados através da multiplicação de uma constante igual à frequência de amostragem. A partir dessa ordem ocorre o cancelamento de pólos e zeros. Os pólos e zeros encontrados devem ser desnormalizados pela multiplicação por 40000, obtendo:

pólos: $p_1 = -5386$, $p_{2,3} = -363 \pm j344$

zeros: $z_1 = 0$, $z_2 = 4160$, $z_3 = -1962$

ganho: -0.12

O ganho deve ser corrigido para representar a tensão correta do sinal. No caso, utilizando os valores médios de correção, chega-se a -0.3 .

A função de transferência desnormalizada é dada por:

$$H(s) = -0.3 \frac{s^3 - 2198s^2 - 8161920s}{s^3 + 6112s^2 + 4160341s + 1347065530} = -0.3 \frac{s(s-4160)(s+1962)}{(s+5386)[(s+363)^2 + 344^2]} \quad (\text{A.2})$$

A transformada Z de $H(s)$ deve ser tomada lembrando a existência do conversor DA, cujo modelo usual é o Retentor de Ordem Zero, representado por $(1 - \exp(-Ts))/s$. A transformada Z que deve ser calculada é, portanto

$$H(z) = Z \left\{ \frac{1 - e^{-Ts}}{s} H(s) \right\} = \frac{z-1}{z} Z \left\{ \frac{H(s)}{s} \right\} \quad (\text{A.3})$$

A expressão em Z necessita do período de amostragem do sinal para a representação. Usando a taxa de amostragem do MCV normal (com algoritmo de compressão e assíncrono), ou seja, para $T_s = 1.354166e-4$ s ($f_s = 7384.62$ Hz), tem-se a seguinte função em Z:

$$H(z) = -0.3 \frac{z^3 - 3.2271z^2 + 3.3545z - 1.1273}{z^3 - 2.3842z^2 + 1.8235z - .4371} = -0.3 \frac{(z-1)(z-1.4493)(z-.7778)}{(z-.4822)[(z-.9510)^2 + .0443^2]} \quad (\text{A.4})$$

Nas figuras a seguir são mostradas as curvas de resposta ao salto adquirida e as simulações feitas utilizando os resultados em S e em Z, para um tempo de 50 ms.

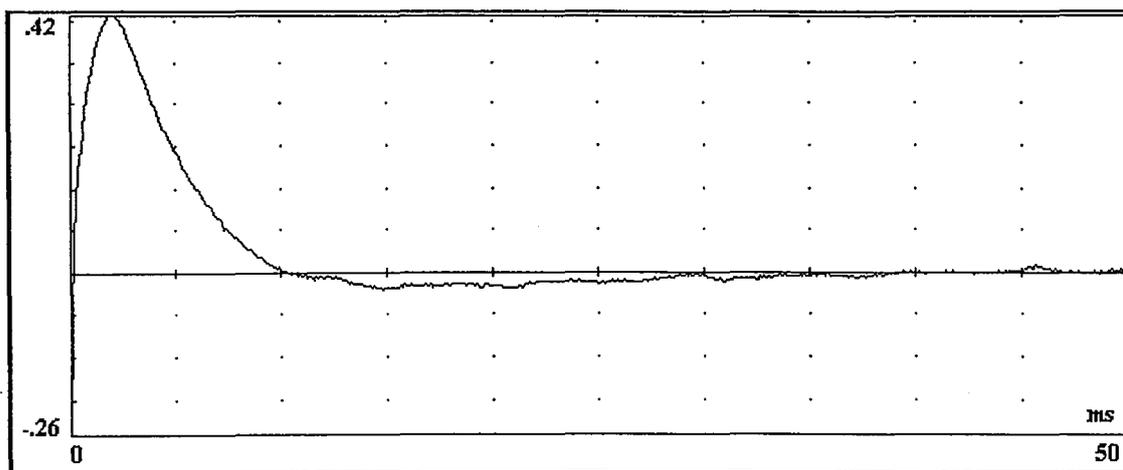


FIGURA A.3: Resposta ao salto digitalizada

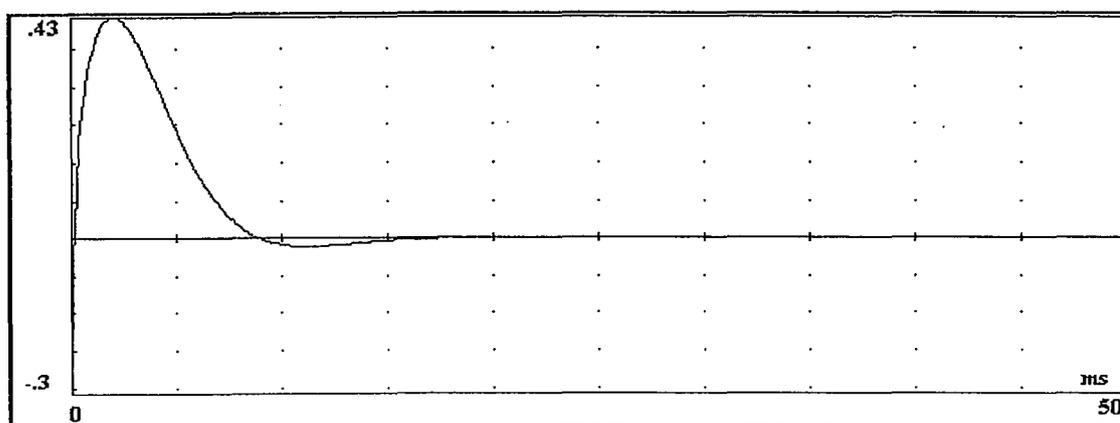


FIGURA A.4: Resposta ao salto simulada em S

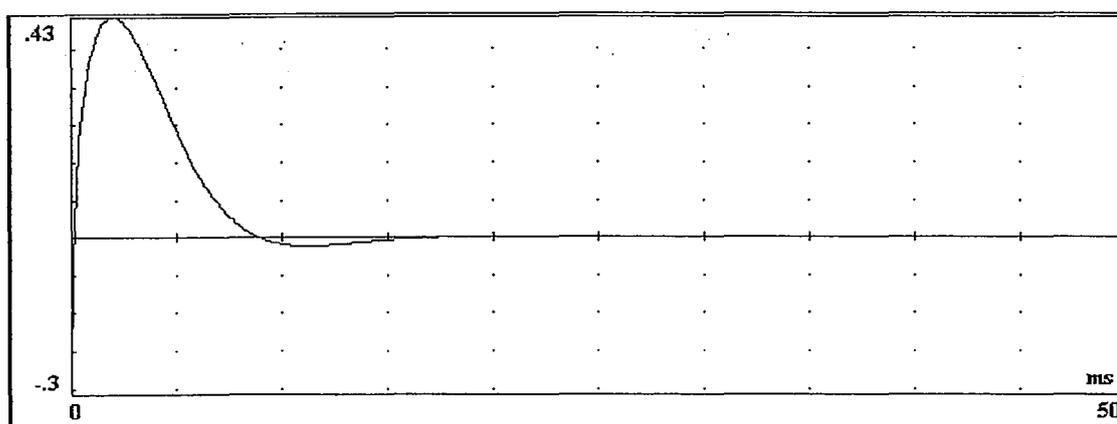


FIGURA A.5: Resposta ao salto simulada em Z

ANEXO B - MEDIDAS DE ATENUAÇÃO SUBJETIVAS

B.1 Objetivos

Conforme discutido no capítulo 4, foi necessário avaliar o caráter subjetivo de tolerância do locutor ao problema de eco. Esta avaliação também foi feita sobre o desempenho do sistema de cancelamento de eco. Deste modo, pode-se avaliar quais os níveis de atenuação de eco que podem ser considerados como "mínimos" para a utilização em sistemas de cancelamento, qual a melhor relação desempenho x carga computacional na aplicação do MCV e como é sentida a velocidade de convergência do cancelador.

B.2 Metodologia de teste

Os testes subjetivos foram conduzidos do seguinte modo: o sistema com dois módulos MCV era interligado e o telefone do outro terminal era colocado fora do gancho e num lugar de pouco ruído ambiente. O outro telefone ficava com a pessoa que faria a avaliação. Foram apresentados 4 canceladores de eco para cada pessoa, todos em seqüência crescente de atenuação (a pessoa não sabia qual era a ordem de atenuação). Em cada apresentação, a pessoa era convidada a responder as seguintes perguntas, com as respostas especificadas:

Para o filtro já adaptado (cancelador em situação de máxima atenuação de eco):

1) Como lhe parece o nível de eco?

muito alto (MA), alto (A), baixo (B), muito baixo (MB), baixíssimo (MMB)

2) Como lhe parece o nível de eco em relação ao filtro que foi mostrado antes?

muito melhor (MM), pouco melhor (PM), igual (I), pouco pior (PP), muito pior (MP)

Para o filtro não adaptado previamente (cancelador ainda não atingiu a maior atenuação):

3) Como lhe parece o nível de eco em relação ao caso adaptado?

muito melhor (MM), pouco melhor (PM), igual (I), pouco pior (PP), muito pior (MP)

4) Qual a impressão geral para o caso já adaptado?

o eco atrapalha/ não atrapalha (R/B) a conversação

5) Nota (0-10) para o caso já adaptado?

nível de eco sem cancelador recebe nota 0 e nível de eco inaudível recebe nota 10

A utilização do filtro previamente adaptado foi feita para separar o problema de atenuação máxima do problema de velocidade de convergência. O processo de adaptação era feito submetendo o cancelador a cerca de 5 s de ruído branco gerado com o auxílio da placa de DSP [27], de mesmo modo descrito no anexo C.

Durante o teste, a pessoa podia falar o que quisesse. Eventualmente outra pessoa falava no telefone B para que fosse sentida uma maior semelhança com um telefone comum. Isso ocorria sob pedido da pessoa. Os módulos possibilitavam a reinicialização, habilitação e desabilitação do cancelador de eco.

B.3 Metodologia de análise

Foram feitas medidas para um grupo de 19 pessoas, cujos resultados estão colocados na tabela B.3, em função da pergunta e do número do cancelador apresentado. Um quadro resumido com os resultados para as perguntas 1, 2 e 3 é dado na figura B.1.

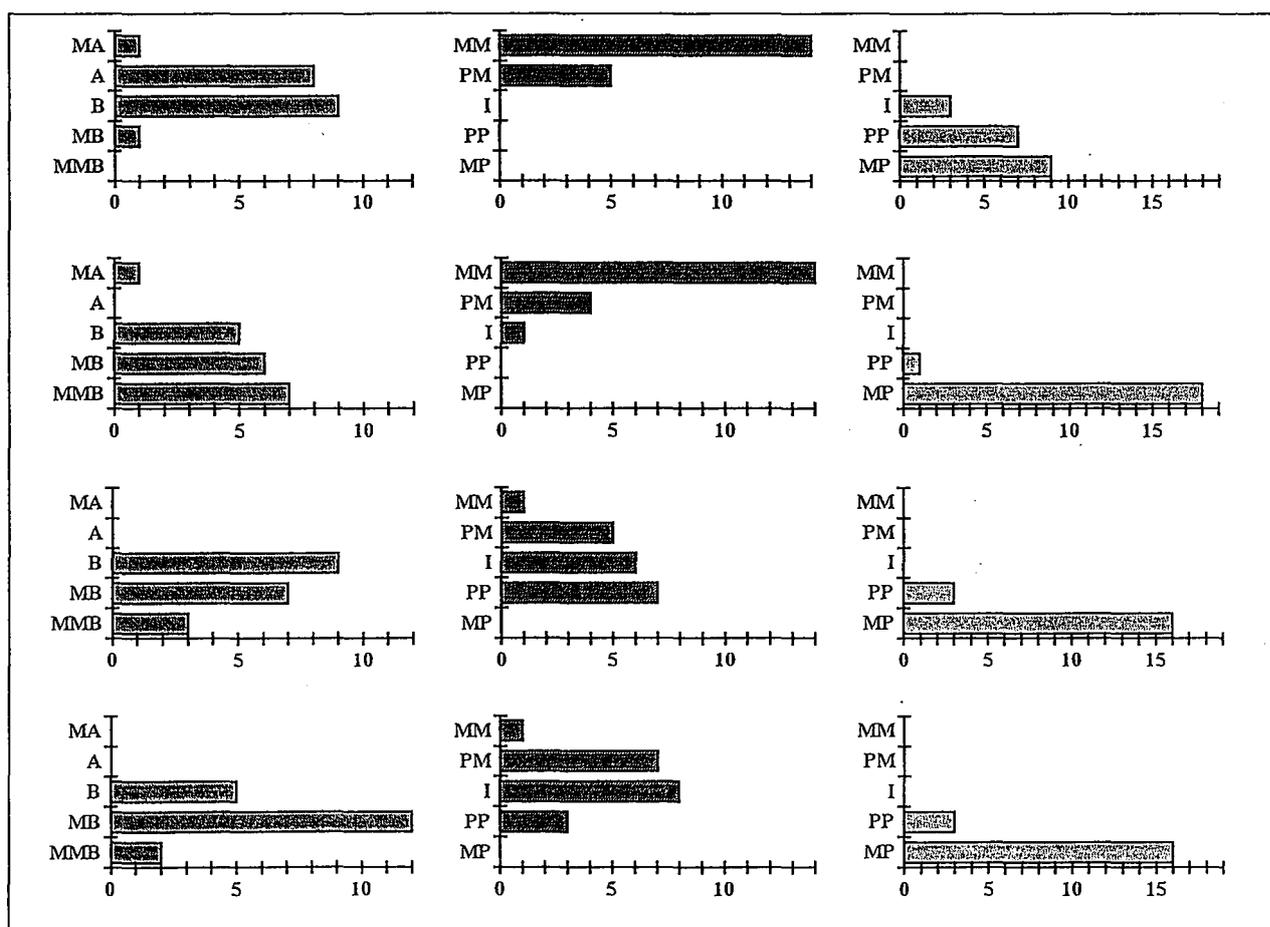


FIGURA B.1: Respostas para as questões 1, 2 e 3, para os filtros 1 (superior), 2, 3 e 4 (inferior)

Nesta figura, o número de respostas para as perguntas 1, 2 e 3 e para o cancelador de 16 coeficientes estão na primeira linha. Os demais resultados (canceladores c/ filtro de 32, 64 e 128 coeficientes) estão colocados nas linhas inferiores. Estes gráficos permitem avaliar comparativamente o desempenho dos diversos canceladores.

Para o caso da impressão geral do filtro (pergunta 4), a tabela B.1 ilustra o número de respostas obtidas para cada cancelador de eco apresentado. A análise é simples e mostra que, para o caso 1 (filtro de 16 coeficientes), não houve consenso se o nível de eco atrapalha ou não a conversação. Já para os demais casos, os níveis de atenuação obtidos parecem não atrapalhar uma conversação telefônica, na opinião do grupo.

TABELA B.1: Resultados para a pergunta 4 em termos do número de respostas

CASO	1	2	3	4
adequado	10	18	19	19
não adequado	9	1	0	0

A análise das medidas subjetivas (perguntas 1 e 2) foi dificultada devido ao caráter subjetivo das mesmas. Para que esse problema fosse resolvido, foi feito um mapeamento de conceitos para números, para que se pudesse fazer uma análise numérica desses dados. Esse mapeamento obedeceu ao critério mostrado na tabela B.2.

TABELA B.2: Mapeamento de conceitos para números

MA	A	B	MB	MMB
MP	PP	I	PM	MM
1	3	5	7	9

A análise das notas (pergunta 5) foi feita de acordo com o mesmo procedimento seguido para a análise das perguntas subjetivas.

O procedimento de análise foi utilizar programas para a comparação múltipla de médias. Estes programas (projex.exe, compvar.exe,..) são utilizados em disciplina do curso de Pós-Graduação em Engenharia Civil da UFRGS, com apostila específica [32].

B.4 Resultados

Para o nível de eco absoluto (pergunta 1), o primeiro caso (cancelador de eco com filtro de 16 coeficientes) teve nível de eco considerado como médio (entre alto e baixo). Os demais casos (16, 32 e 64 coeficientes) tiveram nível de eco considerado muito baixo. Não pôde ser comprovada diferença entre os três últimos algoritmos pela análise das médias.

Os casos 1 e 2 tiveram um desempenho considerado como muito melhor em relação aos casos anteriores, ou seja, o cancelador com 16 coeficientes é "muito melhor" que nenhum cancelador, e o cancelador com 32 coeficientes é "muito melhor" que o cancelador de 16 coeficientes. Os canceladores de 64 e 128 coeficientes tiveram desempenho considerado igual aos anteriores. Essas medidas indicam que não foi percebida diferença entre os canceladores de 32, 64 e 128 coeficientes, confirmando o resultado da análise da pergunta 1.

Em relação às notas (pergunta 5), o caso 1 recebeu a menor nota de todos, ao passo que as notas dos outros casos não diferiram significativamente entre si, mas foram maiores que para o caso 1.

Esta análise subjetiva (perguntas 1, 2 e 5) indica que o grupo analisado de 19 pessoas não notou diferença significativa no desempenho de um cancelador de eco de 32 coeficientes ou mais.

A análise da pergunta 4 leva ao mesmo resultado, indicando que este nível de atenuação (aprox. 24 dB) poderia ser adequado ao uso. Já a análise da pergunta 3 indica que o desempenho dinâmico de todos os casos deve ser melhorado, pois os casos que alcançam os níveis de atenuação mais altos demoram para atingir esta situação.

Deve ser salientado que estas medidas foram feitas com um grupo pequeno (19 pessoas) e tem um caráter informal. O fato de não ter sido feita uma distinção entre os casos de atenuação maior indica que deveriam ser feitos testes mais sensíveis. O uso de uma situação não usual (telefone na outra ponta em silêncio e pessoa esperando ouvir o sinal de eco) também pode contribuir para uma impressão negativa do sistema.

ANEXO C - MEDIDAS DE ATENUAÇÃO ABSOLUTAS DO SINAL DE ECO

C.1 Objetivos

As medidas de atenuação absolutas do sinal de eco tem por objetivo comparar o desempenho dos diversos filtros adaptativos usados no cancelador de eco. Também são úteis para avaliar a sensibilidade dos algoritmos a variações de seus parâmetros.

Foram feitas medidas de atenuação para filtros FIR utilizando o algoritmo *block update* LMS. Também foi medido o nível de atenuação para uma implementação simplificada de filtros *lattice* LMS. Para filtros IIR foi medida a atenuação com uma implementação dos algoritmos SHARF e *Recursive* LMS.

C.2 Metodologia

Para o filtro FIR com algoritmo *block update* LMS, que teve mais importância neste trabalho, foram feitas medidas visando avaliar o efeito da variação do número de coeficientes, do fator de convergência e do número de blocos sobre a atenuação do sinal de eco e sobre o tempo de convergência do cancelador.

As medidas do algoritmo *block update* LMS foram efetuadas usando os módulos MCV em operação normal e com *hardware* modificado (sem o algoritmo de compressão, módulos sincronizados), com os seguintes parâmetros:

- 1) μ , 2μ e $\mu/2$ ou μ variável (dependente de potência);
- 2) filtros com 16, 32 e 64 coeficientes (128 e 256 também, para o caso modificado);
- 3) 8 blocos, para apenas um fator de convergência (32 também, para o caso modificado);
- 4) sinais de voz (/aeiou/, /dois/) e ruído branco.

Foram feitas as seguintes medidas:

- a) tensão de *offset* em silêncio;
- b) tensão sem atenuação;
- c) tensão com atenuação estável ou após 3 minutos, no máximo;

Foi feita a reinicialização do cancelador de eco entre cada medida. O tempo de convergência do algoritmo foi avaliado para o caso de ruído branco. As medidas para os demais algoritmos tiveram procedimento semelhante, mas se restringiram aos casos mostrados.

C.3 Configuração usada para a medição de nível e tempo de atenuação.

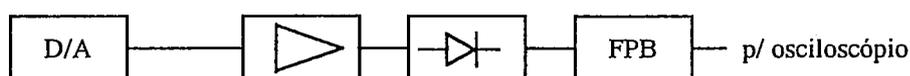


FIGURA C.1: Configuração usada para medição de parâmetros

Na figura C.1, D/A é a saída do conversor D/A do módulo MCV, que vai para o telefone e contém o sinal de eco que é escutado pelo locutor. Quando o cancelador de eco está acionado, este sinal sofre atenuação e esta é que será medida.

Como é difícil medir a atenuação de sinais de voz e ruído devido à variação sofrida na amplitude destes sinais, foi feito um condicionador de sinal que produz em sua saída um valor DC proporcional ao valor médio do sinal presente na saída do D/A. Este sinal é amplificado, retificado (onda completa) e passa por um filtro passa-baixas passivo.

Para manter o sinal de entrada para o sistema em níveis aceitáveis e que assegurassem sua repetibilidade para que fossem feitas comparações, foi utilizada uma placa de DSP [27] com o módulo ADSERIAL ligado a uma caixa de som, conforme ilustra a figura C.2. O telefone foi colocado próximo a esta caixa, e o nível sonoro corresponde ao de uma conversa normal. Os sons foram gravados e repetidos durante o tempo de medida.

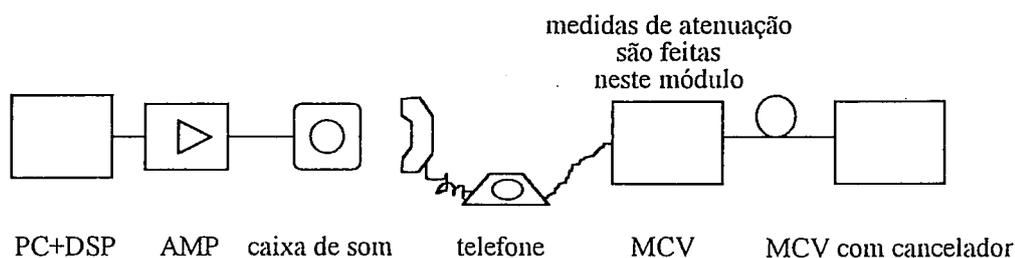


FIGURA C.2: Sistema usado para excitar o telefone

As medidas de atenuação máxima foram feitas utilizando sinais de voz (as vogais /aeiou/ e o número /dois/) e ruído branco, tendo sido feita a média da atenuação para sinal de voz. resultante destas duas medidas. Foram considerados para cada medida o valor do sinal sem o cancelador e o menor valor com o cancelador ligado por um máximo de 3 minutos. A constante de tempo usada no filtro passa-baixas de saída foi de $RC=15k\Omega \times 470\mu F$.

Para efetuar medidas dos tempos de atenuação foi mudada a constante de tempo do filtro para $RC=15k\Omega \times 470nF$, cuja influência na resposta chega a cerca de 2% em 28 ms, suficientemente rápido para as medidas em questão. Estas medidas foram feitas utilizando-se de um osciloscópio de retenção, onde os níveis limites correspondiam à máxima atenuação obtida pelo cancelador em uso e à atenuação com o cancelador desligado (ou seja, sem atenuação). Devido à pior visualização dos níveis do sinal de saída (sinal instável devido à menor constante de tempo do filtro de saída), estas medidas foram repetidas 3 vezes para cada caso, tendo sido tomada sua média.

A seguir são apresentadas as medidas efetuadas. Cabe ressaltar que os valores de atenuação são imprecisos, principalmente quando a atenuação é grande, devido a erros de medida de offset, da ordem de 4 a 5 mV, medidos com um osciloscópio e com nível de ruído sobre o sinal de 5 mV.

C.4 Resultados obtidos

A tabela C.1 apresenta os resultados das medidas de atenuação e velocidade de convergência feitas para o algoritmo *block update* LMS. São mostrados os resultados para os módulos com o hardware modificado e com funcionamento normal.

TABELA C.2: Medidas de atenuação para o algoritmo *lattice* LMS

nº estágios	15	15	15	15	15
aten. voz lin e dB	1.78 x 4.98 dB	1.89 x 5.51 dB	1.60 x 4.06 dB	1.86 x 5.37 dB	1.85 x 5.34 dB
aten. ruído lin e dB	3.47 x 10.82 dB	3.50 x 10.87 dB	1.62 x 4.16 dB	3.63 x 11.19 dB	3.58 x 11.07 dB
tempo 50%	133 ms	-	-	43 ms	50 ms
tempo 40%	188 ms	-	-	57 ms	67 ms
tempo 30%	243 ms	-	-	67 ms	85 ms
μ	1000	10000	100	5000	5000 máx
carga MIPS	4.84	4.84	4.84	4.84	4.94

Os resultados obtidos para o algoritmo *lattice* LMS simplificado e módulos MCV modificados (sem compressão e síncronos) estão colocados na tabela C.2. A simplificação consistiu em fazer a constante de adaptação dos coeficientes *lattice* e *ladder* iguais e constantes. Para o caso variável, o coeficiente era mudado de acordo com a potência do sinal de entrada do filtro uma vez por iteração, e permanecia o mesmo para todos os estágios. Note-se que a atenuação alcançada é inferior à do algoritmo LMS convencional de 16 coeficientes.

TABELA C.3: Medidas de atenuação dos algoritmos SHARF e *Recursive LMS*

	RecLMS	RecLMS	SHARF	SHARF
n° pólos	3	8	3	8
n° zeros	3	8	3	8
aten. voz lin e dB	1.52 x 3.63 dB	2.46 x 7.84 dB	1.43 x 3.11 dB	2.38 x 7.53 dB
aten. ruído lin e dB	1.11 x 0.91 dB	3.33 x 10.46 dB	1.13 x 1.04 dB	3.35 x 10.49 dB
c_1	0	0	-8	-8
c_2	0	0	0	0
μ	1000	1000	1000	1000
carga MIPS	1.04	1.91	1.13	1.99

Para os algoritmos IIR e módulos sem compressão, os resultados estão colocados na tabela C.3. Não foram feitas medidas de tempo, e o programa utilizou-se de memória externa (mais lenta). A implementação utilizou-se de notação de menor precisão para os coeficientes do filtro, devido à possibilidade de serem maiores que a unidade em módulo. O algoritmo SHARF apresenta c_1 diferente de zero, ao passo que o algoritmo *Recursive LMS* apresenta c_1 igual a zero. Novamente, a atenuação alcançada é inferior à alcançada pelo filtro FIR de 16 coeficientes, em ambos os casos.

Cabe ressaltar que as medidas de velocidade representam o instante em que a amplitude média do sinal de eco residual (retificado e filtrado) atingia o percentual especificado em relação à amplitude média do máximo eco (sem cancelador). Portanto, o tempo indicado no quadro para, por exemplo, 30%, representa o tempo que a amplitude média do eco leva para ser reduzida a 30%. Esse valor não corresponde ao valor exato de 70% da atenuação máxima do sinal de eco (especificada em termos de potências), mas pode ser usado como uma aproximação grosseira.

TABELA C.1: Medidas de atenuação para o algoritmo *block update* LMS

n° coef.	módulos sem compressão e síncronos					módulos normais		
	16	16	16	16	16	16	16	16
n° blocos	8	8	8	16	8	8	8	8
aten. voz lin e dB	2.62 x 8.37 dB	3.38 x 10.58 dB	3.08 x 9.76 dB	2.73 x 8.71 dB	3.44 x 10.72 dB	3.13 x 9.90 dB	3.19 x 10.06 dB	3.30 x 10.37 dB
aten. ruído lin e dB	4.82 x 13.66 dB	4.12 x 12.30 dB	4.12 x 12.30 dB	5.01 x 14.00 dB	4.09 x 12.23 dB	3.39 x 10.61 dB	2.92 x 9.29 dB	3.29 x 10.35 dB
tempo 50%	247 ms	137 ms	-	147 ms	63 ms	307 ms	87 ms	-
tempo 40%	340 ms	187 ms	-	193 ms	87 ms	460 ms	110 ms	-
tempo 30%	473 ms	273 ms	-	280 ms	127 ms	753 ms	130 ms	-
μ	1000	2000	500	1000	5000 máx	1000	5000 máx	2000
carga MIPS	.9936	.9936	.9936	1.454	1.145	1.425	1.580	1.425
n° coef.	32	32	32	32	32	32	32	32
n° blocos	8	8	8	32	8	8	8	8
aten. voz lin e dB	12.50 x 21.94 dB	17.94 x 25.08 dB	11.48 x 21.20 dB	17.24 x 21.20 dB	17.24 x 24.73 dB	10.86 x 20.72 dB	11.32 x 21.08 dB	12.74 x 22.10 dB
aten. ruído lin e dB	16.42 x 24.31 dB	14.48 x 23.21 dB	11.02 x 20.84 dB	15.15 x 23.61 dB	13.48 x 22.59 dB	5.67 x 15.07 dB	6.55 x 16.32 dB	8.06 x 18.13 dB
tempo 50%	600 ms	393 ms	-	160 ms	130 ms	580 ms	93 ms	-
tempo 40%	853 ms	527 ms	-	260 ms	173 ms	807 ms	150 ms	-
tempo 30%	1207 ms	713 ms	-	367 ms	260 ms	1167 ms	193 ms	-
μ	1000	2000	500	1000	5000 máx	1000	5000 máx	2000
carga MIPS	1.109	1.109	1.109	2.491	1.260	1.898	2.053	1.898
n° coef.	64	64	64	64	64	64	64	64
n° blocos	8	8	8	32	8	8	8	8
aten. voz lin e dB	17.43 x 24.82 dB	24.05 x 27.62 dB	9.19 x 19.26 dB	33.44 x 30.49 dB	25.50 x 28.13 dB	14.68 x 23.33 dB	20.96 x 26.43 dB	18.68 x 25.43 dB
aten. ruído lin e dB	17.17 x 24.70 dB	34.45 x 30.74 dB	9.76 x 19.79 dB	15.73 x 23.94 dB	31.21 x 29.89 dB	6.32 x 16.02 dB	7.43 x 17.42 dB	7.05 x 17.08 dB
tempo 50%	1367 ms	573 ms	-	273 ms	280 ms	1400 ms	247 ms	-
tempo 40%	1967 ms	840 ms	-	453 ms	400 ms	1750 ms	387 ms	-
tempo 30%	3033 ms	1213 ms	-	653 ms	620 ms	2700 ms	633 ms	-
μ	1000	2000	500	1000	5000 máx	1000	5000 máx	2000
carga MIPS	1.339	1.339	1.339	2.722	1.490	2.843	2.998	2.843
n° coef.	128	128	128	128	128			
n° blocos	8	8	8	32	8			
aten. voz lin e dB	15.76 x 23.95 dB	20.89 x 26.40 dB	7.91 x 17.96 dB	24.29 x 27.71 dB	21.54 x 26.66 dB			
aten. ruído lin e dB	27.97 x 28.93 dB	40.23 x 32.09 dB	15.61 x 23.87 dB	57.25 x 35.16 dB	35.94 x 31.11 dB			
tempo 50%	2667 ms	1400 ms	-	467 ms	840 ms			
tempo 40%	3600 ms	2033 ms	-	680 ms	1120 ms			
tempo 30%	6133 ms	3167 ms	-	1013 ms	1707 ms			
μ	1000	2000	500	1000	5000 máx			
carga MIPS	1.800	1.800	1.800	3.182	1.951			
n° coef.	256	256	256	256	256			
n° blocos	8	8	8	32	8			
aten. voz lin e dB	12.20 x 21.72 dB	16.85 x 24.53 dB	5.94 x 15.48 dB	8.01 x 18.07 dB	18.45 x 25.32 dB			
aten. ruído lin e dB	12.11 x 21.66 dB	39.43 x 31.92 dB	12.65 x 22.04 dB	20.92 x 26.41 dB	39.53 x 31.94 dB			
tempo 50%	5467 ms	2933 ms	-	913 ms	1533 ms			
tempo 40%	8400 ms	4400 ms	-	1247 ms	2267 ms			
tempo 30%	12667 ms	7000 ms	-	1873 ms	3900 ms			
μ	1000	2000	500	1000	5000 máx			
carga MIPS	2.722	2.722	2.722	4.104	2.873			

ANEXO D - O ALGORITMO *BLOCK UPDATE* LMS

Este anexo busca apresentar o algoritmo *block update* LMS na forma utilizada neste trabalho. Este algoritmo está baseado simplesmente em uma adaptação dos coeficientes em blocos, ou seja, o vetor de coeficientes do filtro FIR é dividido em um dado número de blocos, e apenas um coeficiente de cada bloco é adaptado por iteração do algoritmo. Iterações subsequentes permitem alterar os valores dos outros coeficientes do bloco. Esse tipo de estratégia permite reduzir a carga computacional do algoritmo.

O algoritmo *block update* LMS original [31] apresenta uma diferença em relação ao algoritmo LMS normal, já que utiliza uma média para o cálculo do gradiente ao invés de usar o gradiente instantâneo. Esta média faz com o cálculo da correção aplicada a cada coeficiente seja um pouco mais complicado. A fórmula básica apresentada é dada por:

$$w_i[n+1] = w_i[n] + \mu \sum_{m=0}^{L-1} e[n-m]u[n-m-i] \quad (D.1)$$

Onde i é o coeficiente adaptado e L é o período usado na média para estimar o gradiente. O próximo coeficiente a ser adaptado será $i+LB$, depois $i+2xLB$, e assim por diante, onde LB é a largura de cada bloco, obtida pela divisão do número de coeficientes do filtro (M) pelo número de blocos (NB). A figura D.1 ilustra este processo de adaptação para um filtro FIR de 12 coeficientes no total, subdivididos em 4 blocos de 3 coeficientes cada. São mostradas 3 iterações consecutivas (ciclo completo). O número de coeficientes do filtro deve ser um múltiplo inteiro de NB . Note que, para $L=1$, o algoritmo se reduz ao LMS.

Neste trabalho, esta abordagem foi simplificada devido à necessidade de usar uma carga computacional pequena. O uso da média para o cálculo do gradiente foi eliminado, sendo usada a estimativa instantânea. Com isso, o algoritmo é reduzido a:

$$w_i[n+1] = w_i[n] + 2\beta e[n]u[n-i] \quad (D.2)$$

Que é a mesma fórmula para o algoritmo LMS usual. O algoritmo utilizado é mostrado na figura D.1, onde b é o contador de blocos (1 a NB) e ib é o contador de coeficiente do bloco (0 a $LB-1$). Se $NB=M$, o algoritmo também se reduz ao LMS.

TABELA D.1: Algoritmo *block update* LMS utilizado

INICIALIZAÇÃO:

- 1) $\mathbf{W}[0]=\mathbf{0}$, $ib=0$, $i=0$

LOOP:

- 2) $y[n] = \hat{\mathbf{W}}^T[n]\mathbf{u}[n]$
- 3) $e[n]=d[n]-y[n]$
- 4) $i=ib$, $b=1$

ENQUANTO $b \leq NB$ FAÇA:

- 5) $w_i[n+1] = w_i[n] + \mu e[n]u[n-i]$
- 6) $b=b+1$
- 7) $i=i+LB$

- 8) $ib=ib+1$

SE $ib > LB-1$ ENTÃO $ib=0$

O valor de μ pode ser mudado de acordo com uma estimativa de potência do sinal de entrada do filtro, conforme a equação D.3, onde $P[n]$ é uma estimativa para a potência do sinal.

$$\mu = \frac{\mu_0}{P[n]} \quad (D.3)$$

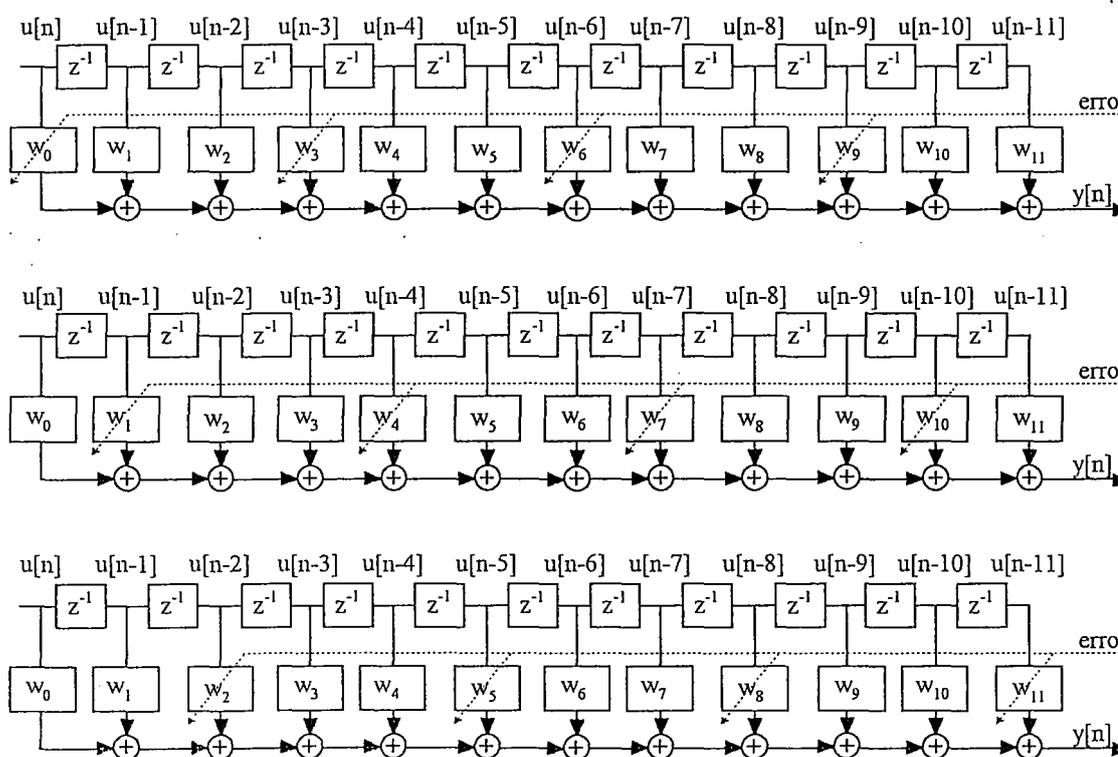


FIGURA D.1: Processo de adaptação dos coeficientes do filtro usando o algoritmo *block update* LMS

ANEXO E - NOTAÇÃO VETORIAL

O uso da notação vetorial torna mais direta a análise e o desenvolvimento de equações. No entanto, para o caso de vetores cujos elementos são números complexos, algumas definições devem ser estendidas. O produto interno (*inner product*) dos vetores \mathbf{x} e \mathbf{y} é definido como [39] (pág. 323) (o domínio onde é feita a operação será representado por subscrito: R ou C - reais ou complexos):

$$\mathbf{x} = [x_0 \quad x_1 \quad \cdots \quad x_n]^T \quad (\text{E.1})$$

$$\mathbf{y} = [y_0 \quad y_1 \quad \cdots \quad y_n]^T \quad (\text{E.2})$$

$$\mathbf{x} \cdot \mathbf{y} \Big|_C \triangleq x_0^* y_0 + x_1^* y_1 + \dots + x_n^* y_n \quad (\text{E.3})$$

Algumas propriedades (domínio complexo):

$$\mathbf{x} \cdot \mathbf{y} = (\mathbf{y} \cdot \mathbf{x})^* \quad (\text{E.4})$$

$$(\mathbf{x} + \mathbf{y}) \cdot \mathbf{z} = \mathbf{x} \cdot \mathbf{z} + \mathbf{y} \cdot \mathbf{z} \quad (\text{E.5})$$

$$\mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z} \quad (\text{E.6})$$

$$(k\mathbf{x}) \cdot \mathbf{y} = k^* (\mathbf{x} \cdot \mathbf{y}) \quad (\text{E.7})$$

$$\mathbf{x} \cdot (k\mathbf{y}) = k (\mathbf{x} \cdot \mathbf{y}) \quad (\text{E.8})$$

onde k é um escalar.

De forma a tornar a operação de conjugado complexo mais visível, adota-se como definição de produto interno dos vetores \mathbf{x} e \mathbf{y} [18]:

$$\mathbf{x} \cdot \mathbf{y} \Big|_C \triangleq \mathbf{x}^* \cdot \mathbf{y} \Big|_R = \mathbf{x}^H \mathbf{y} = \mathbf{y}^T \mathbf{x}^* \quad (\text{E.9})$$

onde H representa as operações de transposição e conjugado complexo.

O produto vetorial é definido por [18]:

$$\mathbf{x} \times \mathbf{y} \Big|_C \triangleq \mathbf{x} \mathbf{y}^H \quad (\text{E.10})$$

Um filtro FIR representa uma aproximação à soma de convolução:

$$y[n] = \sum_{k=-\infty}^{+\infty} h[k]x[n-k] = h[n] * x[n] \quad (\text{E.11})$$

onde os limites do somatório estão definidos. A resposta de um filtro FIR com N+1 coeficientes h a um sinal x[n] é dada por

$$y[n] = \sum_{k=0}^N h[k]x[n-k] \quad (\text{E.12})$$

que é igual ao produto interno do vetor formado pelos coeficientes conjugados do filtro pelo vetor formado pelo sinal x[n] e suas amostras atrasadas no tempo, x[n-1], ..., x[n-N]:

$$y[n] = \mathbf{h}^* \cdot \mathbf{x}[n] \Big|_C = \mathbf{h}^T \mathbf{x}[n] \quad (\text{E.13})$$

Para o caso real, o vetor utilizado no produto interno é igual à resposta ao impulso do filtro.

REFERÊNCIAS BIBLIOGRÁFICAS

1. AHMED, N. Adaptive filtering. In: ELLIOT, D.F. **Handbook of digital signal processing - engineering applications**. San Diego, Ca: Academic Press, 1987. p.857-897
2. AHMED, N.; FOGLER, R.J. A matrix bordering approach for deriving lattice models. **IEEE Transactions on Aerospace and Electronic Systems**, v.20, n.6, p.835-838, Nov. 1984.
3. BORCHARDT, I.G.; BRITO, R.M. **Instrumentação para monitoração e controle de processos**. Porto Alegre: PPGEMM-UFRGS, 1991. 49p. (Caderno Técnico)
4. CARAYANNIS, G.; MANOLAKIS, D.G.; KALOUPTSIDIS, N. A fast sequential algorithm for least-squares filtering and prediction. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v.31, n.6, p.1394-1402, Dec. 1983.
5. CIOFFI, J.M.; KAILATH, T. Fast, recursive-least-squares transversal filters for adaptive filtering. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v.32, n.2, p.304-337, Apr. 1984.
6. DOEBELIN, E.O. **Measurement systems application and design**. 4.ed. Singapore: McGraw-Hill, 1990. 960p.
7. DORNELES, R.V. **Um estudo sobre processamento adaptativo de sinais utilizando redes neurais**. Porto Alegre, 1993. 116p. Dissertação (Mestrado em Informática) - Instituto de Informática, Curso de Pós-Graduação em Ciência da Computação da UFRGS.
8. FALCONER, D.D.; LJUNG, L. Application of fast Kalman estimation to adaptive equalization. **IEEE Transactions on Communications**, v.26, n.10, p.1439-1446, Oct. 1978.
9. FAN, H. A structural view of asymptotic convergence speed of adaptive IIR filtering algorithms: part I - infinite precision implementation. **IEEE Transactions on Signal Processing**, v.41, n.4, p.1493-1517, Apr. 1993.

10. FAN, H.; JENKINS, W.K. An investigation of an adaptive IIR echo canceller: advantages and problems. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v.36, n.12, p.1819-1834, Dec. 1988.
11. FAN, H.; LIU, X. Gal and LSL revisited: new convergence results. **IEEE Transactions on Signal Processing**, v.41, n.1, p.55-66, Jan. 1993.
12. FEDER, M.; OPPENHEIM, A.V.; WEINSTEIN, E. Maximum likelihood noise cancellation using the EM algorithm. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v.37, n.2, p.204-216, Feb. 1989.
13. FEINTUCH, P.L. An adaptive recursive LMS filter. **Proceedings of the IEEE**, v.64, n.11, p.1622-1624, Nov. 1976.
14. FRIEDLANDER, B. Lattice filters for adaptive processing. **Proceedings of the IEEE**, v.70, n.8, p.829-867, Aug. 1982.
15. GRAY Jr., A.H.; MARKEL, J.D. Digital lattice and ladder filter synthesis. **IEEE Transactions on Audio and Electroacoustics**, v.21, n.6, p.491-500, Dec. 1973.
16. GRIFFITHS, L.J. A continuously-adaptive filter implemented as a lattice structure. In: **IEEE INTERNATIONAL CONFERENCE ON ACOUSTIC, SPEECH AND SIGNAL PROCESSING, 2., 1977, Hartford, Conn. Proceedings...** New York: IEEE, 1977. p.683-686.
17. GRIFFITHS, L.J. Adaptive structures for multiple-input noise cancelling applications. In: **IEEE INTERNATIONAL CONFERENCE ON ACOUSTIC, SPEECH AND SIGNAL PROCESSING, 4., 1979, Washington, DC. Proceedings...** New York: IEEE, 1979. p.925-928.
18. HAYKIN, S. **Adaptive filter theory**. 2.ed. Englewood Cliffs, N.J.: Prentice-Hall, 1991. 854p.
19. HSIA, T.C. A simplified adaptive recursive filter design. **Proceedings of the IEEE**, v.69, n.9, p.1153-1155, Sept. 1981.
20. INTERNATIONAL TELEGRAPH AND TELEPHONE. Consultive Committee. **Recommendation G.165: echo cancellers**. In: _____. **Blue book**, ITT: Melbourne, 1988. v.3, fasc.3.1, p.221-243.
21. JOHNSON Jr., C.R. et al. SHARF convergence properties. **IEEE Transactions on Circuits and Systems**, v.28, n.6, p.499-510, June 1981.

22. JOHNSON Jr., C.R.; LARIMORE, M.G. Comments on and additions to "An adaptive recursive LMS filter". **Proceedings of the IEEE**, v.65, n.9, p.1399-1402, Sept. 1977.
23. KAILATH, T. A view of three decades of linear filtering theory. **IEEE Transactions on Information Theory**, v.20, n.2, p.146-181, Mar. 1974.
24. LARIMORE, M.G.; TREICHLER, J.R.; JOHNSON Jr., C.R. SHARF: an algorithm for adapting IIR digital filters. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v.28, n.4, p.428-440, Aug. 1980.
25. LEE, J.; MATHEWS, V.J. A fast recursive least-squares adaptive second-order Volterra filter and its performance analysis. **IEEE Transactions on Signal Processing**, v.41, n.3, p.1087-1102, Mar. 1993.
26. LING, F.; MANOLAKIS, D.; PROAKIS, J.G. Numerically robust least-squares lattice-ladder algorithms with direct updating of the reflection coefficients. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v.34, n.4, p.837-845, Aug. 1986.
27. LUFT, J. A.; NEGREIROS, M.; WEIHMANN, T. **Sistema de processamento de sinais usando TMS320C25**. Porto Alegre: Laboratório de Processamento de Sinais, Departamento de Engenharia Elétrica da UFRGS, 1993. 118p. Documentação interna.
28. LUZ, L.T.O. da **Sistema de identificação, modelagem, análise e projeto de sistemas de controle para sistemas elétricos de potência**. Florianópolis, 1990. 158p. Dissertação (Mestrado em Engenharia) - Escola de Engenharia, Curso de Pós-Graduação em Engenharia Elétrica da UFSC.
29. MAKHOUL, J. A class of all-zero lattice digital filters: properties and applications. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v.26, n.4, p.304-314, Aug. 1978.
30. MAKHOUL, J. Linear prediction: a tutorial review. **Proceedings of the IEEE**, v.63, n.4, p.561-580, Apr. 1975.
31. MESSERSCHMITT, D. et al. Digital voice echo canceller with a TMS32020. In: TEXAS INSTRUMENTS. **Digital signal processing applications with the TMS320 family**. [S.n.] 1986. p.415-454
32. NANNI, L.F.; RIBEIRO, J.L. **Planejamento e avaliação de experimentos**. Porto Alegre: CPGEC-UFRGS, 1992. 194p. (Caderno Técnico)

33. OPPENHEIM, A.V.; SCHAFER, R.W. **Discrete time signal processing**. Englewood Cliffs, N.J.: Prentice-Hall, 1989. 879p.
34. PAPOULIS, A. **Probability, random variables, and stochastic processes**. 3.ed. Singapore: Mc Graw-Hill, 1991. 666p.
35. PARIKH, D.; AHMED, N. On an adaptive algorithm for IIR filters. **Proceedings of the IEEE**, v.66, n.5, p.585-588, May 1978.
36. PARIKH, D.; AHMED, N.; STEARNS, S.D. An adaptive lattice algorithm for recursive filters. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v.28, n.1, p.110-111, Feb. 1980.
37. REGALIA, P.A. Stable and efficient lattice algorithms for adaptive IIR filtering. **IEEE Transactions on Signal Processing**, v.40, n.2, p.375-388, Feb. 1992.
38. SATORIUS, E.H.; ALEXANDER, S.T. Channel equalization using adaptive lattice algorithms **IEEE Transactions on Communications**, v.27, n.6, p.899-905, June 1979.
39. SOKOLNIKOFF, I.S.; REDHEFFER, R.M. **Mathematics of physics and modern engineering**. Tokyo, Japan: McGraw-Hill, 1958. 812p.
40. STEARNS, S.D. Error surfaces of recursive adaptive filters. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v.29, n.3, p.763-766, June 1981.
41. TEXAS INSTRUMENTS. **TMS320C25 user's guide**. [S.n.] 1986.
42. THERRIEN, C.W. **Discrete random signals and statistical signal processing**. Englewood Cliffs, N.J.: Prentice-Hall, 1992. 727p.
43. TREICHLER, J.R.; JOHNSON Jr., C.R.; LARIMORE, M.G. **Theory and design of adaptive filters**. New York: John Wiley, 1987. 342p.
44. WANG, L.; MENDEL, J.M. Fuzzy adaptive filters, with applications to nonlinear channel equalization. **IEEE Transactions on Fuzzy Systems**, v.1, n.3, p.161-170, Aug. 1993.
45. WEIHMANN, T. **Processamento digital de sinais aplicado à transmissão de voz**. Porto Alegre, 1992. 156p. Dissertação (Mestrado em Engenharia) - Escola de Engenharia, Curso de Pós-Graduação em Engenharia Metalúrgica e dos Materiais da UFRGS.

46. WIDROW, B. et al. Adaptive noise cancelling: principles and applications. **Proceedings of the IEEE**, v.63, n.12, p.1692-1716, Dec. 1975.
47. WIDROW, B. et al. Stationary and nonstationary learning characteristics of the LMS adaptive filter. **Proceedings of the IEEE**, v.64, n.8, p.1151-1162, Aug. 1976.
48. WIDROW, B.; MCCOOL, J.M. Comments on "An adaptive recursive LMS filter". **Proceedings of the IEEE**, v.65, n.9, p.1402-1404, Sept. 1977.
49. WIDROW, B.; STEARNS, S.D. **Adaptive signal processing**. Englewood Cliffs, N.J.: Prentice-Hall, 1985. 474p.
50. WILLIAMSON, G.A.; ZIMMERMANN, S. **Globally convergent adaptive IIR filters based on fixed pole locations**. Chicago, Il.: Department of Electrical and Computer Engineering, Illinois Institute of Technology, 1994. 28p. Não publicado.
51. YIN, L.; ASTOLA, J.; NEUVO, Y. A new class of nonlinear filters - neural filters. **IEEE Transactions on Signal Processing**, v.41, n.3, p.1201-1222, Mar. 1993.

APÊNDICE I - PROBABILIDADE, VARIÁVEIS ALEATÓRIAS E PROCESSOS ESTOCÁSTICOS

I.1 Introdução

O objetivo deste apêndice é prover um panorama sobre os aspectos relevantes da teoria de probabilidade para o estudo de sinais e otimização de filtros lineares. Busca-se apresentar os resultados e definições importantes para o estudo e caracterização de processos estocásticos e variáveis aleatórias. O livro de Papoulis [34] é recomendado fortemente em caso de dúvidas.

I.2 Probabilidade e variáveis aleatórias

Considere um experimento como a jogada de um dado ou de uma moeda. A saída (resultado) deste experimento é a face do dado voltada para cima ou a cara/coroa da moeda. Uma saída em particular deste experimento é ζ_i . O conjunto formado por todas as saídas possíveis é o espaço S . Subconjuntos de S (conjuntos formados com as saídas ζ_i) são chamados de eventos. O espaço S é o evento certo (contém todos os eventos). O conjunto vazio $\{\}$ ou \emptyset é o evento impossível (não contém eventos). Se dois eventos A e B são mutuamente exclusivos, $A \cap B = \emptyset$.

A probabilidade de um evento A é um número $P(A)$ que satisfaz as seguintes condições (axiomas) (figura I.1):

- i) $P(A) \geq 0$
- ii) $P(S) = 1$
- iii) Se $A \cap B = \emptyset$ então $P(A+B) = P(A) + P(B)$

Conclui-se :

$$P(\emptyset) = 0 \tag{I.1}$$

$$P(A) = 1 - P(\bar{A}) \leq 1 \tag{I.2}$$

onde \bar{A} é o complemento de A (ou $S-A$).

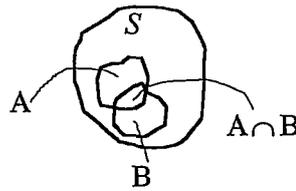


FIGURA I.1: O espaço S e dois eventos A e B , onde $A \cap B \neq \emptyset$

Se dois eventos A e B são independentes, pode-se dizer que a probabilidade de que eles ocorram conjuntamente é dada por $P(A)P(B)$, ou seja, $P(A \cap B) = P(A)P(B)$. Caso contrário, a probabilidade de que eles ocorram conjuntamente não pode ser expressa em termos das probabilidades individuais de ocorrência dos eventos.

Em geral, eventos não são conjuntos formados por números. Se a cada saída ζ_i for atribuído um número de acordo com uma função $x(\zeta_i)$ que obedeça a certas condições, chama-se esta função de variável aleatória. Ao ser especificado o conjunto $\{x \leq x\}$, significa dizer $\{\zeta_i \mid x(\zeta_i) \leq x, \text{ para todo } i\}$, ou seja, o conjunto formado por todas as saídas experimentais ζ_i tais que o valor da variável aleatória $x(\zeta_i)$ é menor ou igual ao valor do número x .

Uma variável aleatória real x é uma função cujo domínio é o espaço S , tal que:

- i) O conjunto $\{x \leq x\}$ é um evento para todo número real x .
- ii) A probabilidade dos eventos $\{x = +\infty\}$ e $\{x = -\infty\}$ é zero:

$$P\{x = +\infty\} = P\{x = -\infty\} = 0 \quad (\text{I.3})$$

Se a variável aleatória for complexa, a função é complexa: $z(\zeta_i) = x(\zeta_i) + jy(\zeta_i)$.

I.2.1 Função distribuição

A função distribuição de uma variável aleatória x é a função $F_x(x) = P\{x \leq x\}$, definida para qualquer x de $-\infty$ a $+\infty$. Portanto, para um dado x , $F_x(x)$ iguala a probabilidade do evento $\{x \leq x\}$. Propriedades da função de distribuição:

- i) $F(+\infty) = 1$ e $F(-\infty) = 0$.
- ii) É uma função não-decrescente de x , ou seja, se $x_1 \leq x_2$, então $F(x_1) \leq F(x_2)$.
- iii) É contínua à direita: $F(x^+) = F(x)$, onde

$$F(x^+) = \lim_{\varepsilon \rightarrow 0} F(x + \varepsilon), \text{ com } \varepsilon > 0 \quad (\text{I.4})$$

I.2.2 Função densidade

A derivada da função distribuição é chamada função densidade de probabilidade da variável aleatória x . Esta função pode conter impulsos (variáveis aleatórias do tipo discreto), podendo ser definida mesmo que $F(x)$ contenha um número finito de pontos não-diferenciáveis (figura I.2). Propriedades da função densidade:

- i) $f(x) \geq 0$, para todo x
- ii) $\int_{-\infty}^{+\infty} f(x) dx = 1$
- iii) $P\{x \leq x\} = \int_{-\infty}^x f(x) dx$

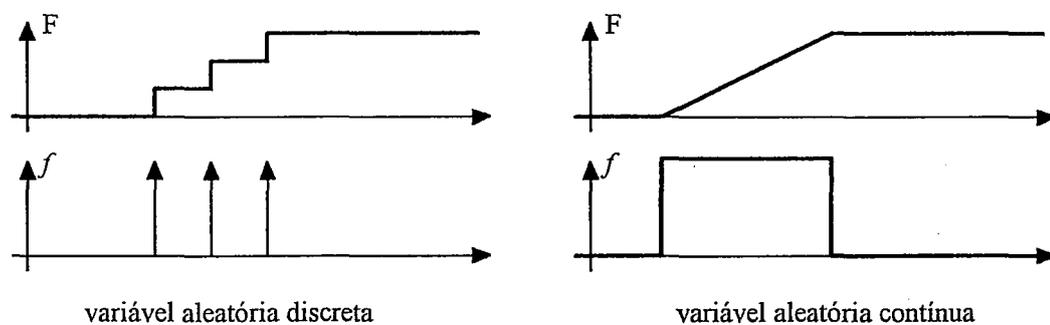


FIGURA I.2: Exemplo de funções distribuição e densidade de variáveis aleatórias

O teorema da existência permite que se faça o uso de variáveis aleatórias sem definir um experimento específico, contanto que se saiba a função distribuição ou a função densidade das variáveis aleatórias.

I.2.3 Função de variável aleatória

Se $g(x)$ é uma função de uma variável real x , pode-se definir a expressão $y=g(x)$, onde x é uma variáveis aleatória e y passa a ser também uma variável aleatória. O domínio da função composta $g \circ x$ é o conjunto formado pelas saídas do experimento ζ_i , ou seja, a variável aleatória y é $y(\zeta_i) = g(x(\zeta_i))$. A função distribuição de y é dada por $F_y(y) = P\{y \leq y\} = P\{g(x) \leq y\}$.

I.2.4 Caracterização de variáveis aleatórias

Define-se as seguintes quantidades para variáveis aleatórias do tipo contínuo:

a) valor esperado ou média

$$\varepsilon\{x\} = \int_{-\infty}^{+\infty} x f(x) dx = \eta_x = \eta \quad (\text{I.5})$$

b) variância

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - \eta)^2 f(x) dx \quad (\text{I.6})$$

onde σ é o desvio padrão de x .

c) momentos

$$m_n = \varepsilon\{x^n\} = \int_{-\infty}^{+\infty} x^n f(x) dx \quad (\text{I.7})$$

d) momentos centrais

$$\mu_n = \varepsilon\{(x - \eta)^n\} = \int_{-\infty}^{+\infty} (x - \eta)^n f(x) dx \quad (\text{I.8})$$

I.2.5 Duas variáveis aleatórias

Dadas duas variáveis aleatórias x e y e os eventos $\{x \leq x\}$ e $\{y \leq y\}$, têm-se que a probabilidade destes eventos é dada por $P\{x \leq x\} = F_x(x)$ e $P\{y \leq y\} = F_y(y)$, funções de distribuição das variáveis aleatórias x e y . Estas funções determinam as suas estatísticas separadas (marginais), mas não a sua estatística conjunta, isto é, a probabilidade do evento $\{x \leq x\} \cap \{y \leq y\} = \{x \leq x, y \leq y\}$ não pode ser expressa em termos de $F_x(x)$ e $F_y(y)$. A função conjunta de distribuição das variáveis aleatórias x e y é definida por $F_{xy}(x, y) = P\{x \leq x, y \leq y\}$. As funções distribuição marginais se relacionam à função distribuição conjunta por :

$$F_{xy}(x, \infty) = F_x(x) \quad (\text{I.9})$$

$$F_{xy}(\infty, y) = F_y(y) \quad (\text{I.10})$$

$$F_{xy}(\infty, \infty) = 1 \quad (\text{I.11})$$

$$F_{xy}(-\infty, y) = 0 \quad (\text{I.12})$$

$$F_{xy}(x, -\infty) = 0 \quad (\text{I.13})$$

A função densidade conjunta de x e y é a função:

$$f(x, y) \triangleq \frac{\partial^2 F(x, y)}{\partial x \partial y} \quad (\text{I.14})$$

E se relaciona à função distribuição conjunta por:

$$F_{xy}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(\alpha, \beta) d\alpha d\beta \quad (\text{I.15})$$

Duas variáveis aleatórias são ditas (estatisticamente) independentes se os eventos $\{x \leq x\}$ e $\{y \leq y\}$ são independentes para qualquer x e y , ou seja, se $P\{x \leq x, y \leq y\} = P\{x \leq x\}P\{y \leq y\}$. Segue que $F_{xy}(x, y) = F_x(x)F_y(y)$ e que $f_{xy}(x, y) = f_x(x)f_y(y)$.

Dadas duas variáveis aleatórias x e y e uma função $g(x, y)$ de variáveis reais x e y que satisfaz certas condições gerais, então $z = g(x, y)$ é uma variável aleatória e seu valor $z(\zeta_i)$ é dado por $z(\zeta_i) = g(x(\zeta_i), y(\zeta_i))$. O valor esperado de z é:

$$\varepsilon\{z\} = \int_{-\infty}^{+\infty} z f(z) dz \quad (\text{I.16})$$

ou

$$\varepsilon\{z = g(x, y)\} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dx dy \quad (\text{I.17})$$

de onde se pode concluir que $\varepsilon\{x+y\} = \varepsilon\{x\} + \varepsilon\{y\}$.

A covariância das variáveis aleatórias x e y é definida como:

$$C_{xy} = \int_{-\infty}^{+\infty} (x - \eta_x)(y - \eta_y) f(x, y) dx dy = \varepsilon\{(x - \eta_x) \cdot (y - \eta_y)\} = \varepsilon\{x \cdot y\} - \varepsilon\{x\}\varepsilon\{y\} \quad (\text{I.18})$$

As duas variáveis aleatórias são ditas não-correlacionadas se $C_{xy} = 0$, ou $\varepsilon\{xy\} = \varepsilon\{x\}\varepsilon\{y\}$.

As variáveis são ditas ortogonais se $\varepsilon\{xy\} = 0$.

Se duas variáveis aleatórias são independentes, então elas são não-correlacionadas, ou seja, $C_{xy} = 0$.

Os momentos conjuntos de ordem n de duas variáveis aleatórias x e y são definidos por:

$$m_{kr} = \varepsilon\{x^k y^r\} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^k y^r f(x, y) dx dy \quad (\text{I.19})$$

onde $n=k+r$.

Os momentos de primeira ordem são $m_{10}=\eta_x$ e $m_{01}=\eta_y$, médias de x e y . Os momentos de segunda ordem são $m_{20}=\varepsilon\{x^2\}$, $m_{02}=\varepsilon\{y^2\}$ e $m_{11}=\varepsilon\{xy\}$, médias dos valores quadrados de x e y e correlação das variáveis aleatórias. Os momentos centrais são definidos similarmente, sendo subtraído o valor da média de cada variável aleatória ($x \rightarrow (x-\eta_x)$). Os momentos centrais de primeira ordem são zero e os de segunda ordem são $\mu_{11}=C_{xy}$, $\mu_{20}=\sigma_x^2$, $\mu_{02}=\sigma_y^2$, covariância e variância das variáveis aleatórias x e y .

Para a determinação das estatísticas conjuntas de x e y o conhecimento de sua densidade conjunta é necessário. No entanto, em muitas aplicações, somente os momentos de primeira e segunda ordem são usados, sendo determinados em termos dos seguintes parâmetros: η_x , η_y , σ_x , σ_y e C_{xy} .

Estes conceitos podem ser estendidos para N variáveis aleatórias e para casos onde se considera variáveis aleatórias complexas.

I.3 Processos estocásticos

Uma variável aleatória é uma regra que atribui a cada saída ζ_i de um experimento um número $x(\zeta_i)$. Um processo estocástico $x(t)$ é uma regra que atribui a cada ζ_i uma função $x(t, \zeta_i)$, real ou complexa. Se o domínio de t é o conjunto dos números reais, o processo $x(t)$ é dito contínuo no tempo. Se o domínio é o conjunto dos números inteiros, o processo $x(t)$ é dito discreto no tempo, sendo denotado por $x[n]$. Neste último caso o processo consiste de uma seqüência de variáveis aleatórias. Pode-se interpretar $x(t)$ de quatro maneiras diferentes (figura I.3):

- 1) t e ζ indefinidos. $x(t, \zeta)$ é uma família(conjunto) de funções $x(t, \zeta)$.
- 2) t e ζ definidos. $x(t, \zeta)$ é um número real ou complexo.
- 3) t definido e ζ indefinido. $x(t, \zeta)$ é uma variável aleatória.
- 4) t indefinido e ζ definido. $x(t, \zeta)$ é uma função do tempo.

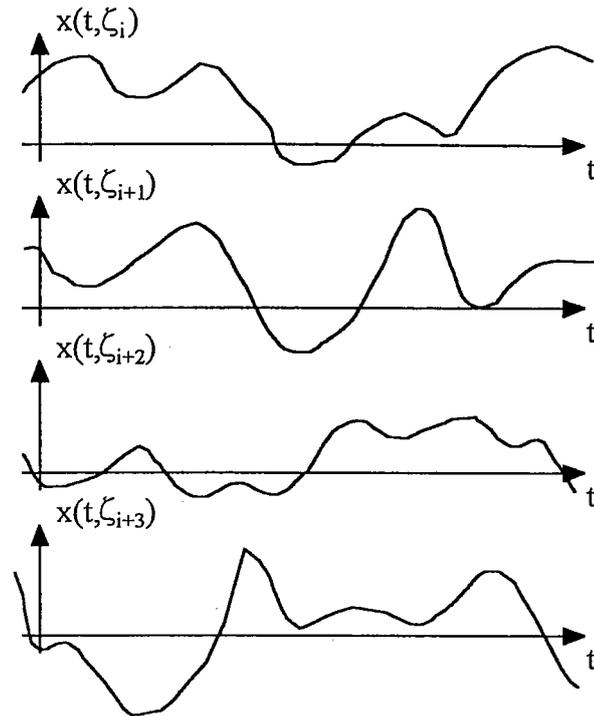


FIGURA I.3: Conjunto (família) de funções que formam um processo estocástico contínuo no tempo

Um processo estocástico é uma infinidade de variáveis aleatórias, uma para cada t . Para um t específico, $x(t)$ é uma variável aleatória com distribuição $F(x,t) = P\{x(t) \leq x\}$. Essa função depende de t e iguala a probabilidade do evento $\{x(t) \leq x\}$, consistindo de todas as saídas ζ tais que, num instante específico t , as amostras $x(t, \zeta)$ do processo não excedam o número x , conforme ilustra a figura I.4. Sendo $F(x,t)$ a função distribuição de primeira ordem do processo $x(t)$, a função densidade de primeira ordem é definida por:

$$f(x,t) = \frac{\partial F(x,t)}{\partial x} \quad (\text{I.20})$$

A distribuição de segunda ordem é a função distribuição conjunta das variáveis aleatórias $x(t_1)$ e $x(t_2)$, $F(x_1, x_2; t_1, t_2) = P\{x(t_1) \leq x_1, x(t_2) \leq x_2\}$. A função densidade correspondente é

$$f(x_1, x_2; t_1, t_2) = \frac{\partial^2 F(x_1, x_2; t_1, t_2)}{\partial x_1 \partial x_2} \quad (\text{I.21})$$

A média do processo $x(t)$ no instante t é o valor esperado da variável aleatória $x(t)$.

$$\eta(t) = \varepsilon\{x(t)\} = \int_{-\infty}^{+\infty} x f(x,t) dx \quad (\text{I.22})$$

A autocorrelação $R(t_1, t_2)$ do processo estocástico $\mathbf{x}(t)$ é o valor esperado do produto $\mathbf{x}(t_1)\mathbf{x}(t_2)$.

$$R(t_1, t_2) = \varepsilon\{\mathbf{x}(t_1)\mathbf{x}(t_2)\} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_1 x_2 f(x_1, x_2; t_1, t_2) dx_1 dx_2 \quad (\text{I.23})$$

A autocovariância é a covariância das variáveis aleatórias $\mathbf{x}(t_1)$ e $\mathbf{x}(t_2)$.

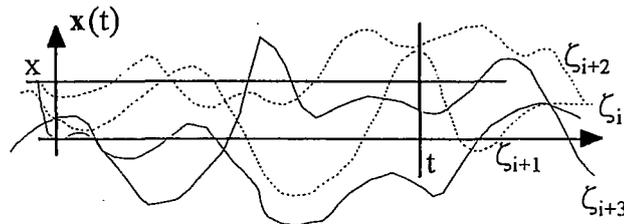


FIGURA I.4: Interpretação da função distribuição de um processo estocástico, dados x e t

I.3.1 Propriedades gerais

As propriedades estatísticas de um processo estocástico real $\mathbf{x}(t)$ são determinadas completamente em termos de sua distribuição de n -ésima ordem: $F(x_1, \dots, x_n; t_1, \dots, t_n) = P\{\mathbf{x}(t_1) \leq x_1, \dots, \mathbf{x}(t_n) \leq x_n\}$. As estatísticas conjuntas de dois processos estocásticos reais $\mathbf{x}(t)$ e $\mathbf{y}(t)$ são determinadas em termos da distribuição conjunta das variáveis aleatórias $\mathbf{x}(t_1), \dots, \mathbf{x}(t_n), \mathbf{y}(t_1'), \dots, \mathbf{y}(t_n')$. O processo complexo $\mathbf{z}(t) = \mathbf{x}(t) + j\mathbf{y}(t)$ é especificado em termos das estatísticas conjuntas dos processos reais $\mathbf{x}(t)$ e $\mathbf{y}(t)$.

A autocorrelação $R_{xx}(t_1, t_2)$ de um processo $\mathbf{x}(t)$, real ou complexo, é dada por:

$$R_{xx}(t_1, t_2) \triangleq \varepsilon\{\mathbf{x}(t_1)\mathbf{x}^*(t_2)\} \quad (\text{I.24})$$

$$R_{xx}(t_1, t_2) = R_{xx}^*(t_2, t_1) \quad (\text{I.25})$$

A autocovariância de $\mathbf{x}(t)$ é a covariância das variáveis aleatórias $\mathbf{x}(t_1)$ e $\mathbf{x}(t_2)$.

$$C_{xx}(t_1, t_2) = R_{xx}(t_1, t_2) - \eta_x(t_1)\eta_x^*(t_2) \quad (\text{I.26})$$

A correlação cruzada de dois processos $\mathbf{x}(t)$ e $\mathbf{y}(t)$ é dada por:

$$R_{xy}(t_1, t_2) \triangleq \varepsilon\{\mathbf{x}(t_1)\mathbf{y}^*(t_2)\} \quad (\text{I.27})$$

$$R_{xy}(t_1, t_2) = R_{yx}^*(t_2, t_1) \quad (\text{I.28})$$

A covariância cruzada é definida similarmente como:

$$C_{xy}(t_1, t_2) = R_{xy}(t_1, t_2) - \eta_x(t_1)\eta_y^*(t_2) \quad (\text{I.29})$$

Dois processos $x(t)$ e $y(t)$ são ditos (mutuamente) ortogonais se $R_{xy}(t_1, t_2) = 0$ para todo t_1 e t_2 . Eles são ditos não-correlacionados se $C_{xy}(t_1, t_2) = 0$ para todo t_1 e t_2 . Estes processos são ditos independentes se as variáveis aleatórias $x(t_1), \dots, x(t_n), y(t_1'), \dots, y(t_n')$ são mutuamente independentes.

I.3.2 Processos estacionários

Um processo estocástico $x(t)$ é chamado estacionário no sentido estrito (SSS - *Strict Sense Stationary*) se suas propriedades estatísticas são invariantes a um deslocamento da origem, ou seja, os processos $x(t)$ e $x(t+c)$ têm as mesmas estatísticas para todo c . Dois processos $x(t)$ e $y(t)$ são conjuntamente estacionários (SSS) se as estatísticas conjuntas de $x(t)$ e $y(t)$ são as mesmas de $x(t+c)$ e $y(t+c)$ para todo c . Um processo complexo $z(t) = x(t) + jy(t)$ é dito estacionário se os processos $x(t)$ e $y(t)$ são conjuntamente estacionários. Segue que a função densidade de primeira ordem de um processo estocástico SSS é independente de t , e a de segunda ordem depende apenas da diferença $\tau = t_1 - t_2$: $f(x; t) = f(x)$ e $f(x_1, x_2; t_1, t_2) = f(x_1, x_2; \tau)$.

Um processo estocástico é dito estacionário no sentido amplo (WSS - *Wide Sense Stationary*) se a sua média é constante, $\varepsilon\{x(t)\} = \eta$, e sua autocorrelação depende somente de $\tau = t_1 - t_2$, $R(\tau) = \varepsilon\{x(t+\tau)x^*(t)\}$. Sua autocovariância é dada por $C(\tau) = R(\tau) - |\eta|^2$. Dois processos estocásticos $x(t)$ e $y(t)$ são conjuntamente WSS se cada um é WSS e suas correlações e covariâncias cruzadas dependem somente de $\tau = t_1 - t_2$: $R_{xy}(\tau) = \varepsilon\{x(t+\tau)y^*(t)\}$ e $C_{xy}(\tau) = R_{xy}(\tau) - \eta_x\eta_y^*$.

I.3.3 Processos estocásticos discretos

Processos estocásticos discretos no tempo são uma seqüência de variáveis aleatórias e serão representados por $x[n]$. Para cada instante de tempo n' , temos uma única variável aleatória $x[n']$. A figura I.5 mostra o processo estocástico discreto de modo similar à figura I.3. Neste trabalho houve mais interesse neste tipo de processo, estendendo os conceitos importantes de processos estocásticos contínuos.

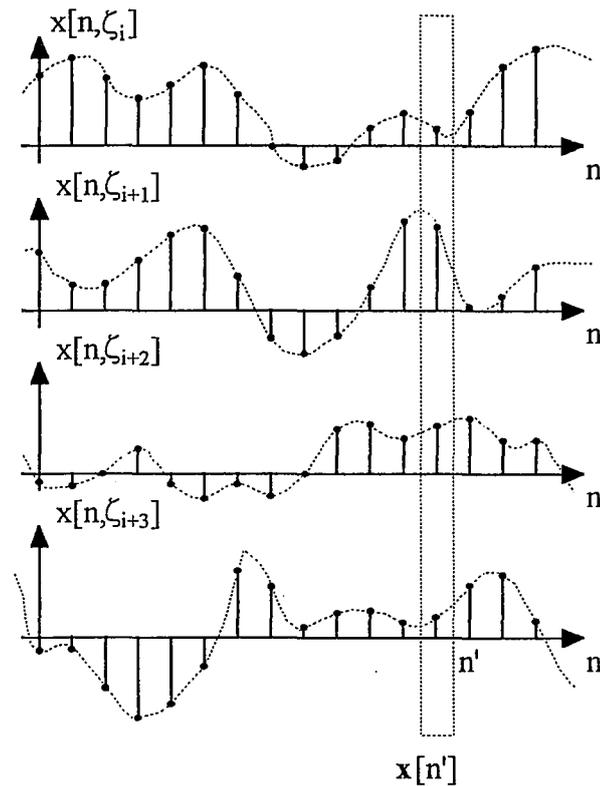


FIGURA I.5: Conjunto (família) de funções que formam um processo estocástico discreto no tempo

A seguir são fornecidos dois quadros com as principais grandezas para processos estocásticos contínuos e discretos no tempo, para o caso geral e para o caso estacionário no sentido amplo (WSS).

TABELA I.1: Caracterização de processos estocásticos contínuos no tempo ($\tau=t_1-t_2$)

grandezas	processo estocástico	processo estocástico WSS
média	$\varepsilon\{\mathbf{x}(t)\}=\eta_x(t)$	$\varepsilon\{\mathbf{x}(t)\}=\eta_x$
autocorrelação	$R_x(t_1, t_2)=\varepsilon\{\mathbf{x}(t_1)\mathbf{x}^*(t_2)\}$	$R_x(\tau)=\varepsilon\{\mathbf{x}(t+\tau)\mathbf{x}^*(t)\}$
autocovariância	$C_x(t_1, t_2)=R_x(t_1, t_2)-\eta_x(t_1)\eta_x^*(t_2)$	$C_x(\tau)=R_x(\tau)- \eta_x ^2$
correlação cruzada	$R_{xy}(t_1, t_2)=\varepsilon\{\mathbf{x}(t_1)\mathbf{y}^*(t_2)\}$	$R_{xy}(\tau)=\varepsilon\{\mathbf{x}(t+\tau)\mathbf{y}^*(t)\}$
covariância cruzada	$C_{xy}(t_1, t_2)=R_{xy}(t_1, t_2)-\eta_x(t_1)\eta_y^*(t_2)$	$C_{xy}(\tau)=R_{xy}(\tau)-\eta_x\eta_y^*$

TABELA I.2: Caracterização de processos estocásticos discretos no tempo ($m=n_1-n_2$)

grandeza	processo estocástico discreto no tempo	processo estocástico discreto no tempo WSS
média	$\varepsilon\{\mathbf{x}[n]\}=\eta_x[n]$	$\varepsilon\{\mathbf{x}[n]\}=\eta_x$
autocorrelação	$R_x[n_1,n_2]=\varepsilon\{\mathbf{x}[n_1]\mathbf{x}^*[n_2]\}$	$R_x[m]=\varepsilon\{\mathbf{x}[n+m]\mathbf{x}^*[n]\}$
autocovariância	$C_x[n_1,n_2]=R_x[n_1,n_2]-\eta_x[n_1]\eta_x^*[n_2]$	$C_x[m]=R_x[m]- \eta_x ^2$
correlação cruzada	$R_{xy}[n_1,n_2]=\varepsilon\{\mathbf{x}[n_1]\mathbf{y}^*[n_2]\}$	$R_{xy}[m]=\varepsilon\{\mathbf{x}[n+m]\mathbf{y}^*[n]\}$
covariância cruzada	$C_{xy}[n_1,n_2]=R_{xy}[n_1,n_2]-\eta_x[n_1]\eta_y^*[n_2]$	$C_{xy}[m]=R_{xy}[m]-\eta_x\eta_y^*$

APÊNDICE II - ESTIMATIVA MÉDIA DO QUADRADO LINEAR E O PRINCÍPIO DA ORTOGONALIDADE

II.1 Introdução

Este apêndice está preocupado em estabelecer a idéia de estimação linear de variável aleatória e o princípio da ortogonalidade. O princípio da ortogonalidade é uma ferramenta muito poderosa e usada no desenvolvimento de diversos resultados no campo de filtragem ótima e predição linear, como o desenvolvimento das equações de Wiener-Hopf.

São mostrados os casos de estimação de uma variável aleatória por uma função de outra variável aleatória e também por uma combinação de variáveis aleatórias. A interpretação geométrica destes resultados é salientada. A extensão para processos estocásticos discretos no tempo e o uso de variáveis complexas é apresentado. Este apêndice está baseado nos livros de Papoulis [34] e Therrien [42].

II.2 Estimação de uma variável aleatória por uma função de outra variável aleatória

O problema é estimar a variável aleatória y em termos da função linear $Ax+B$. O critério de otimização é o que minimiza o valor quadrado médio do erro de estimação.

A estimativa de y é dada por :

$$\hat{y} = Ax + B \quad (\text{II.1})$$

E o erro de estimação é dado pela diferença entre o valor desejado e o valor estimado:

$$J = \varepsilon\{e^2\} \quad (\text{II.2})$$

$$J = \varepsilon\{[y - (Ax + B)]^2\} \quad (\text{II.3})$$

Para encontrar o mínimo de J , faz-se as derivadas de J com relação a A e B iguais a zero.

$$\frac{\partial J}{\partial B} = \varepsilon\{[y - (Ax + B)](-1)\} = 0 \quad (\text{II.4})$$

$$\varepsilon\{y - (Ax + B)\} = 0 \quad (\text{II.5})$$

Esta equação ajusta o valor médio da estimativa.

Desenvolvendo a derivada em relação a A:

$$\frac{\partial J}{\partial A} = \frac{\partial}{\partial A} \varepsilon\{[y - (Ax + B)]^2\} \quad (\text{II.6})$$

$$\frac{\partial J}{\partial A} = \varepsilon\left\{\frac{\partial}{\partial A}[y - (Ax + B)]^2\right\} \quad (\text{II.7})$$

Calculando a derivada:

$$\frac{\partial J}{\partial A} = \varepsilon\{2[y - (Ax + B)](-x)\} = 0 \quad (\text{II.8})$$

Igualando a zero e reconhecendo o erro de estimação:

$$\varepsilon\{[y - (Ax + B)]x\} = 0 \quad (\text{II.9})$$

$$\varepsilon\{ex\} = 0 \quad (\text{II.10})$$

A última equação estabelece que o erro de estimação e a variável de entrada são ortogonais na situação onde a estimativa é ótima no sentido do valor quadrado médio. Esta condição é chamada de "Princípio da Ortogonalidade" (também conhecida por "teorema da projeção").

A interpretação geométrica desta situação pode ser vista na figura II.1. Nesta figura a variável aleatória y é representada por um vetor, em geral não-paralelo ao vetor da variável usada para a estimativa, x . A melhor estimativa no sentido de mínimo valor quadrado médio é dada por Ax , e o erro de estimação correspondente é e . Note que e é o menor possível quando $x \perp e$.

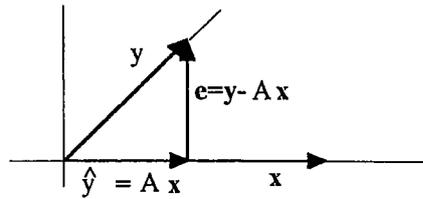


FIGURA II.1: Interpretação geométrica do princípio da ortogonalidade

II.3 Estimação de uma variável aleatória por uma função de N outras variáveis aleatórias

O problema é estimar a variável aleatória s em termos das N variáveis aleatórias x_i , através da função

$$\hat{s} = a_1 x_1 + \dots + a_N x_N \quad (\text{II.11})$$

onde as constantes a_i são escolhidas de forma que o valor quadrado médio do erro de estimação seja mínimo. Para a minimização de J , faz-se a derivada da função em relação a cada coeficiente igual a zero, ou seja.

$$\frac{\partial J}{\partial a_i} = 0 \quad (\text{II.12})$$

para $i=1,2,\dots,N$.

Desenvolvendo, tem-se:

$$\frac{\partial J}{\partial a_i} = \frac{\partial}{\partial a_i} \varepsilon \{ [s - \hat{s}]^2 \} \quad (\text{II.13})$$

Usando II.11:

$$\frac{\partial J}{\partial a_i} = \varepsilon \left\{ \frac{\partial}{\partial a_i} [s - (a_1 x_1 + \dots + a_N x_N)]^2 \right\} \quad (\text{II.14})$$

Calculando a derivada para o i -ésimo termo:

$$\frac{\partial J}{\partial a_i} = \varepsilon \{ 2[s - (a_1 x_1 + \dots + a_N x_N)](-x_i) \} = 0 \quad (\text{II.15})$$

Igualando a zero e reconhecendo o erro de estimação:

$$\varepsilon\{[s - (a_1 x_1 + \dots + a_N x_N)]x_i\} = 0 \quad (\text{II.16})$$

$$\varepsilon\{e x_i\} = 0 \quad (\text{II.17})$$

para $i=1,2,\dots,N$.

A última equação salienta que J é mínimo se o erro de estimação e é ortogonal aos dados de entrada x_i . Esta é a expressão do princípio da ortogonalidade.

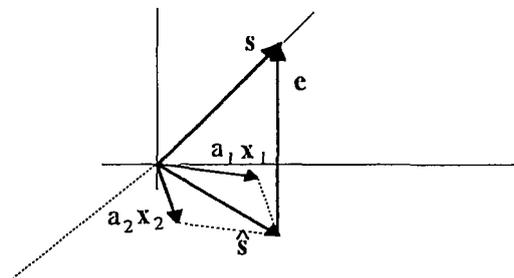


FIGURA II.2: Interpretação geométrica para o caso de $N=2$

A figura acima mostra a interpretação geométrica para o caso de dois coeficientes. Detalhes sobre esta interpretação, que considera variáveis aleatórias como vetores, são dados em Therrien [42].

Para o caso de variáveis aleatórias complexas, o princípio da ortogonalidade pode ser expresso por:

$$\varepsilon\{e x_i^*\} = \varepsilon\{x_i e^*\} = 0 \quad (\text{II.18})$$

para todo i . Esta forma do princípio da ortogonalidade pode ser aplicada diretamente a processos estocásticos discretos no tempo [42] [34].