

AMANDA SOUZA DE PAULA

**SISTEMAS DE SENSORIAMENTO
ESPECTRAL COOPERATIVOS**

São Paulo
2014

AMANDA SOUZA DE PAULA

**SISTEMAS DE SENSORIAMENTO
ESPECTRAL COOPERATIVOS**

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

São Paulo
2014

AMANDA SOUZA DE PAULA

**SISTEMAS DE SENSORIAMENTO
ESPECTRAL COOPERATIVOS**

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

Área de Concentração:
Sistemas Eletrônicos

Orientador:
Prof. Dr. Cristiano Magalhães Panazio

São Paulo
2014

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

São Paulo, de junho de 2014.

Assinatura do autor _____

Assinatura do orientador _____

Catálogo-na-publicação

Paula, Amanda Souza de

Sistemas de sensoriamento espectral cooperativos / A.S. de Paula. - versão corr. - São Paulo, 2014.

146 p.

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

1. Telecomunicações 2. Processamento de sinais 3. Wireless I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Telecomunicações e Controle II.t

Resumo

Esta Tese de Doutorado trata de algoritmos de detecção cooperativa aplicados ao problema de sensoriamento espectral em sistemas de rádios cognitivos. O problema de detecção cooperativa é abordado sob dois paradigmas distintos: detecção centralizada e distribuída. No primeiro caso, considera-se que o sistema conta com um centro de fusão responsável pela tomada de decisão no processo de detecção. Já no segundo caso, considera-se que os rádios cognitivos da rede trocam informações entre si e as decisões são tomadas localmente. No que concerne ao sensoriamento espectral centralizado, são estudados os casos em que os rádios cognitivos enviam apenas um bit de decisão para o centro de fusão (decisão do tipo *hard*) e também o caso em que o detector envia a própria estatística de teste ao centro de fusão (decisão do tipo *soft*). No âmbito de sensoriamento espectral cooperativo com detecção distribuída, são tratados três cenários diferentes. No primeiro, considera-se o caso em que os rádios cognitivos têm conhecimento *a priori* do sinal enviado pelo usuário primário do sistema e do canal entre eles e o usuário primário. No segundo caso, há conhecimento apenas do sinal enviado pelo usuário primário. Já no terceiro, os rádios cognitivos não dispõem de qualquer informação *a priori* do sinal enviado pelo usuário primário. Além do problema de detecção distribuída, a Tese também apresenta um Capítulo dedicado ao problema de estimação, diretamente associado ao de detecção. Esse último problema é abordado utilizando algoritmos derivados da teoria clássica de filtragem adaptativa.

Palavras-chave: Sensoriamento espectral. Rádio cognitivo. Detecção. Estimação. Filtragem adaptativa. Processamento distribuído. Algoritmos de consenso.

Abstract

This Doctorate Thesis deals with cooperative detection algorithms applied to the spectral sensing problem. The cooperative detection problem is approached under two different paradigms: centralized and distributed detection. In the first case, is considered that a fusion center responsible for detection decision is presented in the system. On the other hand, in the second case, is considered that the cognitive radios in the network exchange information among them. Concerning the centralized spectrum sensing system, the case in which the cognitive radios send only one decision bit (hard decision) to the fusion center and the case in which the detector send the statistic test (soft decision) are considered. Regarding the spectrum sensing system with distributed detection, the work analysis three different scenarios. In the first one, where the cognitive radios explore an *a priori* knowledge of the primary user signal and the channel between the primary user and the cognitive radio. In the second one, the cognitive radios use an *a priori* knowledge of only the primary user signal. And, in the last scenario, there is no *a priori* knowledge about the primary user signal. Besides the distributed detection problem, the Thesis also presents a Chapter dedicated to the estimation problem, which is directed related to the detection problem. This last issue is approached using adaptive algorithms derived from the classic adaptive filtering theory.

Keywords: Spectrum sensing. Cognitive radio. Detection. Estimation. Adaptive filtering. Distributed processing. Consensus algorithm.

Agradecimentos

Primeiramente, agradeço profundamente ao Professor e amigo Cristiano Panazio. O Professor Cristiano Panazio me acolheu e confiou no meu potencial, ainda no período de Mestrado. São mais de seis anos de parceria, de aprendizado e amizade. Me sinto privilegiada de ter tido a chance de conhecê-lo e tenho a certeza de que a escolha feita há seis anos atrás, quando resolvi sair da UFPE e começar esse trabalho na Poli-USP, foi a mais acertada possível. Encerro a minha fase de Pós-Graduação tendo consciência de que se trata apenas do começo de uma parceria e de um trabalho em conjunto.

Foram muitos anos na Poli-USP, anos durante os quais tive o prazer de conviver com Professores incríveis que contribuíram enormemente para a minha formação. Gostaria de agradecer, em particular, à Professora Maria Miranda. A Professora Maria sempre me tratou com toda atenção e a admiro por toda dedicação, empolgação e seriedade na Ciência. Também agradeço ao Professor Luiz Antonio Baccalá, com quem tive conversas divertidas e a cada dia aprendia alguma coisa a mais sobre Processos Estocásticos e aos Professores Vitor Nascimento e Magno Silva que sempre me trataram com atenção e que deram contribuições e sugestões valiosas ainda na etapa de Qualificação desse trabalho.

Agradeço aos meus amigos que fizeram com que esse período na Escola Politécnica não se resumisse a dedicação e trabalho. Em especial, agradeço a chance de ter conhecido os amigos Renata Freitas, Luiz Chamon, Wilder Lopes, Murilo Mielle.

A minha família é a grande responsável por todas as minhas conquistas. Tenho consciência que tenho a sorte de ter uma família tão querida, presente e unida. Meus pais, Márcia e Jairo, e meus irmãos, Patrícia e Victor, sempre estiveram ao meu lado, me apoiando com todo o amor e paciência. Meus tios, minhas primas,

avós e avô são meus exemplos de amor e dedicação. Em particular, agradeço à minha prima-irmã Ana Paula. A minha vida em São Paulo ficou muito mais feliz depois de sua chegada. Sem seu apoio, sua companhia e sua alegria a minha fase em São Paulo não teria sido tão divertida.

Finalmente, agradeço e dedico esta Tese a Fernando. Ele esteve presente em todas as fases do trabalho, me animou nas fases mais difíceis e se alegrou comigo nos bons momentos. Fernando chegou com sua doçura, carinho e amor para me tornar mais feliz e realizada.

A todos, muito obrigada.

Lista de Abreviações

ATC	<i>Adapt then Combine</i>
AWGN	Ruído branco e gaussiano (do inglês, <i>Additive White Gaussian Noise</i>)
BSC	<i>Binary Symmetric Channel</i>
BPSK	<i>Binary Phase Shift Keying</i>
CR	Rádio Cognitivo (do inglês, <i>Cognitive Radio</i>)
CTA	<i>Combine then Adapt</i>
DSA	<i>Dynamic Spectrum Access</i>
EMSE	Erro quadrático médio em excesso (<i>Excess Mean Square Error</i>)
FAM	<i>FFT Accumulation Method</i>
FFT	Transformada Rápida de Fourier (do inglês, <i>Fast Fourier Transform</i>)
FCC	<i>Federal Communications Committee</i>
GPS	<i>Global Positioning System</i>
IEEE	<i>Institute of Electrical and Electronics Engineers</i>
LMS	<i>Least Mean Square</i>
LLR	Logarítmo da Razão de Verossimilhança (do inglês, <i>Log Likelihood Ratio</i>)
MF	Filtro Casado (do inglês, <i>Matched Filter</i>)
ML	Máxima Verossimilhança (do inglês, <i>Maximum Likelihood</i>)
MIMO	<i>Multiple Input Multiple Output</i>
MLE	Estimador de Máxima Verossimilhança (do inglês, <i>Maximum Likelihood Estimator</i>)
MMSE	Erro quadrático médio mínimo (do inglês, <i>Minimum Mean Square Error</i>)
MRC	<i>Maximal Ratio Combiner</i>
MSD	Desvio Médio Quadrático, (do inglês, <i>Mean-Squared Deviation</i>)
MSE	Erro quadrático Médio (do inglês, <i>Mean Squared Error</i>)
NP	<i>Neyman-Pearson</i>
PU	Usuário Primário (<i>Primary User</i>)
ROC	Característica de Operação do Receptor (do inglês, <i>Receiver Operating Characteristic</i>)
SM-LMS	<i>Set-membership Least Mean Square</i>
SNR	Razão Sinal-Ruído (do inglês, <i>Signal-to-Noise Ratio</i>)
SU	Usuário Secundário (<i>Secondary User</i>)
WLAN	<i>Wireless Local Area Network</i>

WRAN *Wireless Regional Area Network*

Lista de Símbolos

$\mathbf{A}, \mathbf{B}, \dots$	Matrizes
$\mathbf{a}, \mathbf{b}, \dots$	Vetores coluna
\mathbf{I}_N	Matriz identidade de dimensão N
$\mathbf{1}_N$	Vetor coluna com todos elementos iguais a 1
$\mathbf{0}_N$	Vetor coluna com todos elementos iguais a 0
$(\cdot)^T$	Operador matriz transposta
$(\cdot)^H$	Operador matriz transposta conjugada
\otimes	Operador produto de Kronecker
\mathcal{H}_0	Hipótese nula
\mathcal{H}_1	Hipótese de sinal
\mathbf{H}_k	Matriz de convolução do canal entre o k -ésimo rádio cognitivo e o usuário primário
\mathbf{h}_k	Canal entre o k -ésimo rádio cognitivo e o usuário primário
μ_k	Passo de adaptação do algoritmo LMS aplicado no k -ésimo nó da rede
$\mathbf{u}_k(n)$	Vetor regressor no k -ésimo nó da rede no instante n
$d_k(n)$	Sinal desejado k -ésimo nó da rede no instante n
$\boldsymbol{\psi}_k(n)$	Vetor estimado a partir das estimações locais $\boldsymbol{\phi}_l(n)$ no k -ésimo nó e no instante n
$\boldsymbol{\phi}_k(n)$	Vetor estimado localmente no k -ésimo nó e no instante n
\mathcal{N}_k	Vizinha do nó k
\mathcal{M}	MSD
$\bar{\gamma}$	Limiar utilizado no algoritmo SM-LMS
τ	Constante do algoritmo SM-LMS relacionada com $\bar{\gamma} = \sqrt{\tau\sigma_\eta^2}$
n_k	Cardinalidade de \mathcal{N}_k
P_f	Probabilidade de falso alarme
Q_f	Probabilidade de falso alarme no sistema cooperativo com decisão centralizada
P_m	Probabilidade de perda de detecção
Q_m	Probabilidade de perda de detecção no sistema cooperativo com decisão centralizada
P_e	Probabilidade de erro de um canal BSC
$\mathcal{R}(\cdot)$	Risco de Bayes
$[\cdot]$	Função teto
$Pr\{\mathcal{H}_0\}$	Probabilidade <i>a priori</i> da hipótese nula

$Pr\{\mathcal{H}_1\}$	Probabilidade <i>a priori</i> da hipótese de sinal
$T(\cdot)$	Estatística de teste
\mathcal{L}	Razão de verossimilhança
$\sigma_{\eta,k}^2$	Potência de ruído no k -ésimo nó da rede
γ	SNR
λ	Limiar de decisão do detector
λ_A	Autovalores da matriz \mathbf{A}
$\rho(\mathbf{A})$	Raio espectral da matriz \mathbf{A}

Lista de Figuras

2.1	ROC para canal AWGN, SNR=3dB e $M = 8$	11
2.2	\mathcal{R} em função de λ para diversos valores de α , SNR=3dB e $M = 8$. . .	12
2.3	ROC - canal com desvanecimento Rayleigh, $M = 16$ e SNR=3dB. . . .	14
2.4	Comparação entre os detectores por energia e filtro casado. $M = 8$ e SNR=3dB	15
3.1	Modelo do sistema de fusão centralizada com decisão dura	25
3.2	Função $g(a)$	31
3.3	Campo escalar $\Delta_{1,K}$ considerando $K = 8$, $\alpha = 0,2$, $\beta = 0,8$, $C_f = 0,3$ e $C_m = 0,7$ para diferentes valores de P_e	32
3.4	Risco de Bayes para diferentes valores de n e $K = 4$	34
3.5	Parâmetro n ótimo dado por (3.34).	34
3.6	Comparação entre as regras OU e E com parâmetros $\alpha = 0,2$, $\beta = 0,8$, $C_f = 0,3$ e $C_m = 0,7$	35
3.7	ROC para diferentes valores de n e $K = 4$	36
3.8	Impacto do Risco de Bayes considerando a adoção do critério NP. Os pontos destacados por \star indicam os pontos em que o risco de Bayes é mínimo.	37
3.9	P_f e P_m em função do limiar de decisão λ . Considerando $\gamma = -10$ dB e $M = 32$	40
3.10	P_f e P_m em função do limiar de decisão λ . Considerando $\gamma = 0$ dB e $M = 32$	41
3.11	ROC considerando $M = 32$ e dois valores diferentes de SNR $\gamma = 0$ dB e $\gamma = -10$ dB.	41

3.12	Coeficiente de correlação entre a estatística de teste real e a aproximação linear sob a hipótese de sinal com SNR média $\bar{\gamma} = -3dB$. Os pontos destacados correspondem aos pontos de correlação mínima. Os valores de SNR estão dados em escala linear.	46
3.13	Coeficiente de correlação entre a estatística de teste real e a aproximação linear sob a hipótese de sinal com SNR média $\bar{\gamma} = 6dB$. Os pontos destacados correspondem aos pontos de correlação mínima. Os valores de SNR estão dados em escala linear.	47
3.14	Gráficos de dispersão considerando um sistema com 3 rádios cognitivos e SNR média $\bar{\gamma} = -3dB$	49
3.15	Gráficos de dispersão considerando um sistema com 3 rádios cognitivos e SNR média igual a $\bar{\gamma} = 6dB$	49
3.16	Coeficiente de correlação mínimo em função da SNR média $\bar{\gamma}$ sob a hipótese de sinal.	50
3.17	Curvas de ROC para diversos valores de $\bar{\gamma}$	51
4.1	Raio espectral para as configurações de passo μ_1 e μ_2	60
4.2	Passo de adaptação em cada nó da rede	63
4.3	Raio espectral relativo às matrizes descritas por (4.34) para diferentes valores de K . O valor teórico está indicado pela curva tracejada. . . .	63
4.4	Raio espectral relativo às matrizes descritas por (4.34) para diferentes configurações de passo. O valor teórico está indicado pela curva tracejada.	64
4.5	Raio espectral para as configurações de passo μ_1 e μ_2 para o caso de entrada colorida.	65
4.6	Rede de difusão	66
4.7	Histograma do raio espectral da matriz \mathbf{A} para configurações de passo sorteadas a partir de uma distribuição uniforme com média normalizada $\bar{\mu} = 0,01$	66
4.8	MSD para diferentes configurações de passo de adaptação	68
4.9	Potência do ruído e passo de adaptação em cada nó da rede.	70

4.10	Comparação entre o valor teórico de MSD e o valor dado pela aproximação (4.49) para redes com diferentes número de nós. O passo de adaptação e a potência de ruído em cada nó estão mostrados na Fig.4.9.	70
4.11	Comparação entre o valor teórico de MSD e o valor dado pela aproximação (4.49) para redes com $K = 12$ nós e valores de passos de adaptação diferentes. O passo de adaptação e a potência de ruído em cada nó estão mostrados na Fig.4.9.	71
4.12	Comparação de desempenho entre as versões CTA e ATC do algoritmo LMS, considerando redes com $K = 12$ nós, vários valores de passo médio e matrizes de conectividade dadas por (4.34).	71
4.13	Comparação de desempenho entre as versões CTA e ATC do algoritmo LMS, considerando redes com $K = 36$ nós, vários valores de passo médio e matrizes de conectividade dadas por (4.34).	72
4.14	Rede de difusão	74
4.15	Potência de ruído e passo de adaptação ótimo associado a cada nó da rede mostrada na Fig. 4.14.	75
4.16	Média harmônica estimada.	75
4.17	Passo adaptado considerando a rede mostrada na Fig. 4.14 com potência de ruído mostrada na Fig. 4.15.	76
4.18	Comparação de desempenho entre os algoritmos ATC e CTA para a rede mostrada na Fig. 4.14 e configurações de passo de adaptação e potência de ruído dadas na Fig. 4.15.	77
4.19	MSD para diferentes configurações de passo fixando $\bar{\mu} = 0,01$, entrada branca e rede mostrada na Fig. 4.14.	78
4.20	MSD para diferentes configurações de passo fixando $\bar{\mu} = 0,01$, entrada colorida e rede mostrada na Fig. 4.14.	79
4.21	Rede de difusão com 24 nós.	80
4.22	Potência de ruído e passo de adaptação ótimo associado a cada nó da rede mostrada na Fig. 4.21.	80

4.23	MSD para diferentes configurações de passo fixando $\bar{\mu} = 0,01$, entrada branca e rede mostrada na Fig. 4.21.	81
4.24	MSD para diferentes configurações de passo fixando $\bar{\mu} = 0,01$, entrada colorida e rede mostrada na Fig. 4.21.	81
4.25	Comparação de desempenho entre os algoritmos LMS e SM-LMS, considerando que a potência de ruído em cada nó da rede é a mesma.	84
4.26	Comparação de desempenho entre os algoritmos LMS e SM-LMS, considerando que a potência de ruído em cada nó da rede é a dada por (4.64).	85
4.27	Probabilidade de atualização em função da potência de ruído considerando que a potência de ruído em cada nó da rede é a dada por (4.64).	85
4.28	Comparação de desempenho entre os algoritmos LMS, SM-LMS e sua versão com <i>feedback</i> reduzido RF-SM-LMS.	87
5.1	Rede com $K = 6$ rádios cognitivos	93
5.2	Comparação de desempenho em termos de MSD entre o sistema com passos de adaptação iguais em todos os nós e o sistema com passo otimizado	93
5.3	Comparação de desempenho de detecção entre o sistema com passos de adaptação iguais em todos os nós e o sistema com passo otimizado	94
5.4	Comparação entre a densidade de probabilidade teórica dada na Proposição 5.2.1, considerando $\sigma_{\eta}^2 = 1$, $M = 1$ e vários valores de μ , e o histograma gerado a partir de $N_r = 10^5$ realizações da variável aleatória X sob a hipótese nula.	101
5.5	Comparação entre a densidade de probabilidade teórica dada na Proposição 5.2.1, considerando $M = 1$, $\mu = 0,1$ e vários valores de σ_{η}^2 , e o histograma gerado a partir de $N_r = 10^5$ realizações da variável aleatória X sob a hipótese nula.	102

5.6	Comparação entre a densidade de probabilidade teórica dada na Proposição 5.2.1, considerando $\sigma_\eta^2 = 1$, $\mu = 0,1$, e vários valores de M e o histograma gerado a partir de $N_r = 10^5$ realizações da variável aleatória X sob a hipótese nula.	102
5.7	Comparação entre a densidade de probabilidade teórica dada na Proposição 5.2.2, considerando $\sigma_\eta^2 = 1$, $M = 1$, e vários valores de μ , e o histograma gerado a partir de $N_r = 10^5$ realizações da variável aleatória X sob a hipótese de sinal.	103
5.8	Comparação entre a densidade de probabilidade teórica dada na Proposição 5.2.2, considerando $M = 1$, $\mu = 0,1$ e vários valores de σ_η^2 , e o histograma gerado a partir de $N_r = 10^5$ realizações da variável aleatória X sob a hipótese de sinal.	103
5.9	Comparação entre a densidade de probabilidade teórica dada na dada na Proposição 5.2.2, considerando $\sigma_\eta^2 = 1$, $\mu = 0,1$ e vários valores de M , e o histograma gerado a partir de $N_r = 10^5$ realizações da variável aleatória X sob a hipótese de sinal.	104
5.10	Comparação entre as probabilidade de falso alarme teóricas e simuladas, considerando $\mu = 0 = 0,1$, $M = 1$ e duas configurações de potência de ruído diferentes: $\sigma_\eta^2 = 1$ e $\sigma_\eta^2 = 0,5$	105
5.11	Comparação entre as probabilidade de perda de detecção alarme teóricas e simuladas, considerando $\mu = 0 = 0,1$, $M = 1$ e duas configurações de potência de ruído diferentes: $\sigma_\eta^2 = 1$ e $\sigma_\eta^2 = 0,5$	105
5.12	Convergência de P_f e P_m , considerando $\sigma_\eta^2 = 1$ e $\lambda = 0,4$	106
5.13	Desempenho do detector LMS considerando o canal dado por (5.67), limiar $\lambda = 0,15$, passo de adaptação $\mu = 0,05$ e $\sigma_\eta^2=2$	110
5.14	Desempenho do detector LMS considerando o canal dado por (5.67), limiar $\lambda = 0,03$, passo de adaptação $\mu = 0,005$ e $\sigma_\eta^2=2$	110
5.15	Desempenho do detector LMS considerando o canal dado por (5.67), limiar $\lambda = 0,0003$, passo de adaptação $\mu = 0,0001$ e $\sigma_\eta^2=2$	111

5.16	Desempenho do detector LMS considerando o canal dado por $h = 1$, limiar $\lambda = 0,0003$, passo de adaptação $\mu = 0,001$ e $\sigma_{\eta}^2=2$	111
5.17	Comparação de desempenho entre os sistemas de detecção usando o algoritmo LMS configurado com diferentes passos de adaptação, o detector ótimo ML, o detector sub-ótimo dado em (5.25) e o detector de energia.	112
5.18	Curvas de ROC para o sistema cooperativo em comparação com os rádios cognitivos operando individualmente.	114
5.19	Rede de rádios cognitivos	117
5.20	SNR em cada um dos nós da rede apresentada na Fig. 5.19	117
5.21	Estimativa ponderada de energia do teste sub-ótimo em cada nó da rede apresentada na Fig. 5.19. As estimativa em cada um dos nós está representada com uma cor diferente na figura.	118
5.22	Gráfico de dispersão entre $\lambda(K - 1)$ e α_{opt} para $K = 12$. Curva de ajuste dada por: $\alpha_{opt} = \exp(4,7156 \lambda(K - 1) - 4,6749)$	121
5.23	Gráfico de dispersão entre $\lambda(K - 1)$ e $\log(\alpha_{opt})$ para $K = 12$, $K = 24$ e $K = 36$	121
5.24	Curvas de dispersão entre $\lambda(K - 1)$ e $\log(\alpha_{opt})$ para diferentes ordens do filtro de predição e redes com $K = 12$ nós.	122
5.25	Comparação entre as estimativas de α_{opt} obtidas por ajuste linear e os valores ótimos obtidos experimentalmente para a rede apresentada na 5.19. Também é mostrado, na curva pontilhada, o raio espectral para o preditor de ordem $M_p = 2$	123
5.26	Comparação entre o desempenho do algoritmo de consenso e de sua versão com convergência acelerada na rede apresentada na 5.19. . . .	123
5.27	Convergência de estimativa de energia ponderada no nó 1 da rede da rede apresentada na Fig. 5.19.	124
5.28	ROC do sistema de rádio cognitivo distribuído com rede apresentada na Fig. 5.19.	125

Sumário

1	Introdução	1
1.1	Principais contribuições do Doutorado	3
1.2	Organização da Tese	5
2	Sensoriamento Espectral	7
2.1	Detecção por energia	8
2.1.1	Sinal com desvanecimento	13
2.2	Detector por filtro casado	14
2.3	Detector por ciclo-estacionaridade	16
2.4	Detector por análise de autovalores	18
2.5	Conclusões	21
3	Sensoriamento Espectral Cooperativo Centralizado	23
3.1	Fusão centralizada utilizando decisão dura	24
3.1.1	Risco de Bayes	26
3.1.2	Resultados de simulação	33
3.2	Fusão centralizada usando decisão suave	37
3.2.1	Validade da aproximação gaussiana	39
3.2.2	Uma outra interpretação da fusão linear	42
3.3	Conclusões	51
4	Estimação Distribuída	53
4.1	Topologias de rede	54
4.2	LMS distribuído	55
4.2.1	Otimização da escolha do passo de adaptação	57

4.2.2	Configuração ótima de passo de adaptação	67
4.2.3	Estimação do passo de adaptação ótimo	73
4.2.4	Resultados de simulação	76
4.3	<i>Set-membership</i> distribuído	82
4.3.1	<i>Set-membership</i> distribuído com redução de <i>feedback</i>	86
4.4	Conclusões	87
5	Detecção Distribuída	89
5.1	Técnicas baseadas em filtro casado	89
5.2	Detecção sem conhecimento perfeito de canal	94
5.2.1	Sensoriamento local	95
5.2.2	Detector cooperativo	113
5.3	Detecção de energia usando o método do consenso	115
5.3.1	O algoritmo de consenso	115
5.4	Conclusões	124
6	Conclusões e Perspectivas	127
	Referências Bibliográficas	129
A	Demonstração da expressão para n_{opt}	139
B	Demonstração da aproximação para o MSD	141
B.1	Algoritmo LMS-CTA	141
B.2	Algoritmo LMS-ATC	143
C	Publicações Relevantes	145

Capítulo 1

Introdução

Ultimamente, a demanda nos serviços de transmissão de dados em sistemas sem fio tem crescido consideravelmente. Isto pode ser comprovado pelo fato do número de aparelhos celulares já superar o número de linhas fixas em muitos países, e os serviços de dados WLAN WiFi (IEEE 802.11) estarem sendo instalados em vários lares e estabelecimentos. A fim de suprir tal demanda, várias técnicas que permitem aumentar a eficiência da utilização do espectro foram desenvolvidas. As mais relevantes são a técnica MIMO (*Multiple-Input Multiple-Output*), que usa diversidade espacial para aumentar a capacidade, e os *turbo-codes*, códigos que aumentam a eficiência e a robustez do sistema.

Contudo, essas técnicas só permitem melhorar a capacidade no momento da utilização do espectro, que é limitado tanto fisicamente (condições de propagação) como legalmente (licenças). Um acesso não controlado ao espectro pode gerar demasiada interferência e impossibilitar a comunicação. Deste modo, as agências reguladoras fazem a segmentação do espectro em faixas, destinadas única e especificamente para uma certa aplicação, ou a usuários licenciados.

Tal cenário de alocação, essencialmente estático, gera um aproveitamento pouco eficaz do espectro, assim como mostrado em [1]. Esse fato motiva a busca por procedimentos mais eficientes de alocação espectral.

Nesse contexto, surgem as técnicas de alocação dinâmica de espectro (DSA, do inglês, *Dynamic Spectrum Access*), nas quais a alocação de determinada faixa do espectro não se restringe ao usuário que detém sua licença. Para tanto, os

possíveis usuários da faixa de frequências são classificados em usuários primários e secundários. Os usuários primários são aqueles que detêm a licença e possuem, portanto, prioridade no uso do espectro. Os usuários secundários, por sua vez, podem utilizar a faixa de espectro desde que não interfiram no usuário primário.

Para que tais técnicas possam ser implementadas, é necessário que os sistemas de rádio sejam dotados de propriedades especiais, dando origem à ideia do rádio cognitivo. A definição mais usual de rádio cognitivo é aquela proposta por Mitola em [2], em que é dito que rádio cognitivo é o rádio que consegue identificar as condições do ambiente e a elas se adaptar de forma inteligente.

De acordo com a definição de Mitola, o processo de aplicação do rádio cognitivo pode ser entendido através do ciclo de cognição [3]. No ciclo de cognição, o sistema cognitivo passa pelos seguintes estados: observação, orientação, decisão, ação, planejamento e aprendizado. O rádio cognitivo, portanto, deve, a princípio, ser capaz de observar o meio que o cerca para, a partir de então, utilizar seus recursos da melhor forma possível. Como componente essencial para a implementação do ciclo de cognição, se encontram as técnicas de sensoriamento espectral. São essas ferramentas que tornam as etapas de observação e decisão factíveis.

Ao sensoriamento espectral confiável das faixas disponíveis, está associada uma série de dificuldades advindas dos fenômenos de propagação em sistemas de comunicação sem fio. Caso tal detecção seja feita de forma autônoma, *i.e.*, a informação sobre a ocupação espectral seja realizada por um único rádio cognitivo, tais fenômenos de propagação podem ocasionar o problema de nó oculto (*hidden node*) [4]. Nesse caso, é possível que usuários primários e secundários estejam escondidos de outros usuários secundários, ocasionando interferência. Com o objetivo de contornar esse problema, é possível que usuários secundários cooperem, trocando informações de sensoriamento. Essa abordagem melhora significativamente a qualidade da detecção de faixas não utilizadas [4], [5].

O interesse na temática de rádio cognitivo, entretanto, não é restrito à comunidade acadêmica. Em julho de 2011, o IEEE, em conjunto com o FCC (*Federal Communications Commission*), publicou um novo padrão, o IEEE 802.22 [6], [7],

para implementar redes WRAN (*wireless regional area network*) utilizando os buracos espectrais disponíveis em locais de baixa densidade demográfica na faixa de canais de TV.

Apesar dos esforços em propor técnicas de sensoriamento espectral cada vez mais confiáveis, tais métodos ainda não são suficientemente robustos como indicado em [8], [9]. A falta de robustez de tais métodos, inclusive, afetou o desenvolvimento do padrão IEEE 802.22. Como as informações provenientes das redes de sensoriamento espectral ainda não são suficientemente confiáveis, adota-se um sistema de GPS (*Global Positioning System*) para que cada usuário secundário informe sua localização geográfica e, a partir de então, uma rede centralizadora possa indicar se o espectro está disponível na região do usuário [7].

Com o objetivo de propor métodos de sensoriamento espectral mais confiáveis, apresenta-se esta Tese de Doutorado focada na questão do processamento cooperativo entre os nós da rede formada por rádios cognitivos. Ao longo do documento, serão apresentadas as diferentes técnicas de sensoriamento espectral e as possíveis formas de troca de informação entre os rádios cognitivos. Além do problema de detecção distribuída, diretamente ligado à questão de sensoriamento espectral cooperativo, o trabalho apresentado nesta Tese de Doutorado também analisa a questão de estimação distribuída. O interesse no problema de estimação vem do fato de esse problema ser intrinsecamente relacionado ao problema de detecção. Inclusive, em muitos casos, a estimação constitui um passo intermediário para a detecção.

Também é interessante ressaltar que as análises e desenvolvimentos aqui apresentados podem ser estendidos para redes de sensores aplicadas em qualquer contexto, não sendo restritos ao caso do rádio cognitivo.

1.1 Principais contribuições do Doutorado

1. **Estabelecimento de limitantes no contexto de sensoriamento espectral cooperativo com troca de informação dura em canal imperfeito.**

Como detalhado na Seção 3.1, foram estudados sistemas de sensoriamento espectral nos quais os rádios cognitivos trocam entre si informação dura.

Analisou-se o impacto que canais sujeitos a perda de informação geram na decisão final da rede para as diferentes regras de fusão. Os resultados dessa análise foram publicados em [10].

2. Análise de sistemas de sensoriamento espectral cooperativo com troca de decisão suave.

Na Seção 3.2, o caso de sensoriamento espectral cooperativo com troca de informação suave foi analisado. Como será detalhado, diferentemente do paradigma gaussiano usualmente explorado na literatura, neste trabalho é analisado o teste derivado da razão de verossimilhança gerada a partir das distribuições de probabilidade exatas do sinal recebido em cada rádio cognitivo da rede.

3. Contribuições no campo de filtragem adaptativa distribuída.

No Capítulo 4, foi estudado o algoritmo LMS distribuído, bem como sua versão *set-membership*, no contexto em que os rádios cognitivos apresentam diferentes potências de ruído. Alguns resultados dessa análise já foram publicados em [11] e [12].

4. Aplicação de técnicas de filtragem adaptativa no caso de sensoriamento espectral sem conhecimento de canal.

Na Seção 5.2, é apresentada uma técnica de sensoriamento espectral cooperativa na qual se assume que há conhecimento do sinal transmitido pelo usuário primário, mas não há conhecimento do canal entre os dois usuários. Como detalhado na Seção 5.2, a técnica proposta apresenta complexidade computacional reduzida e desempenho próximo ao detector ideal. São apresentadas expressões teóricas para as probabilidades de falso alarme e perda de detecção para o caso em que o sinal transmitido pelo usuário primário é do tipo BPSK (*Binary Phase Shift Keying*).

5. Utilização de técnicas de consenso acelerado para sensoriamento espectral sem conhecimento *a priori* do canal e do sinal transmitido pelo usuário primário.

Na seção 5.3 é analisado o caso de sensoriamento espectral no qual não há

qualquer conhecimento *a priori* do sinal enviado pelo usuário primário. Neste contexto, o algoritmo de consenso é analisado e é proposta uma forma de estimar os coeficientes de aceleração do algoritmos a partir de [13].

1.2 Organização da Tese

A Tese está organizada nos seguintes capítulos principais:

- **Capítulo 2 - Sensoriamento Espectral**

Neste capítulo são apresentadas as principais técnicas de sensoriamento espectral. São detalhados os métodos de detecção por energia, por filtro casado, cicloestacionariedade e por auto-valores.

- **Capítulo 3 - Sensoriamento Espectral Cooperativo Centralizado**

Já neste capítulo é realizada uma análise do sistema de sensoriamento espectral centralizado no qual o rádio cognitivo envia decisões do tipo dura ao centro de fusão. Tal análise é realizada no contexto em que o envio de decisões ao centro de fusão está sujeito a erros. Nesse cenário, são derivados limitantes de desempenho do sistema em relação ao critério Bayesiano de otimização.

- **Capítulo 4 - Estimação Distribuída**

Este capítulo é dedicado à análise de técnicas de estimação distribuídas baseadas em algoritmos clássicos de filtragem adaptativa. É derivada uma expressão que permite atribuir os passos de adaptação em cada nó da rede a fim de garantir a minimização do erro quadrático médio de estimação. Também é analisado o comportamento de algoritmos do tipo *set-membership* no contexto de redes adaptativas e como eles podem ser utilizados para diminuir o consumo de energia na rede. Os resultados gerados neste capítulo serão aplicados no problema de detecção em rádio cognitivo analisado no capítulo seguinte.

- **Capítulo 5 - Detecção Distribuída**

Este capítulo dedica-se ao estudo de técnicas de detecção distribuídas. São analisados três cenários distintos de detecção. No primeiro, considera-se que o

rádio cognitivo tem conhecimento do sinal enviado pelo usuário primário e do canal entre eles, nestas condições pode aplicar técnicas do tipo filtro casado e utilizar os resultados gerados no capítulo 4. No segundo cenário, considera-se que há conhecimento do sinal enviado pelo usuário primário porém não há conhecimento do canal. Nessas condições, os rádios cognitivos estimam o canal aplicando algoritmos adaptativos e, posteriormente, aplicam técnicas de detecção por filtro casado. Finalmente, é explorado o caso mais geral de detecção em que os rádios cognitivos não dispõem de qualquer informação *a priori* do sinal enviado pelo usuário primário.

Capítulo 2

Sensoriamento Espectral

O problema de sensoriamento espectral constitui uma questão fundamental em sistemas de rádio cognitivo. É através do sensoriamento espectral que os usuários secundários podem detectar as bandas disponíveis para transmissão.

O sensoriamento espectral é formulado a partir do seguinte teste de hipóteses:

$$\begin{cases} \mathcal{H}_0 : x(n) = \eta(n) \\ \mathcal{H}_1 : x(n) = s(n) + \eta(n), \end{cases} \quad (2.1)$$

em que $x(n)$ é o sinal recebido pelo usuário secundário, $s(n)$ é o sinal transmitido pelo usuário primário e $\eta(n)$ é o ruído aditivo branco gaussiano (AWGN, do inglês, *additive white Gaussian noise*).

A partir do sinal recebido, uma dada estatística de teste é aplicada para a formulação do teste de hipóteses. É necessário escolher uma estatística de teste que gere um sinal com as menores probabilidades de perda de detecção (P_m) e probabilidade de falso alarme (P_f), definidas a seguir:

$$P_m = Pr \left\{ \hat{\theta} = \mathcal{H}_0 | \mathcal{H}_1 \right\}, \quad (2.2)$$

$$P_f = Pr \left\{ \hat{\theta} = \mathcal{H}_1 | \mathcal{H}_0 \right\}, \quad (2.3)$$

em que $\hat{\theta}$ é a decisão gerada pela estatística de teste e $Pr\{\cdot\}$ denota probabilidade.

Há um compromisso entre a minimização de P_m e de P_f . Não é possível escolher um algoritmo que minimize simultaneamente P_f e P_m , de modo que é necessário impor outros critérios de otimização. Uma alternativa consiste em minimizar o erro

total, dado por:

$$P_e = Pr\{\mathcal{H}_0\}P_f + Pr\{\mathcal{H}_1\}P_m, \quad (2.4)$$

em que $Pr\{\mathcal{H}_0\}$ e $Pr\{\mathcal{H}_1\}$ correspondem às probabilidades *a priori* de cada uma das hipóteses. Essa abordagem corresponde à minimização do erro segundo o critério de Bayes [14].

Por fim, uma outra possibilidade é a de se utilizar o critério de Neyman Pearson e minimizar P_m condicionado a um certo valor de P_f [14].

2.1 Detecção por energia

Dentre os métodos de sensoriamento espectral, o método de detecção por energia é o mais utilizado. A principal vantagem do uso do sensoriamento espectral através da detecção de energia é a de que não é necessário qualquer conhecimento *a priori* do sinal primário.

A estatística de teste aplicada nesse caso é dada pela estimativa da energia do sinal recebido:

$$T(x) = \frac{1}{\sigma_\eta^2} \sum_{k=0}^{M-1} |x(k)|^2, \quad (2.5)$$

em que σ_η^2 é a variância do ruído $\eta(k)$, a qual é assumida como parâmetro conhecido. Nos casos em que há erro de estimação de σ_η^2 , percebe-se uma perda de desempenho do detector detalhada em [15].

A decisão é então tomada a partir da comparação da estatística de teste com o limiar de decisão λ :

$$\hat{\theta} = \begin{cases} \mathcal{H}_0, & T(x) \leq \lambda \\ \mathcal{H}_1, & T(x) > \lambda \end{cases}. \quad (2.6)$$

Sob a hipótese \mathcal{H}_0 , analisando (2.1) temos que $x(k) = \eta(k)$, portanto:

$$T(x)_{\mathcal{H}_0} = \frac{1}{\sigma_\eta^2} \sum_{k=0}^{M-1} |\eta(k)|^2, \quad (2.7)$$

Considerando que $\eta(k)$ corresponde a amostras de ruído gaussiano representado em

banda base complexa, (2.7) pode ser reescrita como:

$$T(x)_{\mathcal{H}_0} = \sum_{k=0}^{M-1} \left| \frac{\eta_I(k) + j\eta_Q(k)}{\sigma_\eta} \right|^2 = \sum_{k=0}^{M-1} \left(\frac{\eta_I(k)}{\sigma_\eta} \right)^2 + \sum_{k=0}^{M-1} \left(\frac{\eta_Q(k)}{\sigma_\eta} \right)^2. \quad (2.8)$$

em que $\eta_I(k)$ e $\eta_Q(k)$ correspondem, respectivamente, aos termos em fase e quadratura de $\eta(k)$.

A eq. (2.8) mostra que a estatística de teste $T(x)_{\mathcal{H}_0}$ corresponde à soma de $2M$ variáveis aleatórias gaussianas ao quadrado normalizadas. Dessa forma, $T(x)_{\mathcal{H}_0}$ é uma variável aleatória do tipo chi-quadrado centralizada com $2M$ graus de liberdade, a qual será denotada por χ_{2M}^2 . Sendo assim, a probabilidade de falso alarme pode ser expressa como:

$$P_f = Pr \{T(x) > \lambda | \mathcal{H}_0\} = Pr \{T(x)_{\mathcal{H}_0} > \lambda\} = 1 - F_{\chi_{2M}^2}(\lambda), \quad (2.9)$$

em que $F_{\chi_{2M}^2}(x)$ é a função distribuição de probabilidade de χ_{2M}^2 [14], dessa forma:

$$P_f = \frac{\Gamma(M, \frac{\lambda}{2})}{\Gamma(M)}, \quad (2.10)$$

$\Gamma(x, y)$ é a função gama incompleta superior [16], dada por:

$$\Gamma(x, y) = \int_x^\infty t^{y-1} e^{-t} dt \quad (2.11)$$

e $\Gamma(M)$ é a função gama, dada por:

$$\Gamma(M) = \int_0^\infty t^{M-1} e^{-t} dt. \quad (2.12)$$

no caso em que $M \in \mathbb{Z}$, a função gama se reduz a:

$$\Gamma(M) = (M-1)!. \quad (2.13)$$

Já sob a hipótese \mathcal{H}_1 , a estatística de teste pode ser escrita como:

$$T(x)_{\mathcal{H}_1} = \sum_{k=0}^{M-1} \left(\frac{\eta_I(k) + s_I(k)}{\sigma_\eta} \right)^2 + \sum_{k=0}^{M-1} \left(\frac{\eta_Q(k) + s_Q(k)}{\sigma_\eta} \right)^2. \quad (2.14)$$

Nota-se que $T(x)_{\mathcal{H}_1}$ corresponde à soma de $2M$ variáveis aleatórias gaussianas ao quadrado. Cada uma dessas variáveis aleatórias gaussianas tem variância unitária e

média $\mu_k = \frac{s_I(k)}{\sigma_\eta}$ ou $\mu_k = \frac{s_Q(k)}{\sigma_\eta}$.

Portanto, $T(x)_{\mathcal{H}_1}$ é uma variável aleatória χ^2 com $2M$ graus de liberdade e parâmetro de descentralização:

$$\mu = \frac{\sum_{k=0}^{M-1} s_I^2(k)}{\sigma_\eta^2} + \frac{\sum_{k=0}^{M-1} s_Q^2(k)}{\sigma_\eta^2}. \quad (2.15)$$

Como $\sigma_\eta^2 = \text{var} \{ \eta_I^2(k) \} = \text{var} \{ \eta_Q^2(k) \}$, μ pode ser reescrito como:

$$\mu = 2M\gamma, \quad (2.16)$$

em que γ é a SNR média por amostra por dimensão complexa:

$$\gamma = \frac{\mathbb{E} \{ s_I^2(k) \}}{\sigma_\eta^2} = \frac{\mathbb{E} \{ s_Q^2(k) \}}{\sigma_\eta^2}. \quad (2.17)$$

e $\mathbb{E} \{ \cdot \}$ denota o operador de esperança matemática.

A probabilidade de perda de detecção pode ser então determinada:

$$P_m = Pr \{ T(x) < \lambda | \mathcal{H}_1 \} = Pr \{ T(x)_{\mathcal{H}_1} < \lambda \} = F_{\chi_{2\gamma, 2M}^2}(\lambda), \quad (2.18)$$

em que $F_{\chi_{\mu, 2M}^2}(x)$ denota a função distribuição de probabilidade de uma variável aleatória χ^2 descentralizada com parâmetro de descentralização μ e $2M$ graus de liberdade [14]:

$$P_m = 1 - Q_M \left(\sqrt{2\gamma M}, \sqrt{\lambda} \right) \quad (2.19)$$

em que $Q_M(x, y)$ é a função Marcum Q generalizada [14], dada por:

$$Q_M(x, y) = \int_y^\infty t \left(\frac{t}{x} \right)^{M-1} e^{-\frac{t^2+x^2}{2}} I_{M-1}(xt) dt \quad (2.20)$$

e $I_n(x)$ é a função de Bessel modificada de primeira espécie e ordem n :

$$I_n(x) = \frac{1}{2\pi j} \oint_C e^{(x/2)(t-1/t)} t^{-n-1} dt. \quad (2.21)$$

a curva de integração C pode ser definida como qualquer caminho fechado que englobe a origem.

As eqs. (2.10) e (2.19) indicam que P_f é uma função decrescente do limiar de decisão λ , enquanto que P_m é uma função crescente do parâmetro λ , respectivamente.

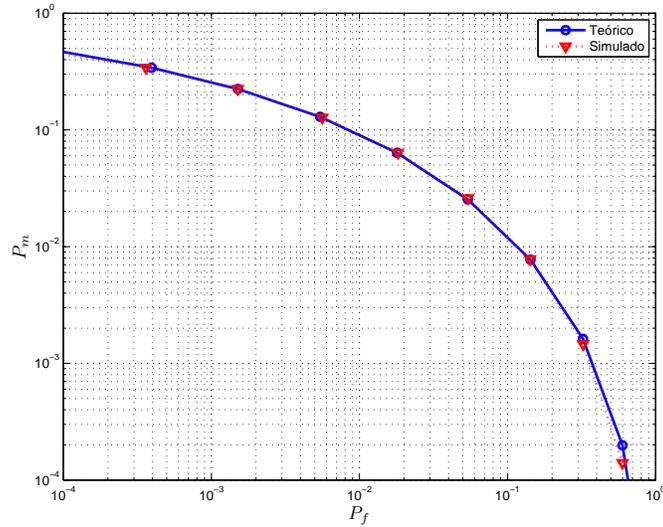


Figura 2.1: ROC para canal AWGN, SNR=3dB e $M = 8$.

A Fig. 2.1 mostra P_m em função de P_f calculadas para um valor fixado de SNR, um certo número de amostras e para vários valores de λ . Esse tipo de curva em que P_m é representada como função de P_f é denominado por ROC (do inglês, *Receiver Operating Characteristics*). A partir da análise de ROC, pode-se perceber o compromisso entre P_f e P_m .

As curvas de ROC são úteis quando há interesse em determinar qual P_f leva a uma dada P_m alvo, ou vice-versa. Entretanto, pode haver situações em que o objetivo seja minimizar a probabilidade de erro total do sistema. Desse modo, definimos o risco de Bayes ou erro total, \mathcal{R} como [17]:

$$\mathcal{R} = \alpha P_f + (1 - \alpha) P_m, \quad (2.22)$$

em que α é uma constante de ponderação que indica o impacto dos erros por falso alarme e perda de detecção no sistema. Em geral, $\alpha < 1/2$, uma vez que uma perda de detecção pode levar à interferência do usuário secundário no usuário primário, o que pode ser mais danoso do que uma perda de oportunidade de uso do buraco espectral pelo usuário secundário gerada por um erro de falso alarme.

A Fig. 2.2 mostra \mathcal{R} em função do limiar de decisão λ para diversos valores de α , considerando ainda SNR=3dB e $M = 8$. Note que o valor de λ que minimiza \mathcal{R} é uma função crescente de α . Essa não é uma constatação inesperada, já que

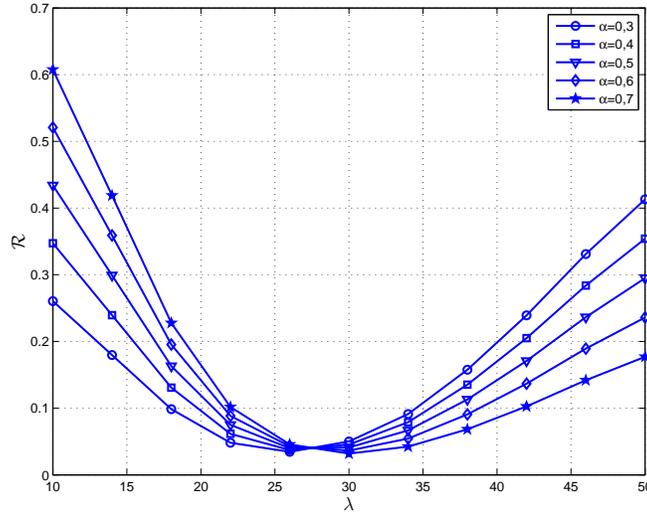


Figura 2.2: \mathcal{R} em função de λ para diversos valores de α , SNR=3dB e $M = 8$

quanto maior o valor de α , maior será a contribuição de P_f para a probabilidade de erro total, \mathcal{R} . Dessa forma, como P_f é função crescente de λ , o valor de λ que minimiza \mathcal{R} também será função crescente de α . Também é interessante notar que se $P_m = P_f$, $\mathcal{R} = P_m$, independentemente do valor de α . No caso da Fig. 2.2, essa situação corresponde a $\lambda = 27,5$.

Com o objetivo de determinar o valor do limiar λ que minimiza o risco de Bayes, deriva-se e iguala-se a zero a eq. (2.22) em relação à variável λ :

$$\frac{\partial}{\partial \lambda} \mathcal{R} = \alpha \frac{\partial}{\partial \lambda} P_f + (1 - \alpha) \frac{\partial}{\partial \lambda} P_m, \quad (2.23)$$

portanto $\frac{\partial}{\partial \lambda} \mathcal{R} = 0$ implica

$$-\frac{\frac{\partial}{\partial \lambda} P_m}{\frac{\partial}{\partial \lambda} P_f} = \frac{\alpha}{1 - \alpha}, \quad (2.24)$$

sendo $P_m = F_{\chi_{2\gamma, 2M}^2}(\lambda)$ e $P_f = 1 - F_{\chi_{2M}^2}(\lambda)$, temos que as derivadas de P_m e P_f são dadas pelas funções densidades de probabilidade:

$$\frac{f_{\chi_{2\gamma, 2M}^2}(\lambda)}{f_{\chi_{2M}^2}(\lambda)} = \frac{\alpha}{1 - \alpha}. \quad (2.25)$$

Substituindo as expressões das funções densidade de probabilidade, conclui-se que o limiar de decisão ótimo (λ_{opt}) pode ser determinado analiticamente, encontrando-

se a solução da seguinte equação:

$$\frac{2^{M-1}e^{-\gamma}}{(2\gamma\lambda_{\text{opt}})^{\frac{M-1}{2}}}\Gamma(M)I_{M-1}\left(\sqrt{2\gamma\lambda_{\text{opt}}}\right) = \frac{\alpha}{1-\alpha}. \quad (2.26)$$

2.1.1 Sinal com desvanecimento

Até então, tratou-se de um caso simples em que o sinal recebido pelo usuário primário é apenas corrompido com ruído AWGN. Entretanto, em geral, o sinal recebido também pode sofrer desvanecimento. Nesse caso, o teste de hipóteses resultante é:

$$\begin{cases} \mathcal{H}_0 : x(n) = \eta(n) \\ \mathcal{H}_1 : x(n) = h(n)s(n) + \eta(n) \end{cases}, \quad (2.27)$$

em que $h(n) \sim \mathcal{N}(0,1)$.

Aplicando a mesma estatística de teste $T(x)$, a probabilidade de perda de detecção é dada por [4], [18], [19]:

$$P_m = 1 - e^{-\frac{\lambda}{2}} \sum_{l=0}^{M-2} \frac{\left(\frac{\lambda}{2}\right)^l}{l!} + \left(\frac{1+2M\gamma}{2M\gamma}\right)^{M-1} \left(e^{-\frac{\lambda}{2+4M\gamma}} - e^{-\frac{\lambda}{2}} \sum_{l=0}^{M-2} \frac{\left(\frac{2M\lambda\gamma}{2+4M\gamma}\right)^l}{l!} \right). \quad (2.28)$$

Evidentemente, a probabilidade de falso alarme não depende da distribuição do sinal transmitido pelo usuário primário, uma vez que ela é determinada sob a hipótese de ausência do sinal do usuário primário e dada por (2.10).

A coerência entre a expressão teórica apresentada em (2.28) e os resultados de simulação está mostrada na Fig. 2.3.

Apesar da simplicidade envolvida na detecção através da estimativa de energia do sinal, o desempenho desse detector, se comparado com as demais técnicas, pode não ser satisfatório, em especial em casos de baixa SNR e incerteza na determinação da potência de ruído [15].

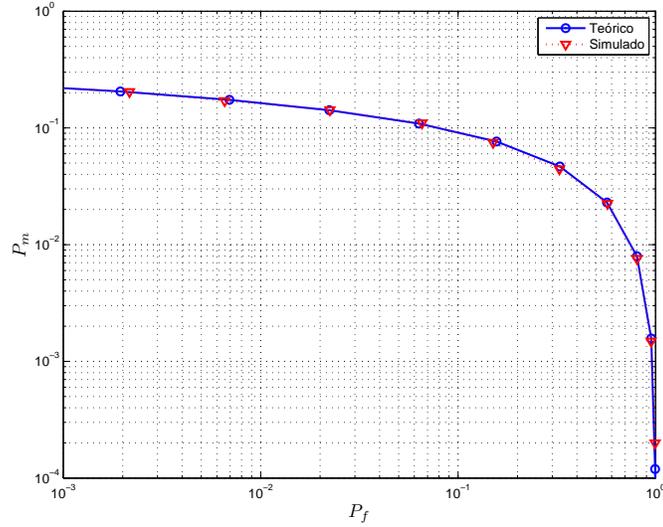


Figura 2.3: ROC - canal com desvanecimento Rayleigh, $M = 16$ e $\text{SNR}=3\text{dB}$.

2.2 Detector por filtro casado

Uma alternativa ao detector por energia é a detecção utilizando o filtro casado, [20, 21]. Nesse caso a estatística de teste é dada por:

$$T(x) = \text{Re} \left\{ \sum_{k=0}^{M-1} x^*(k)s(k) \right\}. \quad (2.29)$$

Com o objetivo de comparar essa técnica com o detector de energia, consideremos o caso do canal AWGN. Nessa situação, sob a hipótese de presença de sinal, a estatística T seria dada por:

$$T(x)_{\mathcal{H}_1} = M\sigma_s^2 + \sum_{k=0}^{M-1} \eta^*(k)s(k). \quad (2.30)$$

Dada a gaussianidade de $\eta(k)$, $T_{\mathcal{H}_1}$ tem distribuição gaussiana de média $\mu = M\sigma_s^2$ e variância $\sigma^2 = M\sigma_s^2\sigma_\eta^2$. Portanto, a probabilidade de perda de detecção seria dada por:

$$P_m = Pr \{T(x) < \lambda | \mathcal{H}_1\} = Pr \{T(x)_{\mathcal{H}_1} < \lambda\} = 1 - Q \left(\frac{\lambda - M\sigma_s^2}{\sigma_s\sigma_\eta\sqrt{M}} \right), \quad (2.31)$$

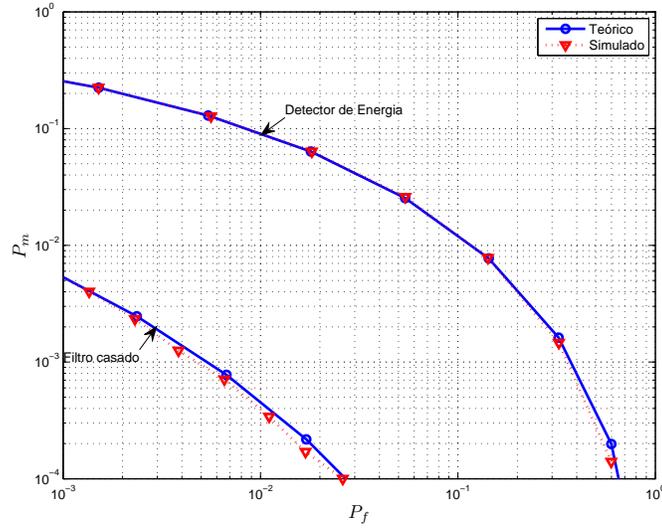


Figura 2.4: Comparação entre os detectores por energia e filtro casado. $M = 8$ e $\text{SNR}=3\text{dB}$

em que $Q(x)$ é a função erro complementar:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{t^2}{2}} dt. \quad (2.32)$$

Sob a hipótese nula, $T(x)$ apresenta distribuição gaussiana com média nula e variância $\sigma^2 = M\sigma_s^2\sigma_\eta^2$. Dessa forma, a probabilidade de falso alarme é dada por:

$$P_f = Pr \{T(x) > \lambda | \mathcal{H}_0\} = Q \left(\frac{\lambda}{\sigma_s\sigma_\eta\sqrt{M}} \right). \quad (2.33)$$

A Fig. 2.4 mostra que o desempenho do detector por filtro casado é superior ao detector por energia. A principal desvantagem dessa técnica de detecção é a necessidade do conhecimento *a priori* do sinal transmitido pelo usuário primário. Em muitas aplicações, essa hipótese não é factível, o que inviabiliza o método de detecção.

Note que no caso de canal com desvanecimento, ainda seria necessário conhecer o canal para que a detecção por filtro casado seja possível. Nesse caso, a estatística de teste passaria a ser dada por:

$$T(x) = \text{Re} \left\{ \sum_{k=0}^{M-1} x^*(k)h(k)s(k) \right\}. \quad (2.34)$$

2.3 Detector por ciclo-estacionaridade

Uma outra possibilidade de detecção é através do uso das propriedades de ciclo-estacionaridade [22, 23, 24, 25, 26] do sinal. Essa classe de sensoriamento espectral explora o fato de o sinal primário, em geral, apresentar característica ciclo-estacionária e o ruído ser estacionário no sentido amplo e decorrelacionado [21].

Um processo estocástico $x(t)$ é dito ciclo-estacionário se existe um período T_0 , tal que a sua função de autocorrelação $R(t_1, t_2) = E\{x(t_1)x(t_2)\}$, satisfaz à seguinte condição:

$$R(t_1, t_2) = R(t_1 + nT_0, t_2 + nT_0), \quad (2.35)$$

para qualquer inteiro n [27]. Em particular:

$$R(t, t + \tau) = R(t + nT_0, t + \tau + nT_0). \quad (2.36)$$

Logo, é possível expressar $R(t, t + \tau)$ por sua expansão em Série de Fourier:

$$R(t, t + \tau) = \sum_{k=-\infty}^{+\infty} \beta_k(\tau) e^{j2\pi \frac{k}{T_0} t}, \quad (2.37)$$

em que β_k são os coeficientes de Fourier dados por:

$$\beta_k(\tau) = \frac{1}{T_0} \int_{-\infty}^{+\infty} R(t, t + \tau) e^{-j2\pi \frac{k}{T_0} t} dt. \quad (2.38)$$

A expressão (2.38) é de pouca utilidade prática, já que é função da autocorrelação estatística do sinal $x(t)$. Como, usualmente, não se tem acesso a essa informação, é interessante explorar a própria medida instantânea $x(t)x(t + \tau)$. Apesar de não podermos afirmar que $x(t)x(t + \tau)$ é uma função periódica, em [28] é discutido que argumentos de ergodicidade em conjunto com (2.36) levam a expressar $x(t)x(t + \tau)$ como função *quase-periódica*. Nessas condições, $x(t)x(t + \tau)$ pode ser expressa como:

$$x(t)x(t + \tau) \approx \sum_{\alpha \in \mathcal{A}} R^\alpha(\tau) e^{j2\pi \alpha t} \quad (2.39)$$

Em que \mathcal{A} é um conjunto contável de frequências e os coeficientes de Fourier $R^\alpha(\tau)$,

são dados pela autocorrelação cíclica [28] definida como:

$$R^\alpha(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t + \tau/2) x^*(t - \tau/2) e^{-j2\pi\alpha t} dt. \quad (2.40)$$

Observe que para $\alpha = 0$, a autocorrelação cíclica coincide com a autocorrelação de um processo ergódico e estacionário em sentido amplo.

Para todo processo ciclo-estacionário, existe uma constante $\alpha \neq 0$, para a qual $R^\alpha(\tau)$ é não-nula. Se o processo não exibir tal característica, $R^\alpha(\tau) \neq 0$ apenas no caso em que $\alpha = 0$.

A densidade espectral de potência periódica pode ser utilizada como estatística de teste no problema de sensoriamento espectral para os casos em que o sinal primário exibe características cíclicas. A título de ilustração, vamos considerar o seguinte teste de hipóteses:

$$x(t) = \begin{cases} \mathcal{H}_0 : \eta(t) \\ \mathcal{H}_1 : \cos(2\pi f_0 t) + \eta(t) \end{cases}, \quad (2.41)$$

Observe que para o ruído $\eta(t)$, temos:

$$R^\alpha(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} \eta(t + \tau/2) \eta^*(t - \tau/2) e^{-j2\pi\alpha t} dt. \quad (2.42)$$

O caso em que $\alpha = 0$, corresponde à própria função de autocorrelação do sinal:

$$R^0(\tau) = \sigma_\eta^2 \delta(\tau). \quad (2.43)$$

Se $\alpha \neq 0$, a periodicidade de $e^{-j2\pi\alpha t}$ fará com que $R^\alpha(\tau) = 0, \forall \tau$.

Já para o termo de sinal, temos:

$$R^\alpha(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} \cos(2\pi f_0(t + \tau/2)) \cos(2\pi f_0(t - \tau/2)) e^{-j2\pi\alpha t} dt, \quad (2.44)$$

o qual pode ser expresso como:

$$R^\alpha(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T/2}^{T/2} (\cos(4\pi f_0 t) + \cos(2\pi f_0 \tau)) e^{-j2\pi\alpha t} dt. \quad (2.45)$$

Analisando (2.45), concluímos que:

$$R^\alpha(\tau) = \begin{cases} \frac{\cos(2\pi f_0 \tau)}{2}, & \alpha = 0 \\ \frac{1}{4}, & \alpha = \pm 2f_0 \end{cases}. \quad (2.46)$$

Portanto, nesse caso, pode-se utilizar como estatística de teste a função de autocorrelação para os valores $\alpha = \pm f_0$. Uma outra grandeza que avalia a ciclo-estacionaridade do processo é a densidade espectral de potência periódica, a qual é dada pela transformada de Fourier da autocorrelação cíclica:

$$S^\alpha(f) = \int_{-\infty}^{+\infty} R^\alpha(\tau) e^{-j2\pi f\tau} d\tau. \quad (2.47)$$

Em geral, é mais simples calcular a densidade espectral de potência periódica do que a função de autocorrelação cíclica. Em [29], é mostrada uma forma eficiente de se calcular a densidade espectral periódica denominada Método de Acumulação de FFT (*Fast Fourier Transform*) denominada FAM (*FFT accumulation method*). A estimativa da densidade espectral de potência periódica, obtida com o algoritmo apresentado em [29], apresenta distribuição circular gaussiana com parâmetros que são funções das hipóteses \mathcal{H}_0 e \mathcal{H}_1 [23].

Como mostrado em [23], o detector por ciclo-estacionaridade consegue prover melhores resultados que o detector de energia a um custo do aumento da complexidade computacional. Além disso, diferentemente do caso de detecção de energia, é necessário um conhecimento *a priori* do sinal primário para que a densidade espectral de potência cíclica seja estimada em frequências adequadas para a aplicação do teste.

Uma outra desvantagem desse método é que o canal precisa permanecer estático durante o sensoriamento, de modo que as características de ciclo-estacionaridade sejam preservadas, e o sinal possa ser detectado.

2.4 Detector por análise de autovalores

Um outro modo de promover o sensoriamento espectral é através da análise dos autovetores da matriz de covariância do sinal recebido [30, 31, 32, 33].

Nesse caso, consideremos um sistema em que o sinal do usuário primário é transmitido por um canal seletivo em frequência com coeficientes $h_0, h_1, \dots, h_{N_h-1}$. O

teste de hipóteses passa a ser dado por:

$$\begin{cases} \mathcal{H}_0 : x(n) = \eta(n) \\ \mathcal{H}_1 : x(n) = \sum_{k=0}^{N_h-1} h_k(n)s(n-k) + \eta(n). \end{cases} \quad (2.48)$$

Considere blocos compostos por M amostras consecutivas:

$$\mathbf{x}_M(n) = [x(n) \quad x(n-1) \quad \cdots \quad x(n-M+1)]^T. \quad (2.49)$$

A técnica de detecção consiste em aplicar um teste de hipóteses a alguma estatística baseada na matriz de autocorrelação do vetor de sinais recebidos $\mathbf{x}_M(n)$.

Sob a hipótese de ausência de sinal, as amostras recebidas são independentes e identicamente distribuídas (i.i.d.). Dessa forma, a matriz de covariância será dada por:

$$\mathbf{C}_0 = \sigma_\eta^2 \mathbf{J} \quad (2.50)$$

e essa matriz tem os autovalores $\lambda_1 = \lambda_2 = \cdots = \lambda_M = \sigma_\eta^2$.

Assumindo a hipótese \mathcal{H}_1 , a matriz de covariância será dada por:

$$\mathbf{C}_1 = \mathbf{H}\mathbf{C}_s\mathbf{H}^H + \mathbf{C}_0, \quad (2.51)$$

em que \mathbf{H} é a matriz de convolução do canal e $(\cdot)^H$ indica a operação de transposto conjugado.

Denotando os autovalores de $\mathbf{H}\mathbf{C}_s\mathbf{H}^H$ por ρ_i e os autovalores de \mathbf{C}_1 por λ_i , temos:

$$\lambda_i = \rho_i + \sigma_\eta^2. \quad (2.52)$$

Em tais métodos, os autovalores da estimativa da matriz de covariância do sinal recebido são obtidos e algumas estatísticas de teste podem ser geradas. A estatística mais imediata seria a relação entre o maior e o menor autovalor, ou seja:

$$T = \frac{|\lambda|_{max}}{|\lambda|_{min}}. \quad (2.53)$$

O desempenho teórico desse tipo de detector pode ser analisado através da teoria de matrizes aleatórias [34]. Para tanto, seria interessante determinar a função distribuição de probabilidade dos autovalores da matriz de autocorrelação.

Como não se tem acesso direto à matriz de autocorrelação, é necessário fazer uso de algum estimador. Denote por $\hat{\mathcal{R}}$ a estimativa da matriz de autocorrelação:

$$\hat{\mathcal{R}} = \frac{1}{L} \sum_{k=0}^{L-1} \mathbf{x}_M(n-k) \mathbf{x}_M^T(n-k), \quad (2.54)$$

em que L é um fator de suavização que indica o tamanho de janelamento da estimativa de autocorrelação.

Sob a hipótese de ausência de sinal primário, a matriz $\hat{\mathcal{R}}$ é conhecida como matriz Wishart. A distribuição conjunta dos autovalores de uma matriz de Wishart é dada pela distribuição de Fisher-Hsu-Roy [35], [36], [37]. Apesar de a distribuição conjunta dos autovalores ser conhecida, as distribuições marginais não as são. Então, os valores teóricos para as probabilidade de falso alarme são de difícil obtenção [30].

Uma alternativa para a análise teórica do desempenho de tais detectores consiste na análise assintótica, *i.e.*, $L \rightarrow \infty, M \rightarrow \infty$. Nessas condições, define-se a variável aleatória:

$$\Lambda_1 = \frac{\lambda_{max} - \left(\sqrt{L} + \sqrt{N}\right)^2}{\left(\sqrt{L} + \sqrt{M}\right) \left(\sqrt{\frac{1}{L}} + \sqrt{\frac{1}{M}}\right)^{\frac{1}{3}}}, \quad (2.55)$$

cuja distribuição é do tipo Tracy-Widom de ordem 2 [38].

A análise da probabilidade de perda de detecção é ainda mais complicada uma vez que a estimativa da matriz de autocorrelação $\hat{\mathcal{R}}$, sob a hipótese de sinal, não é mais uma matriz de Wishart e a distribuição da variável aleatória gerada a partir da ordenação de seus autovalores não é conhecida.

Apesar da dificuldade em se obter estimativas teóricas de tais técnicas, elas são interessantes por permitirem a obtenção de testes que não dependem de estimativas de potência de ruído para o cálculo dos limiares de decisão [30], apesar de apresentar uma complexidade computacional elevada quando comparadas com técnicas de detecção por energia. Entretanto, a principal desvantagem desse método consiste no fato de ele depender de hipótese de seletividade em frequência do canal entre os usuários primários e secundários. Essa hipótese pode ser obedecida no caso de canais seletivos em frequência, como foi mostrado ao longo da Seção, ou considerando o caso em que os rádios cognitivos são dotados de múltiplas antenas.

Uma outra vantagem do método vem do fato de que desde que a hipótese de correlação do sinal primário seja obedecida, ele pode ser aplicado a canais variantes no tempo.

2.5 Conclusões

Nesse Capítulo foram apresentadas algumas das possíveis técnicas de sensoriamento espectral. Mostrou-se que no caso em que não há qualquer conhecimento *a priori* do sinal transmitido pelo usuário primário, o detector deve ser do tipo de energia. Para esse tipo de detector, foram apresentadas as probabilidades de falso alarme e perda de detecção tanto para o caso de canal AWGN quanto para o caso de canal Rayleigh. Foi apresentado também o detector ótimo no caso em que há conhecimento *a priori* do sinal transmitido pelo usuário primário, o filtro casado, também foram mostradas as expressões de probabilidade de perda de detecção e falso alarme para esse detector. Posteriormente, foram mostradas duas alternativas de detecção para as situações em que, apesar de não se ter acesso ao sinal transmitido pelo usuário primário, se conhece alguma característica desse sinal. O detector por ciclo-estacionaridade é adequado para os casos em que o sinal apresenta alguma característica de periodicidade. Já o detector por análise de autovalores, é adequado para o caso de detecção de sinais correlacionados na presença de ruído AWGN.

Capítulo 3

Sensoriamento Espectral Cooperativo Centralizado

No processo de sensoriamento, a presença de desvanecimento temporal no canal dificulta consideravelmente a obtenção de estimativas confiáveis. Além disso, o efeito de sombreamento pode fazer com que o rádio cognitivo fique impossibilitado de detectar o sinal do usuário primário, ocasionando o problema conhecido na literatura por *hidden node* [39].

Difícilmente todos os rádios estarão sujeitos simultaneamente a esses problemas. Portanto, tais deficiências podem ser mitigadas se eles forem utilizados para o sensoriamento espectral, de modo que seja possível obter uma estimativa através da fusão das estimativas obtidas pelos diversos rádios cognitivos [4], [40], [41]. O fato de cada rádio cognitivo estar sujeito a desvanescimentos diferentes gera uma diversidade espacial que pode ser devidamente explorada com o uso de técnicas cooperativas de detecção. Neste e nos próximos dois Capítulos, serão analisados métodos de implementação de tais métodos.

Há dois modos clássicos de se implementar uma rede cooperativa de sensores. O primeiro modo consiste em definir um nó centralizador que compile as informações providas pelos vários sensores da rede, gerando então a decisão final. Essa é a chamada rede centralizadora. A segunda abordagem reside na rede distribuída, na qual cada usuário secundário processa as informações obtidas dos nós vizinhos, consistindo assim num centro de fusão individual. Neste Capítulo, os métodos de cooperação centralizada serão analisados e, no Capítulo seguinte, serão detalhados

os métodos de cooperação distribuída.

As redes de cooperação centralizadas também podem ser classificadas pela natureza da informação trocada entre os nós. Os nós podem trocar informação dura entre si, ou ainda, é possível que as próprias estatísticas sejam trocadas entre os nós, gerando as redes baseadas em decisões suaves.

Na Seção 3.1, será estudado o caso de troca de informação dura entre os nós e, na Seção 3.2, será analisado o caso de redes de fusão centralizada utilizando decisão suave.

3.1 Fusão centralizada utilizando decisão dura

As redes de cooperação centralizadas baseadas em decisão dura são as de implementação mais simples. O modelo desse sistema está apresentado na Fig. 3.1. Para implementar tal tipo de rede, cada nó pode enviar um bit b_i , indicando sua decisão ao nó decisor. Por exemplo, pode-se convencionar que $b_i = 1$, se o sensor i detectar a presença do usuário primário. Considerando que há K sensores na rede, pode-se estabelecer como regra de decisão o evento de n ou mais sensores indicarem a presença do usuário primário. Essa regra será denotada por *regra n em K* . Fazendo $n = 1$, a decisão equivale à aplicação de uma lógica do tipo OU aos bits provenientes dos diferentes sensores [42]. Pode-se, ainda, fazer $n = K$, dando origem à regra do tipo E.

Em [4] e [43], é mostrado que, na maioria dos casos de interesse, a regra do tipo OU fornece os melhores resultados. Entretanto, as análises efetuadas por [4] e [43] são restritas ao caso no qual os canais entre os rádios cognitivos e o centro de fusão não geram erros. Ao longo deste documento, será mostrado que a inserção de erros no envio das leituras dos rádios cognitivos ao centro de fusão pode levar a mudanças na escolha da regra de fusão ótima.

Na situação em que o canal entre o rádio cognitivo e o centro de fusão insere erros, as probabilidades de falso alarme e perda de detecção são dadas por [44]:

$$Q_f = \sum_{i=0}^{K-n} \binom{K}{i} [(1 - P_f)(1 - P_e) + P_f P_e]^i [P_f(1 - P_e) + (1 - P_f)P_e]^{K-i}, \quad (3.1)$$

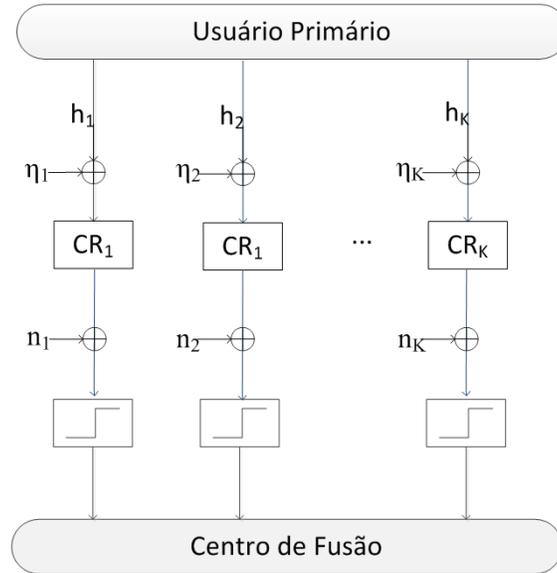


Figura 3.1: Modelo do sistema de fusão centralizada com decisão dura

$$Q_m = \sum_{i=K-n+1}^K \binom{K}{i} [P_m (1 - P_e) + (1 - P_m) P_e]^i [(1 - P_m) (1 - P_e) + P_m P_e]^{K-i}, \quad (3.2)$$

em que P_e é a probabilidade de haver erro de transmissão entre o rádio cognitivo e o centro de fusão, a qual depende do tipo de sinalização utilizada para a transmissão da decisão do rádio cognitivo para o centro de fusão e da distribuição estatística do ruído n_k entre o k -ésimo rádio cognitivo e o centro de fusão. P_f e P_m são as probabilidades de falso alarme e perda de detecção, respectivamente, relativas a cada rádio cognitivo individual e Q_f e Q_m são as probabilidades de falso alarme e perda de detecção do sistema cooperativo.

Há várias formas de se avaliar o desempenho do detector. Quando não se tem conhecimento das probabilidades *a priori* das hipóteses nula e de sinal, pode-se utilizar o critério NP (Neyman Pearson) apresentado no Capítulo 2. Já em situações em que são conhecidas as probabilidades *a priori* das hipóteses nula e de sinal, uma possibilidade seria utilizar uma abordagem Bayesiana, como será explicado em maiores na Subseção seguinte.

Na próxima Subseção, através do detalhamento de desempenho do sistema tendo em vista o critério de otimização de Bayes, serão mostrados limitantes de desempe-

no que indicarão quais seriam as melhores regras a serem utilizadas em diferentes cenários de probabilidade de erro e probabilidades *a priori* das hipóteses nula e de sinal.

3.1.1 Risco de Bayes

O desempenho do detector pode ser avaliado a partir da métrica dada pelo Risco de Bayes [14], [17]:

$$\mathcal{R}(n) = Q_f(n) Pr\{\mathcal{H}_0\} C_f + Q_m(n) Pr\{\mathcal{H}_1\} C_m, \quad (3.3)$$

em que $Pr\{\mathcal{H}_0\}$ é a probabilidade *a priori* associada à hipótese nula, *i.e.*, a probabilidade de que o usuário primário não esteja transmitindo, $Pr\{\mathcal{H}_1\}$ é a probabilidade *a priori* associada à hipótese de sinal e C_f e C_m são constantes positivas que indicam os custos associados aos eventos de falso alarme e perda de detecção, respectivamente.

As probabilidades *a priori* são definidas a partir do modelo de ocupação do canal. Em [45], a ocupação do canal é modelada como um processo de renovação alternando entre os estados livre e ocupado. É importante enfatizar que o tratamento analítico aqui desenvolvido é facilmente generalizado para qualquer modelo de ocupação de canal, uma vez que esse modelo afetará somente as probabilidades *a priori* envolvidas na definição do Risco de Bayes.

Cada um desses estados é modelado como uma variável aleatória de Poisson de parâmetros α e β . Portanto, os seus valores médios são dados por $\frac{1}{\alpha}$ e $\frac{1}{\beta}$, respectivamente. Logo, as probabilidades *a priori* serão dadas por:

$$Pr\{\mathcal{H}_0\} = \frac{\frac{1}{\beta}}{\frac{1}{\beta} + \frac{1}{\alpha}} = \frac{\alpha}{\alpha + \beta}, \quad (3.4)$$

$$Pr\{\mathcal{H}_1\} = \frac{\frac{1}{\alpha}}{\frac{1}{\beta} + \frac{1}{\alpha}} = \frac{\beta}{\alpha + \beta}. \quad (3.5)$$

Dessa forma, o Risco de Bayes passa a ser escrito como:

$$\mathcal{R}(n) = Q_f(n) \frac{\alpha}{\alpha + \beta} C_f + Q_m(n) \frac{\beta}{\alpha + \beta} C_m. \quad (3.6)$$

Antes de proceder com a minimização do Risco de Bayes, é interessante investigar o impacto da probabilidade de erro P_e . Analisando (3.1) e (3.2), pode-se notar que se a condição $P_f \ll P_e$ for obedecida, $Q_f(n)$ pode ser aproximada por:

$$Q_f(n) \approx \overline{Q}_f(n) = \sum_{i=0}^{K-n} \binom{K}{i} (1 - P_e)^i P_e^{K-i}. \quad (3.7)$$

De modo semelhante, se a condição $P_m \ll P_e$ for obedecida, a seguinte aproximação sobre $Q_m(n)$ pode ser feita:

$$Q_m(n) \approx \overline{Q}_m(n) = \sum_{i=K-n+1}^K \binom{K}{i} P_e^i (1 - P_e)^{K-i}. \quad (3.8)$$

Caso essas duas condições sejam simultaneamente satisfeitas, $\mathcal{R}(n)$ pode ser aproximado como:

$$\overline{\mathcal{R}}(n) = \overline{Q}_f(n) \frac{\alpha}{\alpha + \beta} C_f + \overline{Q}_m(n) \frac{\beta}{\alpha + \beta} C_m. \quad (3.9)$$

Sabe-se que devido ao compromisso entre P_f e P_m , as duas condições $P_f \ll P_e$ e $P_m \ll P_e$ podem apenas ser satisfeitas em regime de SNR extremamente elevada e com o uso de um método de detecção muito eficiente. Portanto, neste contexto, (3.9) é um limitante inferior conservador para $\mathcal{R}(n)$. Neste ponto é importante enfatizar que P_e indica a probabilidade de erro introduzida no canal entre o rádio cognitivo e o centro de fusão e a SNR a qual nos referimos diz respeito ao canal entre os usuários primário e secundários, portanto, por estarmos tratando de canais diferentes, é possível que P_e seja elevada mesmo em condições de alta SNR. O limitante inferior mostrado em (3.9) indica que mesmo em cenários extremamente favoráveis de detecção, os erros introduzidos pelo canal entre o centro de fusão e o rádio cognitivo podem degradar consideravelmente o desempenho do sistema.

É interessante notar que (3.9) pode ser um limitante não-conservador mesmo que uma das condições não seja obedecida. Uma situação interessante ocorre para $n = 1$, *i.e.*, a regra OU. Nessa condição se a condição $P_f \ll P_e$ é satisfeita, tem-se:

$$Q_f(1) \approx \overline{Q}_f(1) = 1 - (1 - P_e)^K, \quad (3.10)$$

e aplicando (3.2) para $n = 1$, obtém-se:

$$Q_m(1) = (P_m(1 - P_e) + P_e(1 - P_m))^K. \quad (3.11)$$

Nessa situação, ou seja, se $n = 1$ e $P_f \ll P_e$, pode-se relaxar a condição em P_m , desde que a seguinte desigualdade seja válida:

$$\mathcal{C}(P_m) = \frac{\alpha C_f (1 - (1 - P_e)^K)}{\beta C_m (P_m(1 - P_e) + P_e(1 - P_m))^K} \gg 1, \quad (3.12)$$

Nesse caso, $\mathcal{R}(1) \approx \overline{\mathcal{R}}(1)$, como será confirmado com resultados de simulação apresentados mais adiante.

Com o objetivo de comparar o Risco de Bayes para duas regras diferentes, n_1 e n_2 , definiremos o seguinte campo escalar:

$$\Delta_{n_1, n_2} = \mathcal{R}(n_1) - \mathcal{R}(n_2), \quad (3.13)$$

Observando que (3.1) e (3.2) podem ser reescritas como:

$$Q_f = \sum_{i=0}^{K-n} \binom{K}{i} [1 - P_f^e]^i [P_f^e]^{K-i}, \quad (3.14)$$

$$Q_m = \sum_{i=K-n+1}^K \binom{K}{i} [P_m^e]^i [1 - P_m^e]^{K-i}, \quad (3.15)$$

em que os novos parâmetros P_f^e e P_m^e são dados por:

$$P_f^e = P_f(1 - 2P_e) + P_e, \quad (3.16)$$

$$P_m^e = P_m(1 - 2P_e) + P_e. \quad (3.17)$$

É importante notar que para uma probabilidade de erro P_e fixada, os novos parâmetros são definidos nos seguintes intervalos:

$$P_e \leq P_f^e, P_m^e \leq 1 - P_e. \quad (3.18)$$

Portanto, pode-se analisar a diferença entre os Riscos de Bayes para duas dife-

rentes regras como um campo escalar bivariado:

$$\Delta_{n_1, n_2} (P_f^e, P_m^e) = \mathcal{R}_{n_1} (P_f^e, P_m^e) - \mathcal{R}_{n_2} (P_f^e, P_m^e). \quad (3.19)$$

Para os casos em que $\Delta_{n_1, n_2} (P_f^e, P_m^e) < 0$, a regra $n = n_1$ é mais favorável do que a regra $n = n_2$. Similarmente, se $\Delta_{n_1, n_2} (P_f^e, P_m^e) > 0$, a regra $n = n_2$ é mais vantajosa.

Duas das mais utilizadas regras de fusão são as regras OU e E, *i.e.*, as regras para $n = 1$ e $n = K$. Considerando essas duas regras de fusão, o campo escalar $\Delta_{n_1, n_2} (P_f^e, P_m^e) > 0$ pode ser reescrito como:

$$\Delta_{1, K} (P_f^e, P_m^e) = \frac{1}{\alpha + \beta} \left[\alpha C_f \left(1 - (1 - P_f^e)^K - (P_f^e)^K \right) + \beta C_m \left(-1 + (1 - P_m^e)^K + (P_m^e)^K \right) \right]. \quad (3.20)$$

O gradiente do campo escalar $\Delta_{1, K}$ é dado por:

$$\nabla (\Delta_{1, K}) = \frac{K}{\alpha + \beta} \left[\alpha C_f \left[-(P_f^e)^K + (1 - P_f^e)^K \right] \right. \\ \left. \beta C_m \left[(P_m^e)^K - (1 - P_m^e)^K \right] \right]. \quad (3.21)$$

Portanto, o único ponto crítico de $\Delta_{1, K}$, *i.e.*, o ponto no qual o gradiente é nulo, é dado por $(P_{f, s}^e, P_{m, s}^e) = (\frac{1}{2}, \frac{1}{2})$. A matriz Hessiana relacionada a esse ponto crítico é dada por:

$$\mathcal{H} = \begin{bmatrix} -2 \frac{\alpha}{\alpha + \beta} C_f K (K - 1) \left(\frac{1}{2}\right)^{K-2} & 0 \\ 0 & 2 \frac{\beta}{\alpha + \beta} C_m K (K - 1) \left(\frac{1}{2}\right)^{K-2} \end{bmatrix}. \quad (3.22)$$

Como o determinante da matriz Hessiana é negativo, independentemente do valor de K , concluímos que esse ponto crítico corresponde a um ponto de sela. Os pontos extremos serão, portanto, situados nas bordas do domínio do campo escalar.

Fixando $P_f^e = P_{f, s}^e = \frac{1}{2}$ e $P_m^e = P_e$, temos que:

$$\Delta_{1, K}^{\max} = \frac{\alpha C_f}{\alpha + \beta} \left(1 - \frac{1}{2^{K-1}} \right) + \frac{\beta C_m}{\alpha + \beta} \left(-1 + P_e^K + (1 - P_e)^K \right). \quad (3.23)$$

Já fixando $P_m^e = P_{m, s}^e = \frac{1}{2}$ e $P_f^e = P_e$, temos que:

$$\Delta_{1, K}^{\min} = \frac{\alpha C_f}{\alpha + \beta} \left(1 - P_e^K - (1 - P_e)^K \right) + \frac{\beta C_m}{\alpha + \beta} \left(-1 + \frac{1}{2^{K-1}} \right). \quad (3.24)$$

Considerando que $1 - P_e^K - (1 - P_e)^K \approx 1 - (1 - P_e)^K$, temos que o campo escalar será não-negativo se e somente se a seguinte condição for satisfeita:

$$\alpha C_f \left(1 - (1 - P_e)^K\right) - \beta C_m \left(1 - \frac{1}{2^{K-1}}\right) \geq 0, \quad (3.25)$$

ou equivalentemente:

$$a \left(1 - \frac{1}{2^{K-1}}\right) < 1 - (1 - P_e)^K, \quad (3.26)$$

em que $a = \frac{\beta C_m}{\alpha C_f}$.

Como $1 - (1 - P_e)^K \leq 1$, temos que uma condição necessária para que (3.26) seja satisfeita é que:

$$a \leq \frac{1}{1 - 2^{-(K-1)}}. \quad (3.27)$$

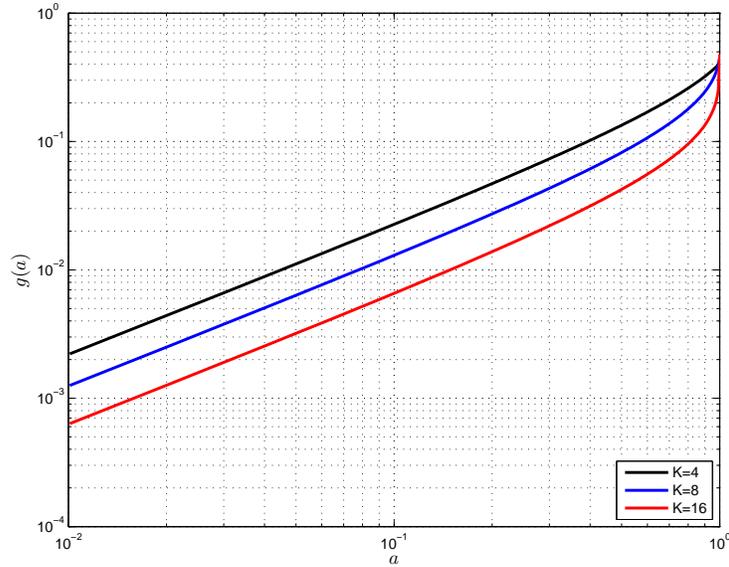
A segunda condição necessária é que:

$$P_e > g(a) = 1 - [1 - a(1 - 2^{-(K-1)})]. \quad (3.28)$$

Portanto, pode-se concluir que, nessas condições, a regra E apresenta melhor desempenho do que a regra do tipo OU. Note ainda que se o sistema contar com muitos usuários secundários, *i.e.*, se $K \gg 1$, (3.27) pode ser simplificada para $\beta C_m < \alpha C_f$. Esse fato pode ser interpretado da seguinte forma: uma condição necessária para que o desempenho da regra E seja superior ao da regra OU é que o Q_f impacte mais no Risco de Bayes do que Q_m . A segunda condição necessária indicada por (3.28) mostra que a probabilidade de erro P_e precisa ser superior que um dado limiar expresso em (3.28). Portanto, caso o canal entre o centro de fusão e o rádio cognitivo seja livre de erros, não é possível afirmar que a regra E supera a regra OU independentemente dos valores de P_f e P_m .

De modo análogo, pode-se mostrar que o campo escalar $\Delta_{1,K}$ será não-positivo se as condições (3.27) e (3.28) forem satisfeitas e se a constante a for definida como $a = \frac{\alpha C_f}{\beta C_m}$. Nesse caso, pode-se afirmar que uma condição necessária para que a regra OU supere a regra E é que a influência de Q_m seja maior do que a influência de Q_f no Risco de Bayes.

A função $g(a)$ expressa em (3.28) é mostrada na Fig. 3.2 para diferentes valores


 Figura 3.2: Função $g(a)$

de K . É possível observar que se o impacto de Q_f e Q_m for aproximadamente o mesmo, *i.e.*, se $a \approx 1$, a desigualdade (3.28) será satisfeita apenas se P_e for proibitivamente elevada ($P_e \approx \frac{1}{2}$). Ainda, pode-se notar que para valores crescentes de K , o limitante inferior para P_e decresce.

É ainda interessante notar que apesar de $(P_{f,s}^e, P_{m,s}^e)$ ser ponto de sela do campo escalar relativo às variáveis transformadas P_f^e e P_m^e , também é ponto de sela do campo relativo às variáveis originais P_f e P_m . Esse fato pode ser constatado observando que aplicando (3.16) em (3.17), tem-se que o ponto crítico $(\frac{1}{2}, \frac{1}{2})$ é mapeado em:

$$P_f = P_m = \frac{\frac{1}{2} - P_e}{1 - 2P_e} = \frac{1}{2}, \quad (3.29)$$

independentemente do valor de P_e e, portanto, $(P_{f,s}^e, P_{m,s}^e) = (\frac{1}{2}, \frac{1}{2})$ é também um ponto de sela do campo escalar $\Delta_{1,n}(P_f, P_m)$ para todos os possíveis valores de P_e .

Esse fato é ilustrado na Fig. 3.3, em que se considera um sistema com $K = 8$ rádios cognitivos e parâmetros $\alpha = 0,2$, $\beta = 0,8$, $C_f = 0,3$, $C_m = 0,7$ e $P_e = 0, 10^{-2}, 10^{-1}$. Note que independentemente do valor de P_e , o ponto de sela é fixado em $(\frac{1}{2}, \frac{1}{2})$, indicado na figura. É também interessante notar que para esses parâmetros, temos $a = \frac{\alpha C_f}{\beta C_m} = 0,1071$, o que satisfaz (3.27). Esse valor de a é mapeado em $g(a) = 0,0140$. Portanto, se $P_e > 0,0140$, o risco de Bayes associado à

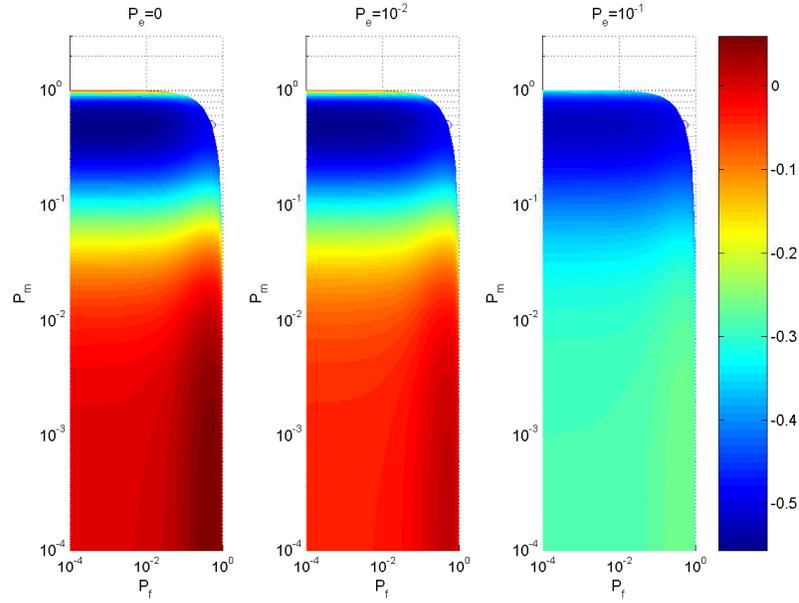


Figura 3.3: Campo escalar $\Delta_{1,K}$ considerando $K = 8$, $\alpha = 0,2$, $\beta = 0,8$, $C_f = 0,3$ e $C_m = 0,7$ para diferentes valores de P_e .

regra OU será inferior ao da regra E quaisquer que sejam os valores de P_f e P_m . De fato, para o caso $P_e = 10^{-1}$, o campo escalar é não-positivo conforme esperado.

Essa comparação se restringe ao caso das regras de fusão E e OU. Para estender a análise a outras regras, determinaremos o valor n que minimiza o Risco de Bayes em (3.3). É importante enfatizar que diferentemente de [46], em que se considerou que o canal entre os rádios cognitivos e o centro de fusão era ideal, aqui considera-se uma probabilidade de erro de transmissão P_e associada a esse canal.

Para determinar o valor ótimo de n , parte-se do pressuposto que $\mathcal{R}(n)$ apresenta um único mínimo. Nessa situação, concluímos que o valor ótimo de n é dado pelo menor inteiro que satisfaz a seguinte desigualdade:

$$\mathcal{R}(n+1) - \mathcal{R}(n) \geq 0. \quad (3.30)$$

Algumas manipulações algébricas mostradas no Apêndice A, verificamos que n deve satisfazer:

$$n = \left\lceil \frac{K - \phi}{1 + \theta} \right\rceil, \quad (3.31)$$

em que

$$\theta = \frac{\log \left(\frac{P_f(1-P_e) + (1-P_f)P_e}{(1-P_m)(1-P_e) + P_e P_m} \right)}{\log \left(\frac{P_m(1-P_e) + (1-P_m)P_e}{(1-P_f)(1-P_e) + P_f P_e} \right)}, \quad (3.32)$$

$$\phi = \frac{\log \left(\frac{\alpha C_f}{\beta C_m} \right)}{\log \left(\frac{P_m(1-P_e) + (1-P_m)P_e}{(1-P_f)(1-P_e) + P_f P_e} \right)}. \quad (3.33)$$

e $\lceil \cdot \rceil$ é a função teto.

Lembrando que n_{opt} deve permanecer no intervalo $1 \leq n_{opt} \leq K$, temos que:

$$n_{opt} = \begin{cases} 1, & \lceil \frac{K-\phi}{1+\theta} \rceil < 1 \\ \lceil \frac{K-\phi}{1+\theta} \rceil, & 1 \leq \frac{K-\phi}{1+\theta} \leq K \\ K, & \frac{K-\phi}{1+\theta} > K. \end{cases} \quad (3.34)$$

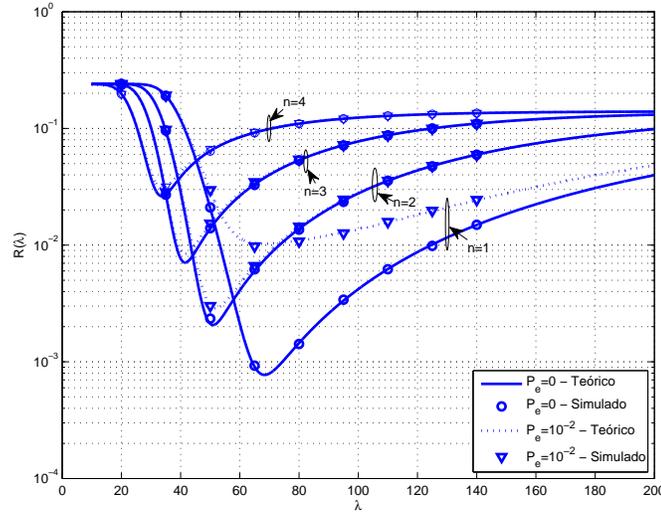
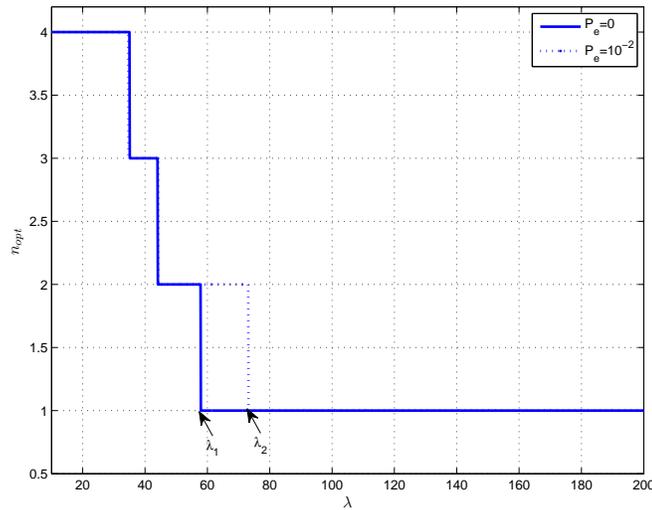
Pode-se notar que n_{opt} , dado por (3.34), é uma função dos parâmetros θ e ϕ , os quais dependem de P_f e P_m . Todavia, essas probabilidades são parametrizadas pelo limiar de detecção λ , definido no Capítulo 2. Portanto, para diferentes valores de λ , nós temos diferentes valores para o parâmetro n_{opt} .

3.1.2 Resultados de simulação

Com o objetivo de corroborar os resultados teóricos, considera-se um sistema com parâmetros $\alpha = 0,8$, $\beta = 0,2$, $C_f = 0,3$ e $C_m = 0,7$. Nesse contexto, em média, o período em que o espectro está livre é maior do que o período de ocupação do espectro e o risco de Bayes é mais afetado pelos eventos de perda de detecção do que os eventos de falso alarme.

A Fig. 3.4 mostra o risco de Bayes em função do parâmetro λ em um sistema com $K = 4$ rádios cognitivos. O canal entre os rádios cognitivos e o usuário primário é Rayleigh com SNR média de 3dB. Considera-se um detector de energia empregando $M = 16$ amostras.

O sistema é analisado considerando os casos de canal entre o rádio cognitivo e usuário primário livre de erros, *i.e.*, $P_e = 0$ e o caso em que o canal insere erros com probabilidade $P_e = 10^{-2}$. A partir da Fig. 3.4, nota-se que o valor de n correspondente ao risco de Bayes mínimo é função de λ , assim como previsto na


 Figura 3.4: Risco de Bayes para diferentes valores de n e $K = 4$.

 Figura 3.5: Parâmetro n ótimo dado por (3.34).

análise teórica. É importante observar que o impacto do erro no canal entre o rádio cognitivo e o centro de fusão é maior para o caso $n = 1$ do que para valores mais elevados de n .

Com o objetivo de explicar esse comportamento observado para o caso $n = 1$, é importante indicar que os valores de P_f e P_m no ponto de mínimo do risco de Bayes são $P_f = 2,0635 \cdot 10^{-4}$ e $P_m = 0,2532$. Nota-se que apenas P_f satisfaz à condição $P_f \ll P_e$, entretanto, P_m é tal que $\mathcal{C}(P_m) = 15,2126 \gg 1$. Portanto, a cota inferior apresentada em (3.9) deve ser uma boa aproximação para \mathcal{R} . De fato,

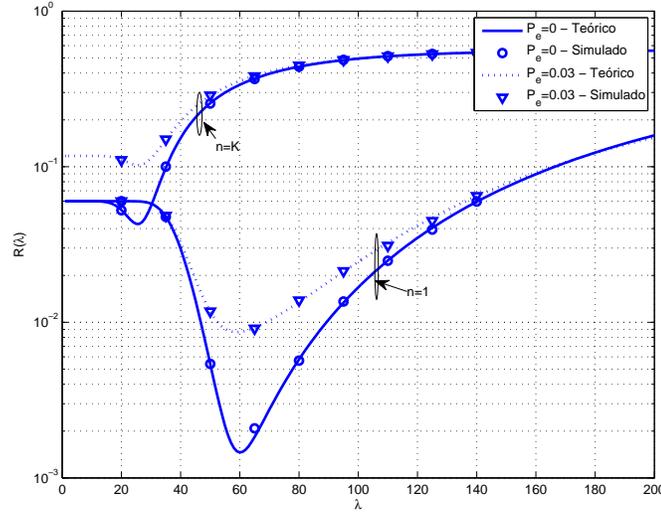
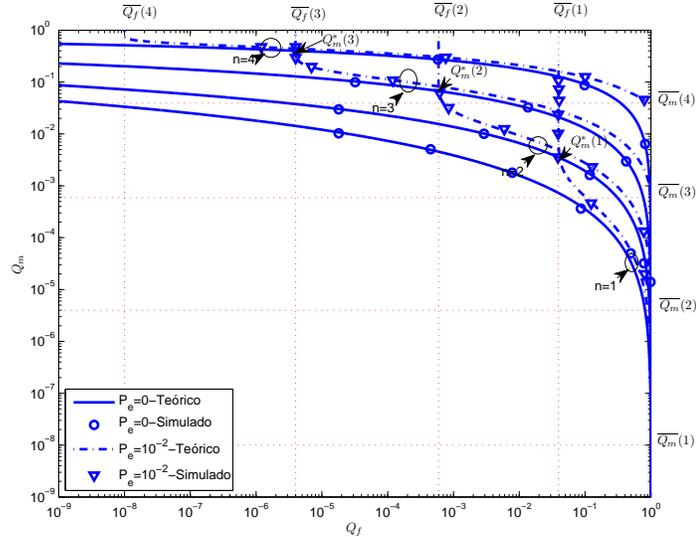


Figura 3.6: Comparação entre as regras OU e E com parâmetros $\alpha = 0,2$, $\beta = 0,8$, $C_f = 0,3$ e $C_m = 0,7$

tem-se $\mathcal{R}(1) = 0,0103$ e $\overline{\mathcal{R}}(1) = 0,0095$ para $P_e = 10^{-2}$. Observa-se também que, no caso em que $P_e = 0$, o valor mínimo do risco de Bayes é dado por $\mathcal{R}(1) = 7,72 \cdot 10^{-4}$. Esse valor é muito inferior àquele obtido com $P_e = 10^{-2}$, $\overline{\mathcal{R}}(1) = 0,0095$. Para os outros valores de n , esse fato não ocorre e, por tal motivo, a degradação do sistema é bem mais sutil. Como consequência, apesar do valor mínimo de \mathcal{R} ser atingido para $n = 1$ no caso em que $P_e = 0$, para o caso em que $P_e = 10^{-2}$, $n = 2$ é o parâmetro que leva ao mínimo global do risco de Bayes.

Na Fig. 3.5, está mostrado o valor ótimo de n obtido aplicando (3.34). A expressão analítica para n_{opt} pode ser validada comparando as Figs. 3.4 e 3.5. Por exemplo, para o caso em que o canal entre o rádio cognitivo e o centro de fusão é livre de erro, se a condição $\lambda > \lambda_1 = 57,9$ é satisfeita, a regra ótima é dada por $n = 1$. De forma semelhante, quando $P_e = 10^{-2}$, $n = 1$ é a regra de fusão ótima para os casos em que $\lambda > \lambda_2 = 73,2$, o que também corresponde ao resultado mostrado na Fig. 3.5.

Pode-se também verificar os resultados teóricos comparando as regras E e OU. Considerando os parâmetros $\alpha = 0,2$, $\beta = 0,8$, $C_f = 0,3$ e $C_m = 0,7$, tem-se $a = \frac{\alpha C_f}{\beta C_m} = 0,1071$. Aplicando (3.28), pode-se afirmar que se $P_e \geq 0,0243$, a regra OU apresenta desempenho superior à regra E para qualquer que seja o limiar de


 Figura 3.7: ROC para diferentes valores de n e $K = 4$.

decisão λ . A Fig. 3.6 mostra que, no caso em que $P_e = 0$, a regra E supera a regra OU para $13 < \lambda < 30$. Entretanto, no caso em que $P_e = 0,03$, a regra OU supera a regra E qualquer que seja o parâmetro λ .

Adicionalmente, é possível analisar o desempenho do detector sob a ótica do critério NP. A Fig. 3.7 mostra a ROC para diferentes valores de n . A partir desses resultados, pode-se notar que no caso em que $P_e = 0$, o valor ótimo de n é $n = 1$ para qualquer que seja a probabilidade de perda de detecção Q_m fixada, o que não ocorre quando $P_e = 10^{-2}$. Denotando essa probabilidade por Q_m^t , tem-se que:

$$n_{opt} = \begin{cases} 1, & Q_m^t \leq Q_m^*(1) \\ n + 1, & Q_m^*(n) < Q_m^t \leq Q_m^*(n + 1) \\ K, & Q_m^t > Q_m^*(K - 1), \end{cases} \quad (3.35)$$

em que $Q_m^*(n)$ corresponde ao valor mínimo de probabilidade de perda de detecção que leva a $\overline{Q_f}(n)$, como mostrado na Fig. 3.7.

Uma outra análise interessante referente a esse problema pode ser feita através da Fig. 3.8, em que o mapa de cores da ROC é influenciado pelo risco de Bayes. Mais uma vez, nota-se que o sistema com $n = 2$ é mais robusto a erros do canal entre o rádio cognitivo e o centro de fusão. Pela escala de cores, percebe-se que o sistema com $n = 1$ é gravemente degradado no caso em que $P_e = 10^{-2}$.

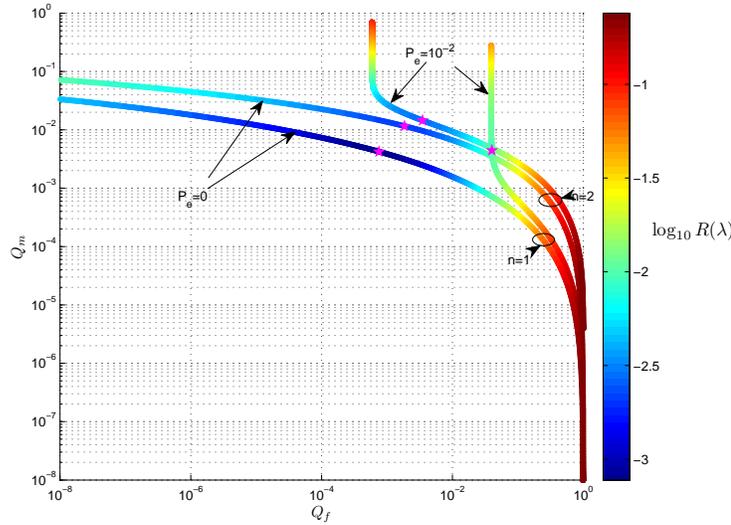


Figura 3.8: Impacto do Risco de Bayes considerando a adoção do critério NP. Os pontos destacados por \star indicam os pontos em que o risco de Bayes é mínimo.

3.2 Fusão centralizada usando decisão suave

Alternativamente ao caso da decisão dura tomada nos nós, é possível que estes enviem as próprias estatísticas de decisão ao centro de fusão [47] e [48], permitindo que seja feita a decisão suave. Nessas condições, o teste ótimo é dado por:

$$L(\mathbf{y}) = \frac{P(\mathbf{y}|\mathcal{H}_1)}{P(\mathbf{y}|\mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda, \quad (3.36)$$

em que $\mathbf{y} = [T(x_1) \cdots T(x_K)]$, λ é o limiar de decisão determinado pela probabilidade de detectar as regiões disponíveis do espectro, $P(\mathbf{y}|\mathcal{H}_1)$ é a função densidade de probabilidade sob a hipótese de sinal, $P(\mathbf{y}|\mathcal{H}_0)$ é a função densidade de probabilidade sob a hipótese nula e $T(x_k)$ é a estimativa de energia dada por:

$$T(x_k) = \sum_{n=0}^{M-1} x_k^2(n). \quad (3.37)$$

Como será visto em detalhes na Subseção 3.2.2, as funções densidades de probabilidade $P(\mathbf{y}|\mathcal{H}_1)$ e $P(\mathbf{y}|\mathcal{H}_0)$ levam a expressões complicadas e de difícil manipulação matemática do teste ótimo dado em (3.36). Por tal motivo, é comum aproximar essas densidades de probabilidade por gaussianas, assim como feito em [49] e [48]. Nesses trabalhos, é aplicado o Teorema Central do Limite, considerando-se que o

vetor \mathbf{y} é assintoticamente gaussiano, *i.e.*, considera-se M em (3.37) um número suficientemente grande, de forma que cada medida energia em (3.37) tende a uma variável aleatória gaussiana. Dessa forma, as densidades de probabilidade $P(\mathbf{y}|\mathcal{H}_1)$ e $P(\mathbf{y}|\mathcal{H}_0)$ são completamente especificadas pelos vetores de média e matrizes de covariância.

Considerando que a variável aleatória x_k é definida como:

$$x_k(n) = \begin{cases} \eta_k(n), & \mathcal{H}_0 \\ s(n) + \eta_k(n), & \mathcal{H}_1 \end{cases}, \quad (3.38)$$

em que $s(n)$ é o sinal de potência p_s transmitido pelo usuário primário e $\eta_k(n) \sim \mathcal{N}(0, \sigma_{\eta,k}^2)$, tem-se que sob a hipótese \mathcal{H}_0 , o vetor \mathbf{y} tem média:

$$\boldsymbol{\mu}_0 = K [\sigma_{\eta_1}^2 \dots \sigma_{\eta_K}^2]^T, \quad (3.39)$$

e matriz de covariância:

$$\boldsymbol{\Sigma}_0 = \text{diag}([\sigma_{\eta_1}^4 \dots \sigma_{\eta_K}^4]). \quad (3.40)$$

Já para a hipótese \mathcal{H}_1 :

$$\boldsymbol{\mu}_1 = K [\sigma_{\eta_1}^2 + p_s \dots \sigma_{\eta_K}^2 + p_s]^T \quad (3.41)$$

e matriz de covariância:

$$\boldsymbol{\Sigma}_1 = \text{diag}([\sigma_{\eta_1}^2 (\sigma_{\eta_1}^2 + p_s) \dots \sigma_{\eta_K}^2 (\sigma_{\eta_K}^2 + p_s)]). \quad (3.42)$$

Portanto, temos:

$$\frac{P(\mathbf{y}|\mathcal{H}_1)}{P(\mathbf{y}|\mathcal{H}_0)} = \frac{\sqrt{\det(\boldsymbol{\Sigma}_0)} \exp\left[-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1}(\mathbf{y} - \boldsymbol{\mu}_1)\right]}{\sqrt{\det(\boldsymbol{\Sigma}_1)} \exp\left[-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1}(\mathbf{y} - \boldsymbol{\mu}_0)\right]}. \quad (3.43)$$

Finalmente, tomando o logarítmo, obtém-se:

$$L_q(\mathbf{y}) = \mathbf{y}^T (\boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\Sigma}_1^{-1}) \mathbf{y} + 2(\boldsymbol{\mu}_1^T \boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\mu}_0^T \boldsymbol{\Sigma}_0^{-1}) \mathbf{y}. \quad (3.44)$$

Note que na expressão acima não foram incluídos os termos constantes, aqueles que não dependem do sinal, já que esses não alteram o detector ótimo. A forma não-

linear da função de verossimilhança traz dificuldades para a otimização do limiar λ . Portanto, é comum propor um teste linear sub-ótimo, na forma de:

$$L_l(\mathbf{y}) = \mathbf{w}^T \mathbf{y}, \quad (3.45)$$

em que \mathbf{w} representa o vetor de pesos.

A distribuição de $L_l(\mathbf{y})$ é gaussiana, supondo a gaussianidade de \mathbf{y} . Dessa forma, a probabilidade de falso alarme é dada por:

$$P_f = Q\left(\frac{\lambda - \boldsymbol{\mu}_0^T \mathbf{w}}{\sqrt{\mathbf{w}^T \boldsymbol{\Sigma}_0 \mathbf{w}}}\right) \quad (3.46)$$

e a probabilidade de perda de detecção é obtida como:

$$P_m = 1 - Q\left(\frac{\lambda - \boldsymbol{\mu}_1^T \mathbf{w}}{\sqrt{\mathbf{w}^T \boldsymbol{\Sigma}_1 \mathbf{w}}}\right). \quad (3.47)$$

3.2.1 Validade da aproximação gaussiana

A literatura envolvendo a combinação suave de decisões no contexto de sensoriamento espectral em rádio cognitivo é fortemente fundamentada na validade da aproximação gaussiana. Entretanto, é importante notar que a função densidade de probabilidade $f_Y(y)$, da variável aleatória composta pela soma de N variáveis aleatórias independentes, *i. e.*, $Y = \sum_{k=0}^{N-1} X_k$, é aproximadamente gaussiana se as duas condições abaixo são simultaneamente satisfeitas:

1. N deve ser suficientemente grande.
2. y deve ser um valor suficiente próximo da média da variável aleatória Y .

A segunda condição indica que a aproximação gaussiana, garantida pelo Teorema Central do Limite, não é uma boa aproximação para a cauda da distribuição.

Portanto, para que (3.46) e (3.47) sejam válidas simultaneamente, é necessário que o limiar de decisão esteja próximo do centro das distribuições das estatísticas de teste sob as duas hipóteses de detecção.

A fim de ilustrar esse fato, considere um sistema com um rádio cognitivo com $M = 32$ amostras na estimação de energia e em dois contextos de SNR distintos: $\gamma = -10$ dB e $\gamma = 0$ dB.

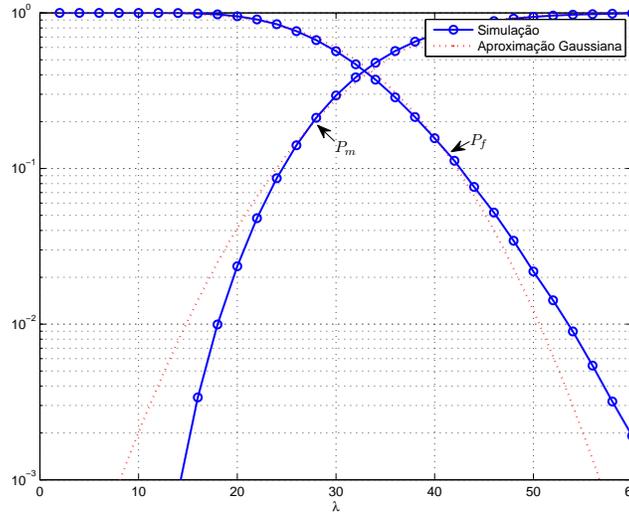


Figura 3.9: P_f e P_m em função do limiar de decisão λ . Considerando $\gamma = -10$ dB e $M = 32$.

As Figs. 3.9 e 3.10 mostram P_f e P_m em função do parâmetro λ . Observa-se que, no caso em que a SNR é muito baixa, $\gamma = -10$ dB, se $\lambda \in [28, 40]$, a aproximação gaussiana é válida tanto para P_f quanto para P_m . Entretanto, observa-se que para $\gamma = 0$ dB, nos casos em que a aproximação gaussiana é válida para P_f , ela não o é para P_m e vice-versa.

A Fig. 3.11 mostra as ROCs correspondentes às SNRs $\gamma = 0$ dB e $\gamma = -10$ dB. Esses resultados comprovam que a validade da aproximação gaussiana está associada também ao valor de SNR. Mesmo para um número grande de amostras, *i.e.*, $M = 32$, a aproximação se mantém válida apenas para valores proibitivamente baixos de SNR.

Conclui-se, portanto, que otimizar o detector a partir das expressões obtidas, explorando uma possível gaussianidade das estatísticas de teste, pode não ser a estratégia ideal.

Com o objetivo de contornar esse problema, na próxima Subseção é apresentada uma alternativa para a otimização do detector, partindo da densidade de probabilidade exata da estatística de teste.

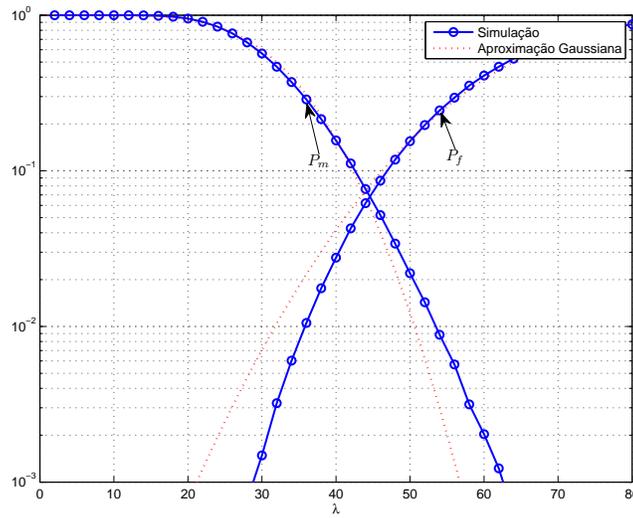


Figura 3.10: P_f e P_m em função do limiar de decisão λ . Considerando $\gamma = 0$ dB e $M = 32$.

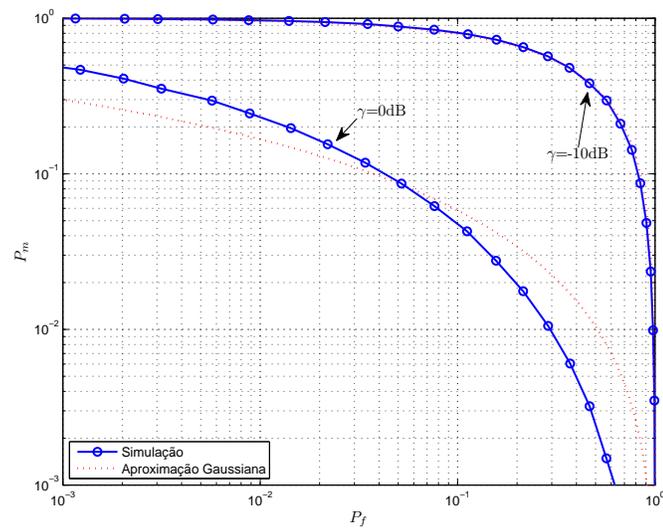


Figura 3.11: ROC considerando $M = 32$ e dois valores diferentes de SNR $\gamma = 0$ dB e $\gamma = -10$ dB.

3.2.2 Uma outra interpretação da fusão linear

Uma outra possibilidade para a análise de fusão de dados do tipo suave seria aplicar o detector ótimo considerando a própria distribuição de probabilidade das estatísticas de teste, sem explorar a hipótese de gaussianidade.

Observando que, sob a hipótese \mathcal{H}_0 , a estatística de teste apresenta distribuição do tipo χ^2 centralizada com M graus de liberdade, e sob a hipótese \mathcal{H}_1 , a estatística de teste apresenta distribuição do tipo χ^2 descentralizada com parâmetro de descentralização $\mu = M\gamma$ e M graus de liberdade, pode-se concluir que a razão de verossimilhança é dada por [14]:

$$\mathcal{L} = A_1 \frac{\prod_{k=0}^{K-1} \exp\left(-\frac{y_k + M\gamma_k}{2}\right) \left(\frac{y_k}{M\gamma_k}\right)^{M/4-1/2} I_{M/2-1}\left(\sqrt{M\gamma_k y_k}\right)}{\prod_{k=0}^{K-1} y_k^{M/2-1} \exp\left(-\frac{y_k}{2}\right)}, \quad (3.48)$$

em que γ_k é a SNR em cada rádio cognitivo e A_1 é uma constante dada por:

$$A_1 = 2^{M/2-K} \Gamma(M/2). \quad (3.49)$$

A eq. (3.48) pode ser reescrita como:

$$\mathcal{L} = A_2 \prod_{k=0}^{K-1} \left(\frac{y_k}{M\gamma_k}\right)^{-M/4+1/2} I_{M/2-1}\left(\sqrt{M\gamma_k y_k}\right), \quad (3.50)$$

em que A_2 é uma outra constante dada por:

$$A_2 = A_1 \prod_{k=0}^{K-1} \left(\frac{1}{M\gamma_k}\right)^{M/4-1/2} \exp\left(-\frac{M\gamma_k}{2}\right). \quad (3.51)$$

Lembrando que a função de Bessel modificada de primeira espécie assume a seguinte expressão em Série de Potências [50]:

$$I_n(z) = \left(\frac{z}{2}\right)^n \sum_{k=0}^{\infty} \left(\frac{z^2}{4}\right)^k \frac{1}{k! \Gamma(n+k+1)}. \quad (3.52)$$

Substituindo (3.52) em (3.50), temos que:

$$\mathcal{L} = A_2 \prod_{k=0}^{K-1} \left(\frac{y_k}{M\gamma_k} \right)^{-M/4+1/2} \left(\frac{\sqrt{M\gamma_k y_k}}{2} \right)^{M/2-1} \sum_{m=0}^{\infty} \left(\frac{M\gamma_k y_k}{4} \right)^m \frac{1}{m! \Gamma(M/2 + m)}, \quad (3.53)$$

Portanto:

$$\mathcal{L} = A_3 \prod_{k=0}^{K-1} \sum_{m=0}^{\infty} \left(\frac{M\gamma_k y_k}{4} \right)^m \frac{1}{m! \Gamma(M/2 + m)}, \quad (3.54)$$

em que A_3 é um terceira constante dada por:

$$A_3 = A_2 \prod_{k=0}^{K-1} \left(\frac{M\gamma_k}{2} \right)^{M/2-1}. \quad (3.55)$$

Expandindo o somatório, tem-se que:

$$\mathcal{L} = A_3 \prod_{k=0}^{K-1} \left(1 + \frac{M\gamma_k y_k}{4\Gamma(M/2)} + \left(\frac{M\gamma_k y_k}{4} \right)^2 \frac{1}{\Gamma(M/2 + 1)} + \dots \right). \quad (3.56)$$

Aproximando em primeira ordem e desprezando os termos cruzados, temos:

$$\mathcal{L} = A_3 \frac{M}{4\Gamma(M/2)} \sum_{k=0}^{K-1} \gamma_k y_k. \quad (3.57)$$

Finalmente, o teste ótimo é dado por:

$$\sum_{k=0}^{K-1} \gamma_k y_k \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda. \quad (3.58)$$

É interessante observar o impacto da aproximação linear na determinação do teste ótimo. O primeiro ponto a ser notado é que a aproximação linear é melhor para valores baixos de SNR ($\gamma_k \ll 1$), assim como ocorre com a aproximação assumindo gaussianidade.

Um segundo ponto interessante a ser observado é que a razão de verossimilhança pode ser escrita da seguinte forma:

$$\mathcal{L} = A_3 \prod_{k=0}^{K-1} \mathcal{L}_k, \quad (3.59)$$

em que:

$$\mathcal{L}_k = \sum_{m=0}^{\infty} \left(\frac{M\gamma_k y_k}{4} \right)^m \frac{1}{m! \Gamma(M/2 + m)}. \quad (3.60)$$

Para o caso em que M é um número par (que é o caso mais frequente, uma vez que é comum transmitir sinais com representação complexa em banda-base, dando origem a um número par de amostras por símbolo), pode-se escrever $\Gamma(M/2 + m) = (M/2 + m - 1)!$. Dessa forma, tem-se que:

$$\mathcal{L}_k = \sum_{m=0}^{\infty} \left(\frac{M\gamma_k y_k}{4} \right)^m \frac{1}{m! (M/2 + m - 1)!} \quad (3.61)$$

ou ainda:

$$\mathcal{L}_k = \frac{1}{(M/2)!} \sum_{m=0}^{\infty} \left(\frac{M\gamma_k y_k}{4} \right)^m \frac{1}{m! \prod_{r=1}^{m-1} \left(\frac{M}{2} + r \right)}. \quad (3.62)$$

Pode-se notar que (3.62) apresenta vários aspectos em comum com a expansão em Série de Taylor de uma função exponencial, o que motiva expressar (3.62) como:

$$\mathcal{L}_k \approx \frac{M/2}{(M/2)!} \sum_{m=0}^{\infty} \left(\frac{M\gamma_k y_k}{4} \right)^m \frac{1}{m! \left(\frac{M}{2} \right)^m}. \quad (3.63)$$

Finalmente:

$$\mathcal{L}_k \approx \frac{\exp\left(\frac{\gamma_k y_k}{2}\right)}{(M/2 - 1)!}. \quad (3.64)$$

Portanto, a razão de verossimilhança pode ser aproximada por:

$$\mathcal{L} \approx \frac{A_3}{(M/2 - 1)!} \prod_{k=0}^{K-1} \exp\left(\frac{\gamma_k y_k}{2}\right) \quad (3.65)$$

e um teste sub-ótimo é dado pelo logaritmo da razão de verossimilhança aproximado por:

$$T_{MRC}(\mathbf{y}) = \sum_{k=0}^{K-1} \gamma_k y_k \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda. \quad (3.66)$$

Note que esse teste é equivalente a uma fusão de dados do tipo *maximal-ratio combining* (MRC), em que é realizada uma combinação linear, cujos pesos de ponderação são determinados pela SNR em cada rádio cognitivo.

Com o objetivo de avaliar o impacto da aproximação linear na precisão da detecção, pode-se analisar o grau de correlação entre a estatística de teste obtida com

a expressão exata da razão de verossimilhança e aquela obtida com a aproximação linear.

Caso seja possível expressar T_{MRC} da seguinte forma:

$$T_{MRC} = \alpha T_{LLR} + \beta, \quad (3.67)$$

o teste ótimo dado por:

$$T_{LLR} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda, \quad (3.68)$$

torna-se equivalente a:

$$T_{MRC} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \underbrace{\alpha\lambda + \beta}_{\lambda_{MRC}}. \quad (3.69)$$

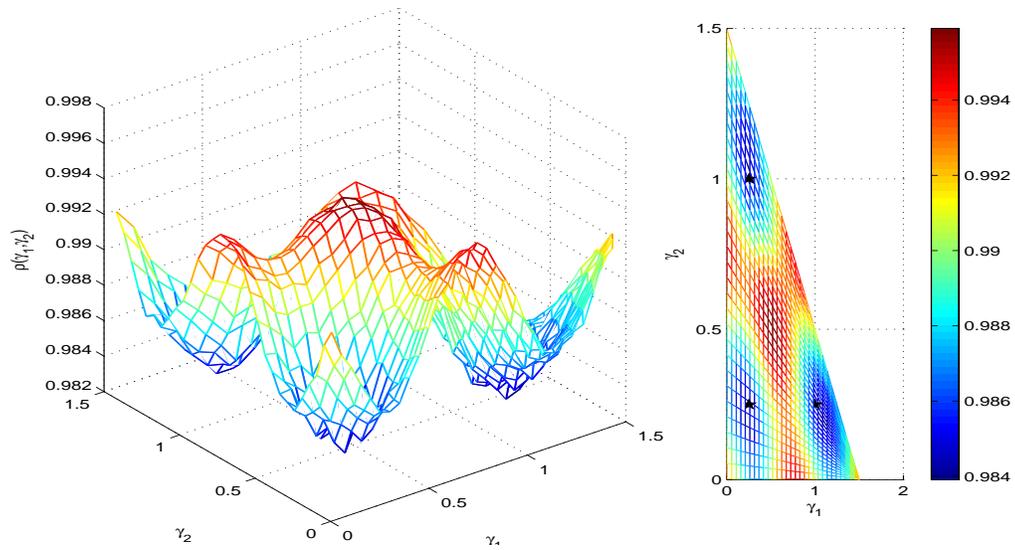
Portanto, o teste dado em (3.69) passa a ser equivalente ao teste ótimo.

Uma condição necessária para garantir a validade de (3.67) é que o coeficiente de correlação seja próximo de 1, sob as duas hipóteses. Com o objetivo de testar essa suposição, primeiramente analisaremos a influência da SNR média na aproximação linear. Para tanto, considera-se um sistema com três rádios cognitivos e SNR média $\bar{\gamma} = -3\text{dB}$ e $M = 8$.

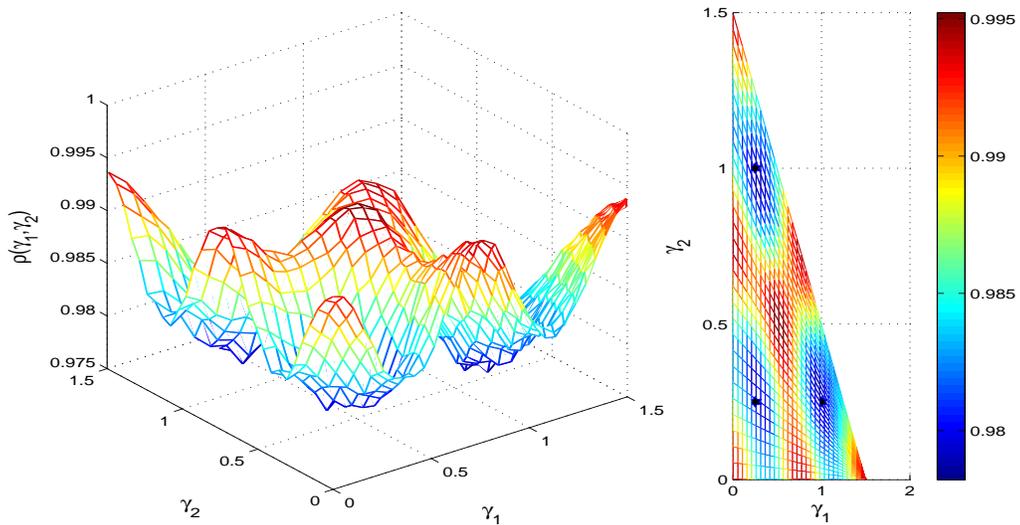
A Fig. 3.12a mostra o coeficiente de correlação estimado entre a estatística de teste obtida com a verossimilhança exata e a estatística de teste gerada com a aproximação linear e sob a hipótese de sinal. Já a Fig. 3.12b mostra o coeficiente de correlação estimado entre as duas estatísticas de teste e sob a hipótese nula.

Note que o coeficiente de correlação é muito próximo da unidade para todas as configurações de SNR. Esse fato é um indicativo que a estatística de teste obtida com a aproximação linear é uma aproximação muito boa da estatística de teste ótima para esse valor de SNR média.

Para as configurações de SNR em que dois dos rádios cognitivos apresentam SNR = -6dB e o terceiro apresenta SNR = 0dB , o coeficiente de correlação apresenta o seu valor mínimo. Essa deve ser a situação em que a diferença de desempenho entre a aproximação linear e o teste ótimo é máxima. Entretanto, nota-se que mesmo nesse caso extremo, o coeficiente de correlação ainda é muito próximo da unidade. Portanto, para valores baixos de SNR, a aproximação linear deve prover resultados

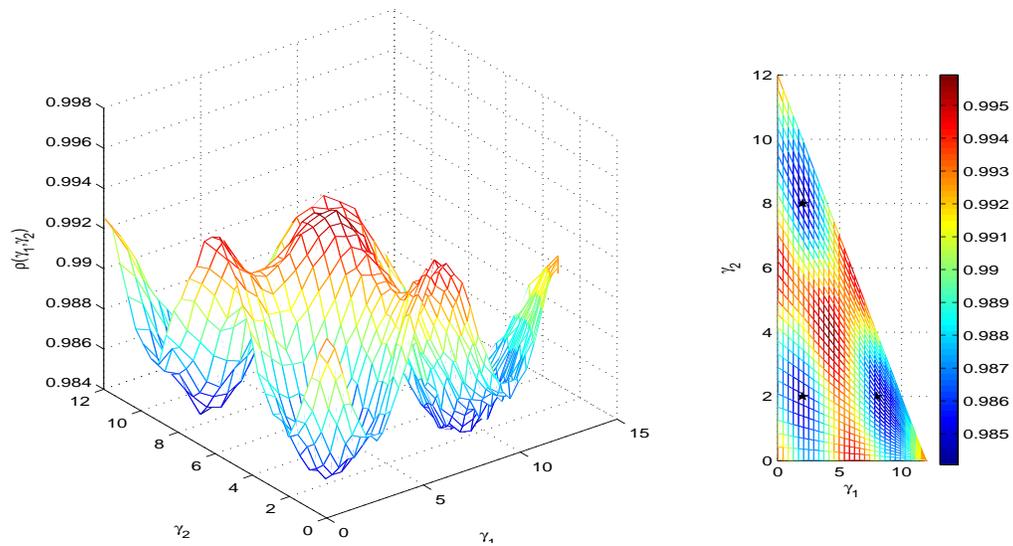


(a) Hipótese nula

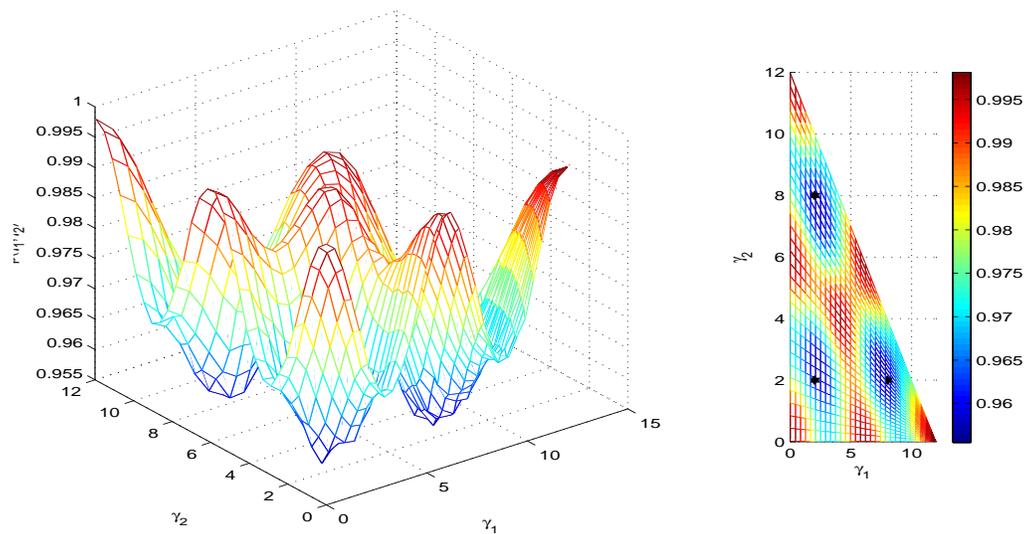


(b) Hipótese de sinal

Figura 3.12: Coeficiente de correlação entre a estatística de teste real e a aproximação linear sob a hipótese de sinal com SNR média $\bar{\gamma} = -3dB$. Os pontos destacados correspondem aos pontos de correlação mínima. Os valores de SNR estão dados em escala linear.



(a) Hipótese nula



(b) Hipótese de sinal

Figura 3.13: Coeficiente de correlação entre a estatística de teste real e a aproximação linear sob a hipótese de sinal com SNR média $\bar{\gamma} = 6dB$. Os pontos destacados correspondem aos pontos de correlação mínima. Os valores de SNR estão dados em escala linear.

de detecção muito próximos do teste ótimo.

Repetindo o mesmo procedimento para um valor de SNR média mais elevado $\bar{\gamma} = 6\text{dB}$, nota-se que o formato das curvas de correlação se mantem aproximadamente constante. Entretanto, nota-se que os valores mínimos de autocorrelação diminuem em relação ao caso de SNR média $\bar{\gamma} = -3\text{dB}$. Portanto, para esse caso, espera-se que a aproximação linear apresente um resultado inferior ao obtido com a estatística de teste ótima.

Portanto, conclui-se que, para valores crescentes de SNR média, em certas configurações, a aproximação linear pode se distanciar do detector ótimo.

Um outro ponto que deve ser observado é o comportamento das curvas que relacionam as duas estatística de teste. Se o coeficiente de correlação é próximo da unidade, pode-se afirmar que há uma relação aproximadamente linear entre as duas estatística de teste. Todavia, para que os testes sejam de fato equivalentes, é necessário que essa relação linear seja a mesma sob as duas hipóteses, *i.e.* os coeficientes α e β em (3.67) devem ser os mesmos tanto sob a hipótese de sinal como também para a hipótese nula. As aproximações lineares sob as duas hipóteses podem ser comparadas analisando as curvas das Figs. 3.14 e 3.15, onde são mostrados gráficos de dispersão das duas variáveis aleatórias para as duas configurações de SNR. Os gráficos de dispersão foram gerados para a situação de pior caso para a aproximação linear destacadas nas Figs. 3.12a, 3.12b, 3.13a e 3.13b.

Os gráficos de dispersão mostram que, para valores baixos de SNR média, há uma relação praticamente linear entre a estatística de teste ótima e a estatística de teste obtida com a aproximação linear sob ambas as hipóteses. Ainda, em ambos os casos, as retas que aproximam as duas estatísticas têm praticamente a mesma inclinação. Já para valores mais elevados de SNR, a inclinação das retas que aproximam as estatísticas de teste são diferentes. Além disso, nota-se que há uma dispersão maior no caso de hipótese de sinal. Esse fato já era esperado, uma vez que as curvas de coeficientes de correlação já mostravam que essa observação se mantém válida apenas sob a hipótese nula. Por tal motivo, espera-se que haja uma degradação do desempenho do detector linear para valores mais elevados de SNR.

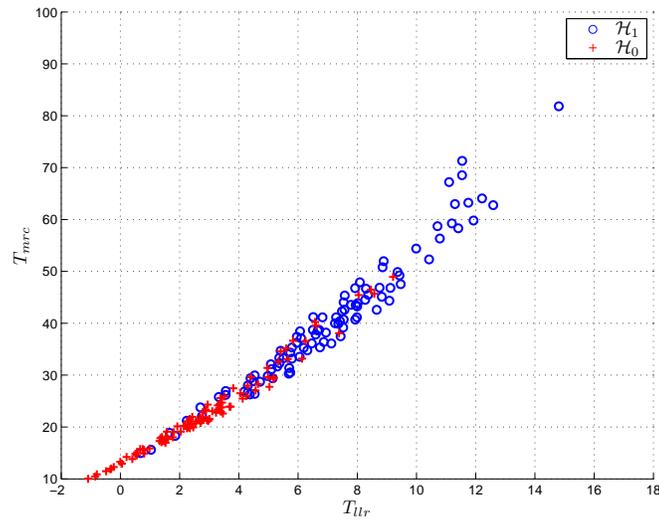


Figura 3.14: Gráficos de dispersão considerando um sistema com 3 rádios cognitivos e SNR média $\bar{\gamma} = -3\text{dB}$

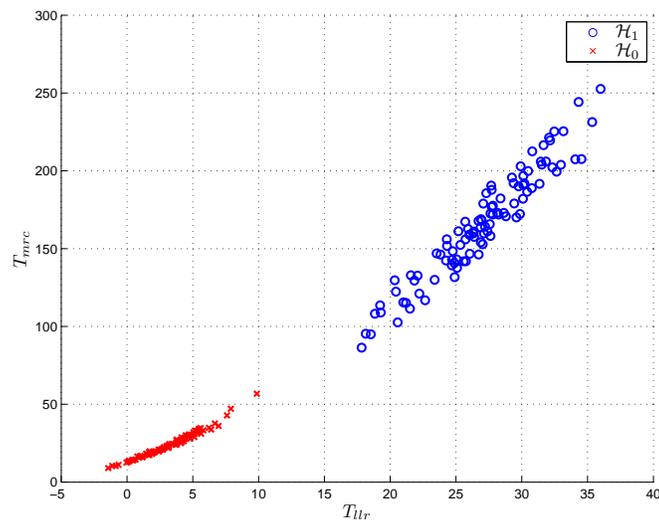


Figura 3.15: Gráficos de dispersão considerando um sistema com 3 rádios cognitivos e SNR média igual a $\bar{\gamma} = 6\text{dB}$.

Um outro fator que merece atenção é a influência do parâmetro M na aproximação linear. Pelo desenvolvimento teórico, espera-se que a aproximação linear seja mais precisa para valores crescentes de M , já que sua validade está associada à aproximação de $(M/2 + m - 1)!$ por $(M/2)!(M/2)^{m-1}$.

A Fig. 3.16 mostra o valor mínimo de correlação em função do parâmetro M e de $\bar{\gamma}$ nos rádios cognitivos, sob a hipótese de sinal. Observa-se que, sob a hipótese de sinal, a aproximação linear realmente é mais precisa para valores mais elevados de M . Entretanto, a influência do parâmetro M é muito mais sutil do que a influencia da SNR média do sistema e, como já esperado, o coeficiente de correlação é uma função decrescente da SNR média do sistema. A tabela 3.1 mostra o coeficiente de correlação em função do parâmetro M , note que o coeficiente de correlação também é maior para valores crescentes de M .

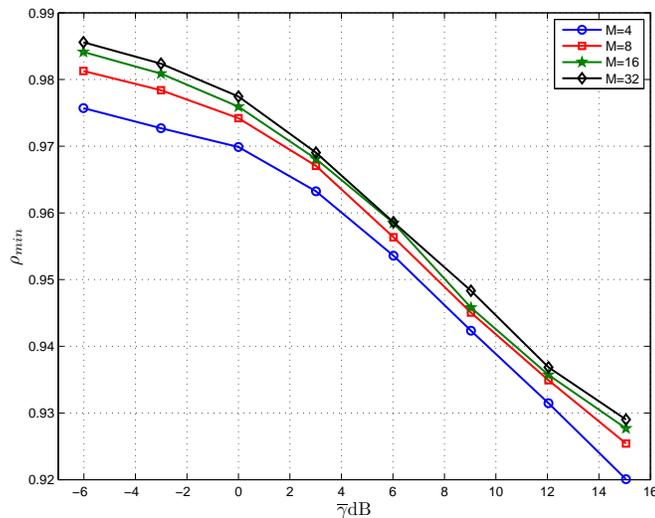


Figura 3.16: Coeficiente de correlação mínimo em função da SNR média $\bar{\gamma}$ sob a hipótese de sinal.

M	4	8	16	32
ρ_{min}	0,9785	0,9841	0,9871	0,9884

Tabela 3.1: Valor Mínimo do coeficiente de correlação, sob a hipótese nula, em função do parâmetro M .

Resta analisar o impacto da aproximação linear nas curvas de ROC. Como o parâmetro M pouco influencia na precisão da aproximação linear, fixou-se $M =$

8 e o sistema foi simulado na condição de pior caso para a aproximação linear. Analisando a Fig. 3.17, pode-se notar que, para valores baixos de $\bar{\gamma}$, o resultado obtido com a aproximação linear é equivalente àquele obtido com o teste de razão de verossimilhança. Para valores altos de SNR, apesar de haver uma perda de desempenho do detector linear em relação ao detector ótimo, o detector linear ainda é comparável ao detector ótimo.

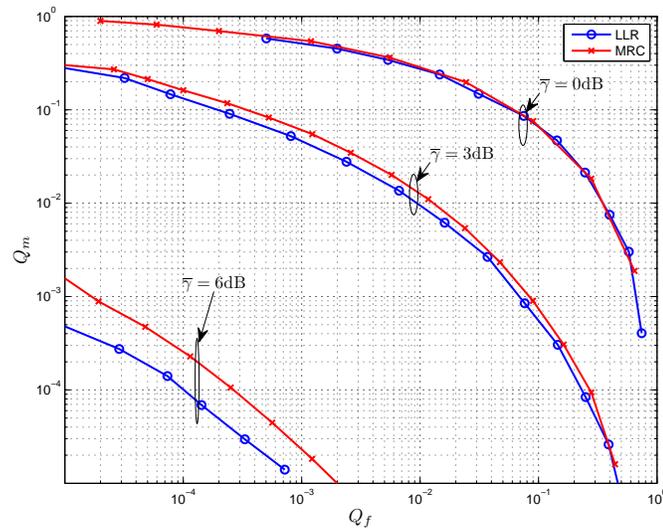


Figura 3.17: Curvas de ROC para diversos valores de $\bar{\gamma}$.

3.3 Conclusões

Este Capítulo tratou da análise de sistemas de sensoriamento espectral cooperativos centralizados. Primeiramente, na Seção 3.1, foi considerado um sistema no qual os rádios cognitivos enviam decisões do tipo dura para o centro de fusão. Para esse tipo de sistema, foi desenvolvida uma extensa análise considerando que o canal entre o rádio cognitivo e o centro de fusão é modelado como um canal BSC com probabilidade de erro P_e . A principal contribuição dessa análise foi prover limitantes de desempenho teóricos para o sistema e discutir o impacto que altas valores de P_e trazem para o sistema a depender do tipo de regra de decisão utilizada. Os principais resultados dessa Seção foram publicados em [10].

Na Seção 3.2 foi feita uma análise do sistema de sensoriamento espectral coope-

rativo considerando envio de decisão suave para o centro de fusão. Primeiramente, mostrou-se a fragilidade da hipótese de gaussianidade comumente aceita na literatura e, posteriormente, analisou-se a estatística de teste ótima obtida com as expressões exatas de verossimilhança. Foi mostrado que em sistemas com baixo valor de SNR média, essa estatística de teste ótimo tende ao decisor do tipo MRC. Resultado semelhante foi obtido em [47], onde assumiu-se a hipótese de gaussianidade.

Capítulo 4

Estimação Distribuída

Em redes de sensores, a estimação pode ser realizada de modo completamente cooperativo. Nessa situação, não é necessária a presença de um centro de fusão, pois os nós da rede trocam informações diretamente entre si e a decisão é realizada nos próprios sensores.

Neste Capítulo, será dada atenção ao caso de estimação distribuída, em que se considera uma rede na qual K sensores estimam um parâmetro em comum. Já o caso de detecção cooperativa, em que K sensores de uma rede têm o objetivo de detectar um parâmetro comum, bem como sua aplicação ao problema de sensoriamento espectral, serão os tópicos abordados no Capítulo 5.

Este Capítulo está organizado em 4 Seções. Na Seção 4.1 será apresentada uma introdução à classificação de topologias de rede. Já na Seção 4.2, será descrito o algoritmo LMS distribuído e será desenvolvida uma análise para atribuição ótima de passos de adaptação para cada nó da rede. Além disso, ainda na Seção 4.2, serão encontrados limitantes de desempenho da rede de adaptação para diferentes topologias e será proposto um método de otimização dos parâmetros da rede. Na Seção 4.3 são apresentados alguns resultados envolvendo o algoritmo *set-membership* distribuído. Finalmente, na Seção 4.4, são apresentadas algumas conclusões referentes ao Capítulo.

4.1 Topologias de rede

Considere uma rede de sensores com K nós (sensores), seja o conjunto \mathcal{N}_k formado por todos os nós conectados ao k -ésimo nó. Dada essa topologia de rede, define-se para cada nó o seu grau: $n_k = |\mathcal{N}_k|$, *i.e.*, n_k corresponde ao número de nós conectados ao k -ésimo nó. Considera-se que os caminhos da rede são canais ideais e que se trata de uma rede bidirecional, *i.e.*, se o nó k está conectado ao nó n , o nó n está conectado ao nó k . A cada rede, associa-se uma matriz de combinação que pode ser definida de várias maneiras [51]:

- Metrópolis

Para a regra Metrópolis, os elementos da matriz de combinação C são dados por:

$$c_{k,l} = \begin{cases} \frac{1}{\max(n_k, n_l)}, & k \neq l, l \in \mathcal{N}_k \\ 0, & l \notin \mathcal{N}_k \\ 1 - \sum_{l \in \mathcal{N}_k, l \neq k} c_{k,l}, & k = l \end{cases} . \quad (4.1)$$

- Laplace

A aplicação da regra de Laplace fornece a seguinte matriz de combinação:

$$\mathbf{C} = \mathbf{I} - \kappa \mathcal{L}_p, \quad (4.2)$$

em que \mathbf{I} é a matriz identidade, $\kappa = 1/n_{max}$ e \mathcal{L}_p , por sua vez, é definida por:

$$\mathcal{L}_p = \mathcal{D} - \mathbf{A}_d, \quad (4.3)$$

sendo $\mathcal{D} = \text{diag}\{n_1, \dots, n_M\}$ e \mathbf{A}_d é a matriz de rede adjacente dada por:

$$\mathbf{A}_{d,k,l} = \begin{cases} 1, & l \in \mathcal{N}_k \\ 0, & l \notin \mathcal{N}_k \end{cases} . \quad (4.4)$$

- Regra do vizinho

Finalmente, a regra do vizinho é tal que o elemento da matriz de combinação é dado por:

$$c_{k,l} = \begin{cases} \frac{1}{n_k}, & l \in \mathcal{N}_k \\ 0, & l \notin \mathcal{N}_k \end{cases} , \quad (4.5)$$

Outras matrizes de combinação podem ser adotadas, desde que as seguintes condições sejam satisfeitas [52]:

$$\begin{cases} \mathbf{C}\mathbf{1} = \mathbf{1} \\ c_{k,l} = 0, \quad l \notin \mathcal{N}_k \end{cases}, \quad (4.6)$$

em que $\mathbf{1}$ é um vetor coluna de dimensão K com todos os elementos iguais a um.

Definida a matriz de combinação, os algoritmos clássicos de filtragem adaptativa [53], [54], [55] podem ser aplicados ao contexto de processamento distribuído de sinais. A seguir, será mostrado como, a partir de uma dada matriz de combinação, estender os resultados de filtragem adaptativa clássica para o caso de redes de adaptação.

4.2 LMS distribuído

Localmente, cada nó na rede pode aplicar vários algoritmos adaptativos para atualizar suas estimativas. Com o objetivo de manter a complexidade computacional baixa, é comum utilizar versões do algoritmo LMS no processamento local.

Dentre as versões de LMS distribuído, é possível utilizar algoritmos do tipo consenso, os quais são compostos por uma fase em que os nós realizam as medidas e uma outra, em que as estimativas são trocadas. Essa é a abordagem utilizada em [56], [57]. Em contraste, é ainda possível combinar as duas fases em uma única, dando origem aos algoritmos de difusão mostrados em [52], [58], [59], [60] e [51]. Em [59], é mostrado que as estratégias de difusão são mais eficientes do que as estratégias de consenso, no sentido de que o algoritmo de consenso leva a um menor erro de estimação para um dado número de iterações.

Nos restringindo às estratégias de difusão, em [52] são apresentadas duas possíveis alternativas para a implementação do algoritmo. A primeira consiste em, para cada iteração do algoritmo, inicialmente combinar as estimativas obtidas nos nós da rede e depois adaptá-las. Essa estratégia é conhecida como *combine-then-adapt* (CTA). A outra possibilidade seria primeiro adaptar o algoritmo e depois combinar as estimativas provenientes dos outros nós da rede, dando origem à técnica *adapt-then-combine* (ATC). Em [59] e [52] é mostrado que a técnica ATC é mais eficiente

LMS distribuído - CTA
$\boldsymbol{\psi}_k(n) = \sum_{l \in \mathcal{N}_k} c_{k,l} \boldsymbol{\phi}_l(n)$
$e(n) = d(n) - \boldsymbol{\psi}_k(n)^T \mathbf{u}_k(n)$
$\boldsymbol{\phi}_k(n+1) = \boldsymbol{\psi}_k(n) + \mu_k e(n) \mathbf{u}_k(n)$

Tabela 4.1: Algoritmo LMS distribuído - CTA

LMS distribuído - ATC
$e(n) = d(n) - \boldsymbol{\psi}_k(n)^T \mathbf{u}_k(n)$
$\boldsymbol{\phi}_k(n+1) = \boldsymbol{\psi}_k(n) + \mu_k e(n) \mathbf{u}_k(n)$
$\boldsymbol{\psi}_k(n) = \sum_{l \in \mathcal{N}_k} c_{k,l} \boldsymbol{\phi}_l(n)$

Tabela 4.2: Algoritmo LMS distribuído - ATC

do que a técnica CTA, principalmente em cenários nos quais o passo de adaptação do algoritmo é pequeno. Uma outra forma de melhorar o desempenho do algoritmo seria promover a troca de medidas entre os nós da rede, além da troca de estimativas. Dessa forma, é possível conseguir estimativas menos ruidosas em troca de um aumento na complexidade computacional.

Ao longo desse Capítulo, a seguinte notação será utilizada: \mathbf{w}_{opt} indica o vetor ótimo comum a todos nós que de ser estimado na rede, μ_k indica o passo de adaptação associado ao k -ésimo nó da rede, $\boldsymbol{\psi}_k(n)$ corresponde ao sinal estimado no nó k e $\boldsymbol{\phi}_k(n)$ corresponde ao vetor obtido pela combinação linear das estimativas geradas nos nós vizinhos ao k -ésimo nó, o sinal regressor no nó k é definido como $\mathbf{u}_k(n)$ e o sinal desejado em cada nó obedece à seguinte relação:

$$d_k(n) = \mathbf{u}_k^T(n) \mathbf{w}_{opt}(n) + \boldsymbol{\eta}_k(n). \quad (4.7)$$

As versões CTA e ATC do algoritmo LMS por difusão estão mostradas nas Tabelas 4.1 e 4.2, respectivamente.

A técnica de difusão além de promover uma diminuição do erro quadrático médio e um aumento da velocidade de convergência, também apresenta melhores condições de estabilidade. Esse fato é mostrado em [51], em que os modos da rede são calculados, e conclui-se que a cooperação leva a autovalores da matriz de autocorrelação de entrada de menor magnitude o que contribui para a estabilidade do filtro.

4.2.1 Otimização da escolha do passo de adaptação

É comum considerar que todos os nós das redes apresentam o mesmo passo de adaptação, *i. e.*, $\mu_k = \mu, \forall k$ [51], [60].

Todavia, em redes heterogêneas, onde cada nó apresenta uma potência de ruído diferente, espera-se que os passos de adaptação nos nós das redes também sejam diferentes.

Com o objetivo de responder a essa questão, analisaremos uma rede completamente conectada. Nesse caso, a matriz de combinação \mathbf{C} deve ser escolhida como:

$$\mathbf{C} = \frac{1}{K} \mathbf{1}\mathbf{1}^T, \quad (4.8)$$

em que K é o número de nós na rede e $\mathbf{1}$ é um vetor coluna K -dimensional com todos os elementos iguais a um.

Uma abordagem para analisar a taxa de convergência consiste em introduzir o vetor de erro de coeficientes da seguinte forma:

$$\tilde{\boldsymbol{\psi}}(n) = \left[\tilde{\boldsymbol{\psi}}_1^T(n) \ \tilde{\boldsymbol{\psi}}_2^T(n) \ \cdots \ \tilde{\boldsymbol{\psi}}_K^T(n) \right]^T, \quad (4.9)$$

em que

$$\tilde{\boldsymbol{\psi}}_k = \boldsymbol{\psi}_k - \mathbf{w}_{opt}. \quad (4.10)$$

Primeiramente, examinaremos o caso do algoritmo CTA, em seguida será mostrado que o resultado obtido pode ser facilmente generalizado para o caso ATC. Considerando a versão CTA do LMS distribuído, a seguinte recursão é válida [51]:

$$\mathbb{E} \left\{ \tilde{\boldsymbol{\psi}}(n) \right\} = (\mathbf{I}_{MK} - \mathbf{D}\mathbf{R}_u) \mathbf{G} \mathbb{E} \left\{ \tilde{\boldsymbol{\psi}}(n-1) \right\}, \quad (4.11)$$

onde

$$\mathbf{G} = \mathbf{C} \otimes \mathbf{I}_M, \quad (4.12)$$

$$\mathbf{D} = \boldsymbol{\mu} \otimes \mathbf{I}_M, \quad (4.13)$$

$$\mathbf{R}_u = \mathbf{I}_K \otimes \mathbf{R}, \quad (4.14)$$

sendo \mathbf{I}_K é a matriz identidade de dimensão K e \mathbb{E} denota o operador de esperança

estatística. Ainda, o símbolo \otimes denota o produto de Kronecker, a matriz $\boldsymbol{\mu}$ é uma matriz diagonal cujos elementos são os passos de adaptação em cada nó e a matriz \mathbf{R} é a matriz de autocorrelação da entrada $\mathbf{R} = E \{ \mathbf{u}_k \mathbf{u}_k^T \}$, que é considerada a mesma para todo nó k .

A taxa de convergência é determinada pelo raio espectral [51], *i.e.*, a máxima magnitude dos autovalores da matriz $\mathbf{A} = (\mathbf{I}_{MK} - \mathbf{D}\mathbf{R}_u) \mathbf{G}$.

Aplicando algumas propriedades do produto de Kronecker, pode ser mostrado que a matriz \mathbf{A} pode ser reescrita como:

$$\mathbf{A} = \mathbf{C} \otimes \mathbf{I}_M - \boldsymbol{\mu} \mathbf{C} \otimes \mathbf{R}. \quad (4.15)$$

Nas subseções seguintes, analisaremos, em separado, os casos de entrada branca e entrada colorida.

Entrada branca

Modelando o vetor regressor como um processo estocástico branco de potência unitária, a matriz \mathbf{A} pode ser reescrita como:

$$\mathbf{A} = (\mathbf{I}_K - \boldsymbol{\mu}) \mathbf{C} \otimes \mathbf{I}_M. \quad (4.16)$$

Definindo $\mathbf{A}_1 = (\mathbf{I}_K - \boldsymbol{\mu}) \mathbf{C}$, pode-se notar que os autovalores de \mathbf{A} são iguais aos autovalores de \mathbf{A}_1 . Portanto, os raios espectrais de \mathbf{A} e \mathbf{A}_1 são iguais [61].

A matriz \mathbf{A}_1 aplicada a um vetor \mathbf{v} com componentes v_i , $i = 1 \cdots K$ resulta em:

$$\mathbf{A}_1 \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_K \end{bmatrix} = \frac{1}{K} \sum_{i=1}^K v_i \begin{bmatrix} 1 - \mu_1 & 0 & \cdots & 0 \\ 0 & 1 - \mu_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 - \mu_K \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \quad (4.17)$$

Portanto:

$$\mathbf{A}_1 \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_K \end{bmatrix} = \frac{1}{K} \sum_{i=1}^K v_i \underbrace{\begin{bmatrix} 1 - \mu_1 \\ 1 - \mu_2 \\ \vdots \\ 1 - \mu_K \end{bmatrix}}_{\mathbf{v}_0}. \quad (4.18)$$

Logo, o vetor \mathbf{v}_0 é um autovetor de \mathbf{A}_1 associado ao autovalor $\lambda = 1 - \frac{1}{K} \sum_{i=1}^K \mu_i$.

Dessa forma, o raio espectral de \mathbf{A} é dado por:

$$\rho = 1 - \frac{1}{K} \sum_{i=1}^K \mu_i. \quad (4.19)$$

Finalmente, podemos concluir que se a média dos passos for constante, a taxa de convergência do algoritmo também será constante.

Apesar da demonstração acima ser válida para a rede completamente conectada, em que a matriz de conectividade é dada por (4.8), o resultado em (4.19) é ainda uma boa aproximação para casos mais gerais, em que \mathbf{C} é uma matriz estocástica qualquer. A fim de ilustrar isso, consideraremos um exemplo com uma rede de 3 nós, em que a matriz de combinação é dada por:

$$\mathbf{C} = \begin{bmatrix} c_1 & c_2 & 1 - c_1 - c_2 \\ 1 - c_1 - c_2 & c_1 & c_2 \\ c_2 & 1 - c_1 - c_2 & c_1 \end{bmatrix}. \quad (4.20)$$

com $0 \leq c_1 \leq 1$ e $0 \leq c_2 \leq 1 - c_1$. Duas configurações de passo de adaptação com a mesma média são analisadas: $\boldsymbol{\mu}_1 = [0,3 \ 0,1 \ 0,2]$ e $\boldsymbol{\mu}_2 = [0,39 \ 0,01 \ 0,2]$. A Fig. 4.1 mostra o raio espectral relativo a matrizes de combinação da forma (4.20). Pode-se notar pela Fig. 4.1 que (4.19) é uma boa aproximação para o raio espectral para uma larga faixa valores de c_1 e c_2 . A expressão teórica perde sua validade se $c_1 \approx 1$, caso em que a matriz de combinação tende à identidade, o que leva o algoritmo a se aproximar do caso não-cooperativo. Nessa situação, o raio espectral será dado por $\rho = 1 - \mu_{min}$, sendo que, para a configuração de passo $\boldsymbol{\mu}_1$, tem-se que $\rho = 0,9$, e para a configuração de passo $\boldsymbol{\mu}_2$, tem-se que $\rho = 0,99$.

Com o objetivo de verificar a validade da hipótese de rede completamente conectada, é necessário analisar os autovalores de $\mathbf{A}_1 = (\mathbf{I}_K - \boldsymbol{\mu}) \mathbf{C}$.

Por definição, $\mathbf{v} = \mathbf{1}$ é um autovetor de \mathbf{C} associado ao autovalor $\lambda = 1$. Portanto, tem-se que:

$$(\mathbf{I}_K - \boldsymbol{\mu}) \mathbf{C} \mathbf{v} = (\mathbf{I}_K - \boldsymbol{\mu}) \mathbf{v}. \quad (4.21)$$

Nesse ponto, apresenta-se a seguinte Proposição:

Proposition 4.2.1 (Decomposição matriz diagonal). *Seja \mathbf{D} uma matriz diagonal*

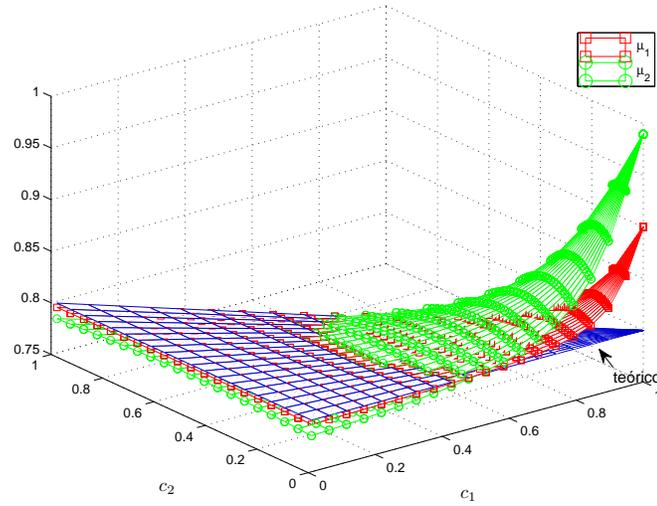


Figura 4.1: Raio espectral para as configurações de passo μ_1 e μ_2

de dimensão 2^M , sendo $M \in \mathbb{Z}$, é possível decompor \mathbf{D} como:

$$\mathbf{D} = \frac{1}{K} \sum_{k=0}^{K-1} \mathbf{h}_k^T \mathbf{d} \cdot \text{diag} \{ \mathbf{h}_k \}, \quad (4.22)$$

em que \mathbf{d} é um vetor coluna cujos elementos são os elementos diagonais de \mathbf{D} e \mathbf{h}_k é a k -ésima coluna da matriz de Walsh-Hadamard \mathbf{H} , definida recursivamente como:

$$\mathbf{H}_{2^n} = \mathbf{H}_2 \otimes \mathbf{H}_{2^{n-1}}. \quad (4.23)$$

Demonstração. Expandindo o termo $\mathbf{h}_k^T \mathbf{d} \cdot \text{diag} \{ \mathbf{h}_k \}$, tem-se que:

$$\mathbf{h}_k^T \mathbf{d} \cdot \text{diag} \{ \mathbf{h}_k \} = \left(\sum_{q=0}^{K-1} d_q h_{q,k} \right) \begin{bmatrix} h_{1,k} & & & \\ & h_{2,k} & & \\ & & \dots & \\ & & & h_{K,k} \end{bmatrix}. \quad (4.24)$$

Portanto, o m -ésimo elemento da diagonal de $\sum_{k=0}^{K-1} \mathbf{h}_k^T \mathbf{d} \cdot \text{diag} \{ \mathbf{h}_k \}$ é:

$$\left\{ \sum_{k=0}^{K-1} \mathbf{h}_k^T \mathbf{d} \cdot \text{diag} \{ \mathbf{h}_k \} \right\}_m = \sum_{k=0}^{K-1} \left(\sum_{q=0}^{K-1} d_q h_{q,k} \right) h_{m,k}. \quad (4.25)$$

Reordenando a ordem do somatório:

$$\left\{ \sum_{k=0}^{K-1} \mathbf{h}_k^T \mathbf{d} \cdot \text{diag} \{ \mathbf{h}_k \} \right\}_m = \sum_{q=0}^{K-1} \left(\sum_{k=0}^{K-1} h_{m,k} h_{q,k} \right) d_q. \quad (4.26)$$

Sabendo que as linhas da matriz de Walsh-Hadamard são ortogonais, conclui-se que:

$$\left\{ \sum_{k=0}^{K-1} \mathbf{h}_k^T \mathbf{d} \cdot \text{diag} \{ \mathbf{h}_k \} \right\}_m = \sum_{q=0}^{K-1} (K \delta_{m,q}) d_q, \quad (4.27)$$

em que $\delta_{m,q}$ é a função delta de Kronecker:

$$\delta_{m,q} = \begin{cases} 1, & m = q \\ 0, & m \neq q \end{cases}. \quad (4.28)$$

Finalmente:

$$d_m = \frac{1}{K} \left\{ \sum_{k=0}^{K-1} \mathbf{h}_k^T \mathbf{d} \cdot \text{diag} \{ \mathbf{h}_k \} \right\}_m. \quad (4.29)$$

□

Aplicando o resultado acima, a matriz $(\mathbf{I}_K - \boldsymbol{\mu})$ pode ser reescrita como:

$$(\mathbf{I}_K - \boldsymbol{\mu}) = \frac{1}{K} \sum_{k=0}^{K-1} \mathbf{h}_k^T \mathbf{d} \cdot \text{diag} \{ \mathbf{h}_k \}, \quad (4.30)$$

em que \mathbf{d} é um vetor coluna cujos elementos são os elementos diagonais de $(\mathbf{I}_K - \boldsymbol{\mu})$.

Como a primeira coluna da matriz de Walsh-Hadamard é dada por $\mathbf{1}$, $(\mathbf{I}_K - \boldsymbol{\mu})$ ainda pode ser expressa como:

$$(\mathbf{I}_K - \boldsymbol{\mu}) = \left\{ \frac{1}{K} \sum_{k=0}^{K-1} d_k \right\} \mathbf{I}_K + \frac{1}{K} \sum_{k=1}^{K-1} \mathbf{h}_k^T \mathbf{d} \cdot \text{diag} \{ \mathbf{h}_k \}. \quad (4.31)$$

Além disso, é importante notar que as demais colunas da matriz de Walsh-Hadamard são compostas por $K/2$ elementos iguais a 1 e $K/2$ elementos iguais a -1 . Portanto, os termos da forma $\mathbf{h}_k^T \mathbf{d}$ para $k \geq 1$ são compostos por diferenças de termos d_k . Lembrando que $d_k = 1 - \mu_k$, conclui-se que se $\mu_k \ll 1$, a relação seguinte é válida:

$$\mathbf{h}_0^T \mathbf{d} = \sum_{k=0}^{K-1} d_k \gg \mathbf{h}_k^T \mathbf{d}, \quad k \geq 1. \quad (4.32)$$

Finalmente, pode-se aproximar (4.21) como:

$$(\mathbf{I}_K - \boldsymbol{\mu}) \mathbf{v} \approx \left(\frac{1}{K} \sum_{k=0}^{K-1} 1 - \mu_k \right) \mathbf{v}. \quad (4.33)$$

Portanto, observa-se que $1 - \bar{\mu}$ aproxima bem o raio espectral de $(\mathbf{I}_K - \boldsymbol{\mu}) \mathbf{C}$,

mesmo em casos em que a rede não é completamente conectada. É importante enfatizar que a aproximação é diretamente ligada à hipótese $\mu_k \ll 1$. Quanto maior o valor do passo de adaptação em cada nó da rede, pior a aproximação.

Esse resultado pode ser verificado se considerarmos uma rede com K nós e matriz de combinação dada por:

$$\mathbf{C} = (1 - K\alpha)\mathbf{I}_K + \alpha\mathbf{1}\mathbf{1}^T, \quad (4.34)$$

em que $0 \leq \alpha \leq \frac{1}{K}$. Se o parâmetro $\alpha = 0$, essa combinação tende ao modo não-cooperativo. Por outro lado, se $\alpha = \frac{1}{K}$, a combinação equivale ao modo totalmente conectado.

A Fig. 4.3 mostra o raio espectral associado às redes cujas matrizes de combinação dadas por (4.34) e com os passos de adaptação mostrados na Fig. 4.2. É possível notar que o raio espectral é, de fato, bem aproximado por (4.19). É ainda interessante notar que a aproximação é melhor para redes com mais nós.

Resta ainda analisar como o valor do passo de adaptação impacta na aproximação proposta. Para tanto, o tamanho da rede é fixado em $K = 36$ e o raio espectral é calculado para diferentes valores de passo de adaptação. O resultado está mostrado na Fig. 4.4. A curva em azul, corresponde ao raio espectral calculado considerando a configuração de passo mostrada na Fig. 4.2. As demais curvas correspondem a versões escalonadas da mesma configuração de passo. Os resultados indicam que a aproximação é melhor para valores pequenos de passo de adaptação, conforme já era esperado.

Entrada colorida

Nesse caso, consideramos uma entrada colorida com matriz de correlação dada por: $\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, em que \mathbf{U} é uma matriz unitária e $\mathbf{\Lambda}$ é uma matriz diagonal com elementos $\mathbf{\Lambda}(k, k) = \lambda_k$.

O raio espectral da matriz \mathbf{A} , nesse contexto, é dado por:

$$\rho = 1 - \lambda_{min} \frac{1}{K} \sum_{i=1}^K \mu_i. \quad (4.35)$$

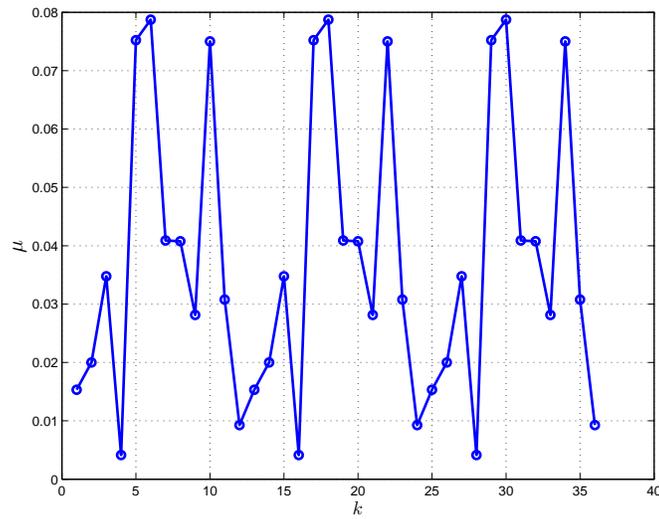


Figura 4.2: Passo de adaptação em cada nó da rede

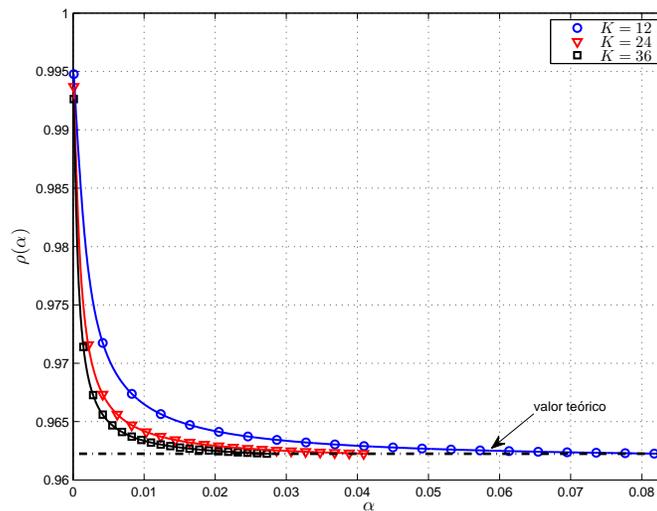


Figura 4.3: Raio espectral relativo às matrizes descritas por (4.34) para diferentes valores de K . O valor teórico está indicado pela curva tracejada.

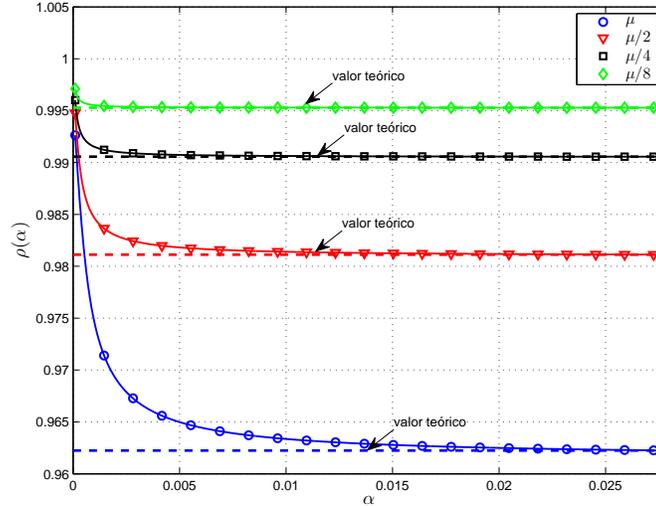


Figura 4.4: Raio espectral relativo às matrizes descritas por (4.34) para diferentes configurações de passo. O valor teórico está indicado pela curva tracejada.

A fim de provar (4.35), note que a matriz \mathbf{R}_U pode ser decomposta em $\mathbf{R}_U = (\mathbf{I}_K \otimes \mathbf{U}) (\mathbf{I}_K \otimes \mathbf{\Lambda}) (\mathbf{I}_K \otimes \mathbf{U}^H)$. Portanto, podemos escrever \mathbf{A} como:

$$\mathbf{A} = (\mathbf{I}_{MK} - \mathbf{D} (\mathbf{I}_K \otimes \mathbf{U}) (\mathbf{I}_K \otimes \mathbf{\Lambda}) (\mathbf{I}_K \otimes \mathbf{U}^H)) \mathbf{G}. \quad (4.36)$$

Nesse ponto, utilizaremos a seguinte propriedade relativa ao produto de Kronecker:

$$(\mathbf{M}_1 \otimes \mathbf{M}_2) (\mathbf{M}_3 \otimes \mathbf{M}_4) = \mathbf{M}_1 \mathbf{M}_3 \otimes \mathbf{M}_2 \mathbf{M}_4, \quad (4.37)$$

em que \mathbf{M}_1 , \mathbf{M}_2 , \mathbf{M}_3 e \mathbf{M}_4 são matrizes com dimensões que permitem os produtos $\mathbf{M}_1 \mathbf{M}_3$ e $\mathbf{M}_2 \mathbf{M}_4$. Aplicando a propriedade (4.37), e lembrando que $\mathbf{D} = \boldsymbol{\mu} \otimes \mathbf{I}_M$ e $\mathbf{G} = \mathbf{C} \otimes \mathbf{I}_M$, as seguintes igualdades são válidas:

$$\mathbf{D} (\mathbf{I}_K \otimes \mathbf{U}) = (\mathbf{I}_K \otimes \mathbf{U}) \mathbf{D}, \quad (4.38)$$

$$(\mathbf{I}_K \otimes \mathbf{U}^H) \mathbf{G} = \mathbf{G} (\mathbf{I}_K \otimes \mathbf{U}^H). \quad (4.39)$$

Portanto, a matriz \mathbf{A} pode ser reescrita como:

$$\mathbf{A} = (\mathbf{I}_K \otimes \mathbf{U}) \mathbf{E} \mathbf{G} (\mathbf{I}_K \otimes \mathbf{U})^H, \quad (4.40)$$

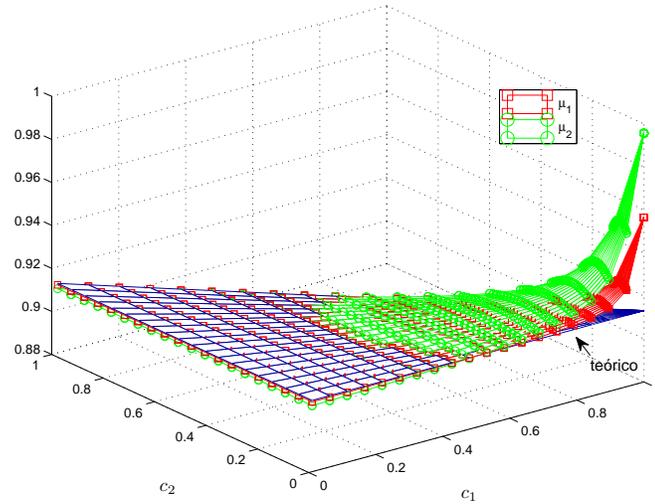


Figura 4.5: Raio espectral para as configurações de passo μ_1 e μ_2 para o caso de entrada colorida.

em que \mathbf{E} é uma matriz diagonal dada por:

$$\mathbf{E} = \mathbf{I}_{KM} - \mathbf{D}(\mathbf{I}_K \otimes \mathbf{\Lambda}). \quad (4.41)$$

Pode-se notar que os vetores da seguinte forma:

$$\mathbf{v} = [\mathbf{0}_{k-1} \ 1 - \mu_1 \lambda_k \ \mathbf{0}_{M-k} \ \cdots \ \mathbf{0}_{k-1} \ 1 - \mu_K \lambda_k \ \mathbf{0}_{M-k}]^T \quad (4.42)$$

são autovetores associados à matriz $\mathbf{E}\mathbf{G}$ e que seus respectivos autovalores são dados por $1 - \lambda_k \frac{1}{K} \sum_{i=1}^K \mu_i$. Os demais $(K - 1)M$ autovalores de $\mathbf{E}\mathbf{G}$ são nulos, uma vez que $\mathbf{C} = \frac{1}{K} \mathbf{1}\mathbf{1}^T$ apresenta $K - 1$ autovalores nulos. Portanto, o raio espectral da matriz \mathbf{A} é de fato dado por (4.35).

A eq. (4.35) mostra que, de modo análogo ao caso de entrada branca, a taxa de convergência permanece a mesma se a média dos passos de adaptação de cada filtro da rede permanecer constante. Para ilustrar esse resultado, considere um exemplo em que a primeira linha da matriz de autocorrelação é dada por $\mathbf{r} = [1 \ 0,2 \ 0,4, 0,1]$.

Assim como no caso de entrada branca, iremos analisar uma rede de três nós e matriz de combinação dada por (4.20). A Fig. 4.5 mostra o raio espectral em função dos parâmetros da matriz de combinação. Pode-se notar que a expressão teórica em (4.35) é uma boa aproximação para várias escolhas da matriz de combinação.

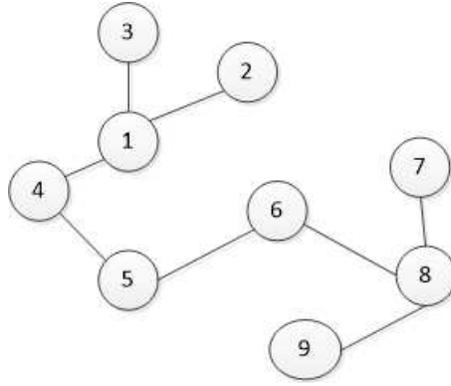
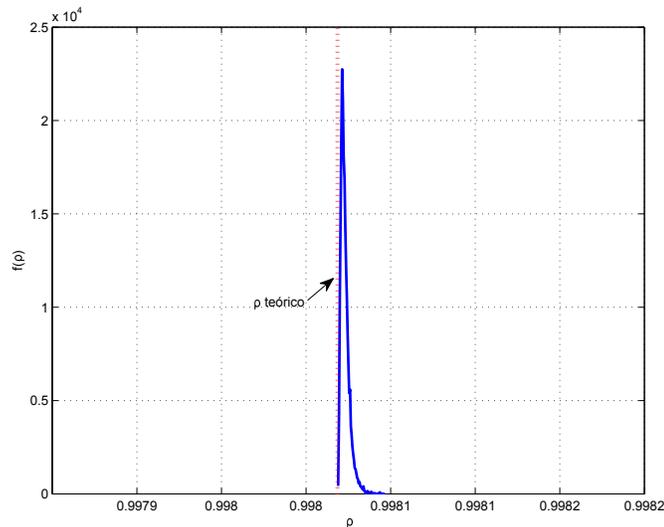


Figura 4.6: Rede de difusão

Figura 4.7: Histograma do raio espectral da matriz \mathbf{A} para configurações de passo sorteadas a partir de uma distribuição uniforme com média normalizada $\bar{\mu} = 0,01$

É ainda interessante analisar cenários mais realistas em que redes de sensores com mais nós são empregadas. Para tanto, o raio espectral da matriz \mathbf{A} associado à rede mostrada na Fig. 4.6 é calculado para diferentes configurações de passo de adaptação. Cada configuração de passo é sorteada a partir de uma distribuição de probabilidade uniforme e normalizada para que a média dos passos seja constante. Considerou-se que o vetor regressor é colorido com função de autocorrelação dada por $r(n) = 0,7^{|n|}$. O histograma dos valores de raio espectral obtidos com esse cenário está mostrado na Fig. 4.7. A partir desse resultado, pode-se notar que o histograma do raio espectral é muito próximo de um impulso localizado no valor teórico calculado a partir da hipótese de rede completamente conectada.

Generalização para o algoritmo ATC

Cabe observar que todo o desenvolvimento apresentado nessa subseção também se aplica ao caso do algoritmo do tipo ATC. A inversão na ordem de realização do algoritmo, leva à seguinte equação de recursão da média:

$$\mathbb{E} \left\{ \tilde{\boldsymbol{\psi}}(n) \right\} = \mathbf{G} (\mathbf{I}_{MK} - \mathbf{D}\mathbf{R}_{\mathbf{u}}) \mathbb{E} \left\{ \tilde{\boldsymbol{\psi}}(n-1) \right\}. \quad (4.43)$$

Logo, a matriz de sistema passa ser dada por:

$$\mathbf{A}_{ATC} = \mathbf{G} (\mathbf{I}_{MK} - \mathbf{D}\mathbf{R}_{\mathbf{u}}). \quad (4.44)$$

Reescrevendo as matrizes $\mathbf{R}_{\mathbf{u}}$ e \mathbf{D} e \mathbf{G} em função das matrizes \mathbf{R} e $\boldsymbol{\mu}$ e \mathbf{C} , respectivamente, temos que:

$$\mathbf{A}_{ATC} = \mathbf{C} (\mathbf{I} - \boldsymbol{\mu}) \otimes \mathbf{R}. \quad (4.45)$$

Utilizando a decomposição dada em (4.31) e tomando apenas o termo dominante, concluímos que a matriz \mathbf{A}_{ATC} pode ser aproximada por:

$$\mathbf{A}_{ATC} \approx \mathbf{C} \left(1 - \frac{1}{K} \sum_{k=1}^K \mu_k \right) \otimes \mathbf{R}, \quad (4.46)$$

de onde concluímos que a aproximação da matriz de sistema \mathbf{A}_{ATC} é a mesma a aproximação da matriz \mathbf{A} . Portanto, o comportamento da média do vetor de erro de estimação é o mesmo para ambos os algoritmos e, como consequência, a velocidade de convergência também é a mesma no caso dos algoritmos ATC e CTA.

4.2.2 Configuração ótima de passo de adaptação

O desempenho do algoritmo adaptativo é frequentemente avaliado em termos do MSD (do inglês, *Mean Square Deviation*), definido como:

$$\text{MSD} = \frac{1}{K} \sum_{k=1}^K E \left\{ \left\| \tilde{\boldsymbol{\psi}}_k \right\|^2 \right\}. \quad (4.47)$$

Na Seção anterior, foi mostrado que se a média dos passos for constante, a taxa de convergência é também constante. Então, cabe perguntar, dada tal restrição, qual configuração de passos minimiza o MSD. É a resposta a tal pergunta que será

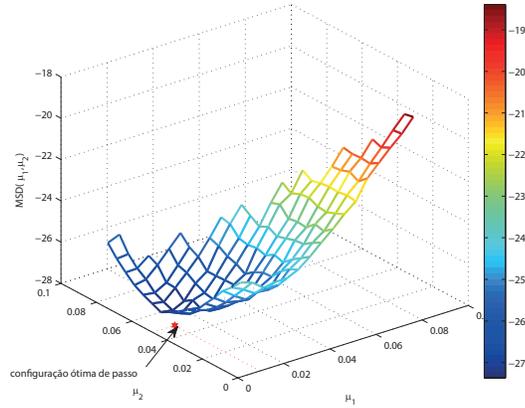


Figura 4.8: MSD para diferentes configurações de passo de adaptação

mostrada nesta Seção.

Para analisar essa questão, será considerada uma rede com três nós, assim como feito na Seção anterior. Como parâmetros de simulação, a potência do vetor de entrada é mantida constante para todos os nós e a potência de ruído de medida em cada nó é dada por: $\sigma_{\eta}^2 = [0,2 \ 0,04 \ 0,06]$. Fixando o passo médio em $\bar{\mu} = 0,03$, variou-se os passos de adaptação para os nós 1 e 2. Os valores de MSD em regime obtidos para cada configuração de passo estão indicados na Fig. 4.8. A partir dela, pode-se notar que o valor mínimo de MSD é alcançado para a configuração de passo $\mu_1 \approx 0,01$, $\mu_2 \approx 0,05$ e, conseqüentemente, $\mu_3 \approx 0,03$. Em função desse resultado, conjectura-se que a configuração ótima de passos de adaptação deve obedecer à seguinte condição:

$$\sigma_{\eta,1}^2 \mu_1 = \sigma_{\eta,2}^2 \mu_2 = \dots = \sigma_{\eta,K}^2 \mu_K. \quad (4.48)$$

Com o objetivo de verificar se esse fato ocorre em qualquer contexto, determinaremos qual a distribuição ótima de passo de adaptação para uma dada velocidade de convergência fixada. Para determinarmos a condição ótima de passo, no Lema 4.2.1 propomos uma expressão simplificada para o MSD da rede adaptativa.

Lema 4.2.1 (Expressão aproximada para o MSD). *Os MSD em regime associados aos algoritmos LMS-CTA e LMS-ATC, denotados respectivamente por \mathcal{M}_{CTA} e \mathcal{M}_{ATC} de uma rede adaptativa com K nós e vetor regressor de comprimento M é*

aproximado pela seguinte expressão:

$$\mathcal{M}_{CTA} = \frac{1}{K} \sum_{n=1}^K \sum_{m=1}^M \frac{\lambda_u(m)}{1 - (\lambda_c(n)(1 - \bar{\mu}\lambda_u(m)))^2} \sum_{k=1}^K \mu_k^2 \sigma_{\eta,k}^2, \quad (4.49)$$

$$\mathcal{M}_{ATC} = \frac{1}{K} \sum_{n=1}^K \sum_{m=1}^M \frac{\lambda_c^2(n)\lambda_u(m)}{1 - (\lambda_c(n)(1 - \bar{\mu}\lambda_u(m)))^2} \sum_{k=1}^K \mu_k^2 \sigma_{\eta,k}^2. \quad (4.50)$$

Demonstração. Ver apêndice B. \square

Lembrando que os autovalores da matriz \mathbf{C} são tais que $\lambda_c(n) \leq 1, \forall n$, a análise de (4.49) e (4.50) leva à conclusão de que $\mathcal{M}_{ATC} \leq \mathcal{M}_{CTA}$, como já observado em [59].

Para ilustrar a validade dessa aproximação, é interessante observar as Figs. 4.10 e 4.11. Esses resultados mostram a aproximação de MSD comparada com a solução real da equação de Lyapunov tanto para a versão CTA quanto para a versão ATC. O MSD foi calculado considerando que os passos de adaptação e as potências do ruído são aqueles mostrados na Fig. 4.9 e a matriz de combinação é a indicada em (4.34). A influência do tamanho da rede é avaliada na Fig. 4.10, enquanto que a influência da magnitude dos passos de adaptação está mostrada na Fig.4.9. Por esses resultados, observa-se que a aproximação é melhor para redes com mais nós e para passos de adaptação de pequena magnitude. Além disso, nota-se que a diferença de desempenho entre os algoritmo ATC e CTA diminui para passos pequenos e para redes com menos nós.

A fim de compreender melhor como o algoritmo ATC se relaciona com o CTA, o MSD de cada algoritmo foi comparado para diversas configurações de passo de adaptação médio e valor de α para a matriz de conectividade mostrada em (4.34). Os resultados relativos a redes com $K = 12$ e $K = 36$ nós estão mostrados nas Figs. 4.12 e 4.13, respectivamente. A análise dos resultados mostra que, em ambos os casos, se o passo médio for pequeno, as duas versões do LMS distribuído são equivalentes em termos de desempenho. Para o caso da rede com $K = 12$ nós, se o passo médio for cerca de $\bar{\mu} = 10^{-2}$, a diferença de desempenho entre eles é inferior a 1 dB, independentemente do grau de conectividade da rede. Já para o caso da rede com $K = 36$ nós, notamos que o passo médio precisa ser inferior a $\bar{\mu} = 5 \cdot 10^{-2}$ para

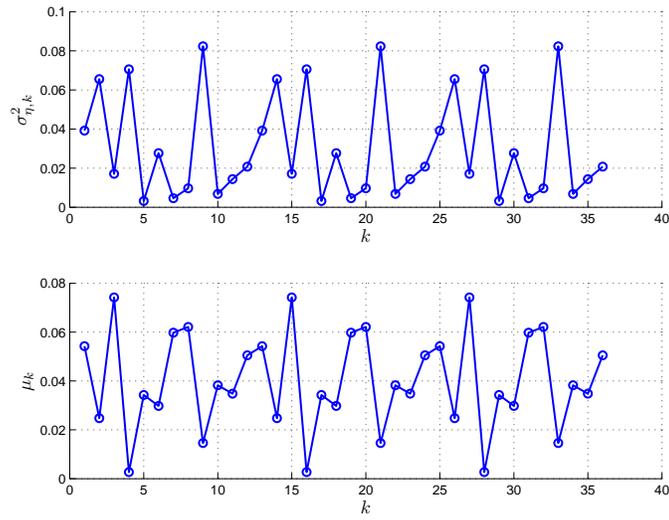


Figura 4.9: Potência do ruído e passo de adaptação em cada nó da rede.

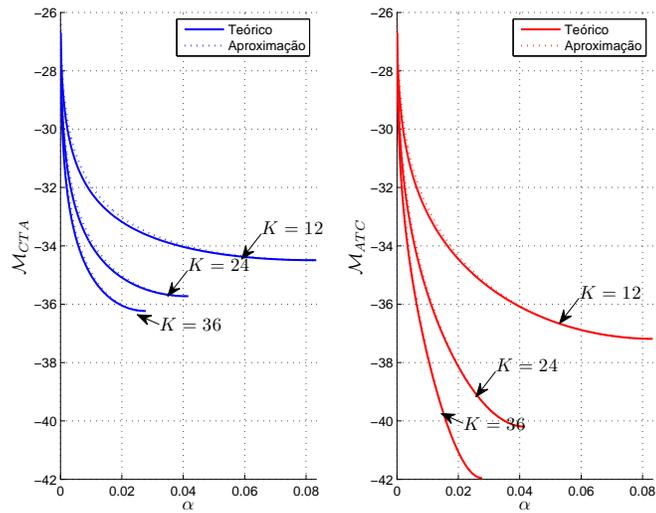


Figura 4.10: Comparação entre o valor teórico de MSD e o valor dado pela aproximação (4.49) para redes com diferentes número de nós. O passo de adaptação e a potência de ruído em cada nó estão mostrados na Fig.4.9.

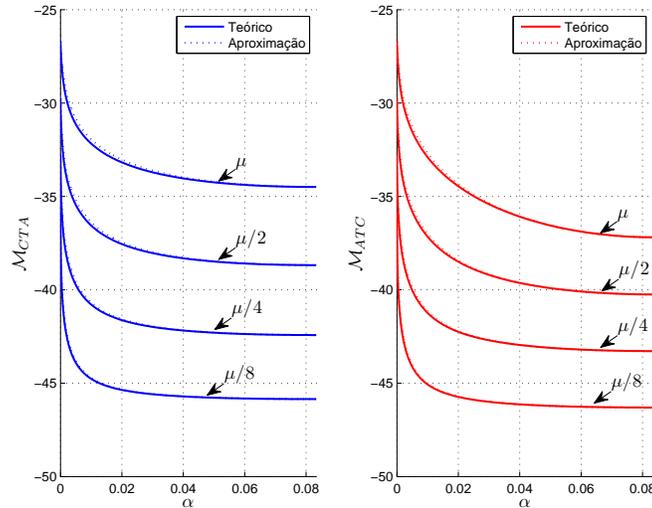


Figura 4.11: Comparação entre o valor teórico de MSD e o valor dado pela aproximação (4.49) para redes com $K = 12$ nós e valores de passos de adaptação diferentes. O passo de adaptação e a potência de ruído em cada nó estão mostrados na Fig.4.9.

que os dois algoritmos apresentem o mesmo desempenho para qualquer que seja o valor de α . Em ambos os casos, nota-se que quanto maior o grau de conectividade da rede (valor de α) maior a diferença de desempenho entre os dois algoritmos.

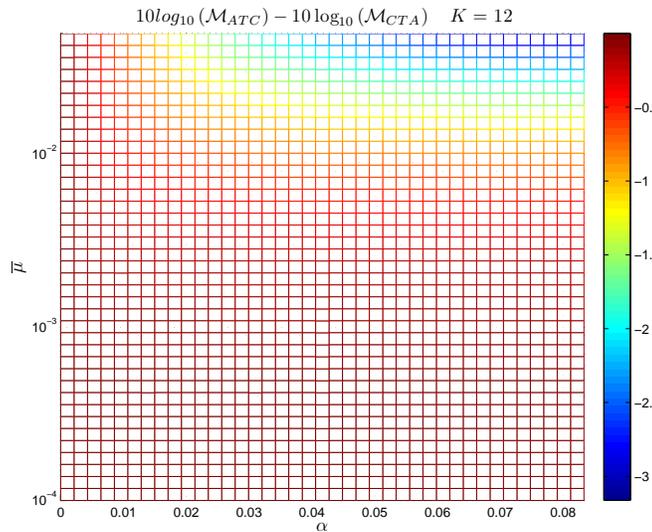


Figura 4.12: Comparação de desempenho entre as versões CTA e ATC do algoritmo LMS, considerando redes com $K = 12$ nós, vários valores de passo médio e matrizes de conectividade dadas por (4.34).

A partir das expressões para MSD fornecidas por esse lema, podemos encontrar a distribuição ótima para os passos de adaptação em cada nó da rede como descrita

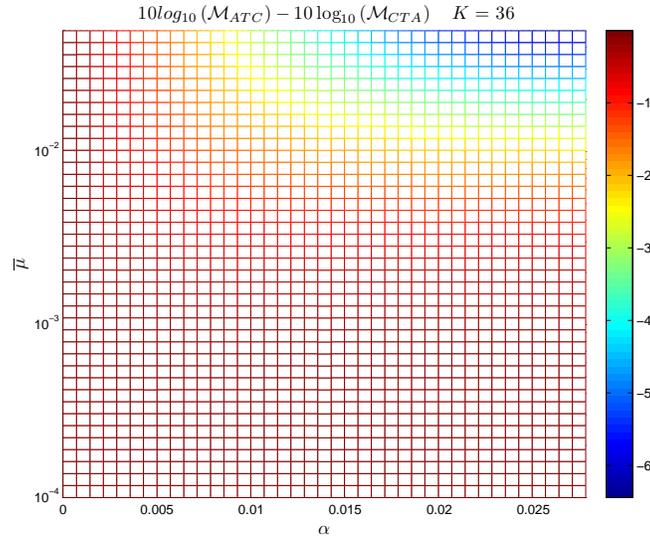


Figura 4.13: Comparação de desempenho entre as versões CTA e ATC do algoritmo LMS, considerando redes com $K = 36$ nós, vários valores de passo médio e matrizes de conectividade dadas por (4.34).

na seguinte Proposição:

Proposition 4.2.2 (Distribuição ótima de passo de adaptação). *Para um dado passo de adaptação médio $\bar{\mu}$, o passo de adaptação em cada nó deve ser dado por:*

$$\mu_i = \frac{\bar{\mu}}{\sigma_{\eta,i}^2} \text{harm}(\boldsymbol{\sigma}_{\eta}^2), \quad (4.51)$$

em que $\text{harm}(\boldsymbol{\sigma}_{\eta}^2)$ representa a média harmônica do vetor de potência de ruído:

$$\text{harm}(\boldsymbol{\sigma}_{\eta}^2) = \frac{1}{\frac{1}{K} \sum_{k=1}^K \frac{1}{\sigma_{\eta,k}^2}}. \quad (4.52)$$

Demonstração. Eliminando os termos constantes em (4.49) e (4.50) relativos aos passos de adaptação, a configuração ótima de passo de adaptação será dada pela solução do seguinte problema de otimização:

$$\begin{aligned} & \underset{\boldsymbol{\mu}=[\mu_1 \ \mu_1 \ \dots \ \mu_K]}{\text{minimizar}} && \sum_{k=1}^K \mu_k^2 \sigma_{\eta,k}^2 \\ & \text{sujeito a} && \frac{1}{K} \sum_{k=1}^K \mu_k = \bar{\mu} \end{aligned}$$

O lagrangiano associado a esse problema é dado por:

$$L(\boldsymbol{\mu}, l) = \sum_{k=1}^K \mu_k^2 \sigma_{\eta,k}^2 + l \left(\frac{1}{K} \sum_{i=1}^K \mu_i - \bar{\mu} \right). \quad (4.53)$$

Igualando o gradiente do lagrangiano a zero, temos que a seguinte relação precisa ser obedecida:

$$\frac{\partial}{\partial \mu_k} L(\boldsymbol{\mu}, l) = 0 \rightarrow \mu_k = -\frac{l}{2K\sigma_k^2}. \quad (4.54)$$

Aplicando (4.54) na equação de restrição, obtém-se (4.51).

□

É interessante e importante observar que esse é um resultado que não depende da matriz de correlação do sinal de entrada. Esse problema de otimização é o mesmo resolvido em [62], no qual foi considerada uma rede LMS do tipo incremental com entrada branca.

4.2.3 Estimação do passo de adaptação ótimo

Na Seção anterior, a configuração ótima de passo do adaptação foi determinada para uma dada configuração da potência do ruído nos nós da rede. Analisando (4.54), entretanto, observa-se que a configuração ótima de passo de adaptação depende do conhecimento da potência do ruído em todos os nós da rede. Com o objetivo de superar essa limitação, é proposta uma técnica para esta finalidade, cujo princípio é que cada nó difunda também a sua própria estimativa de potência de ruído.

A potência do ruído pode ser estimada diretamente a partir do erro quadrático médio (MSE) em cada um dos nós. A variância da estimativa da potência do ruído pode ser deduzida através da aplicação de um filtro de média móvel na estimativa da MSE:

$$\hat{\sigma}_{\eta,k}^2(n) = \frac{1}{N} \sum_{r=0}^{N-1} |e_k(n-r)|^2, \quad (4.55)$$

em que N é a ordem do filtro.

A estimativa da potência do ruído no nó k , então, é passada para os nós em sua vizinhança. Se a rede não for completamente conectada, será necessário mais de uma iteração até que a estimativa seja passada para todos os nós da rede, sendo

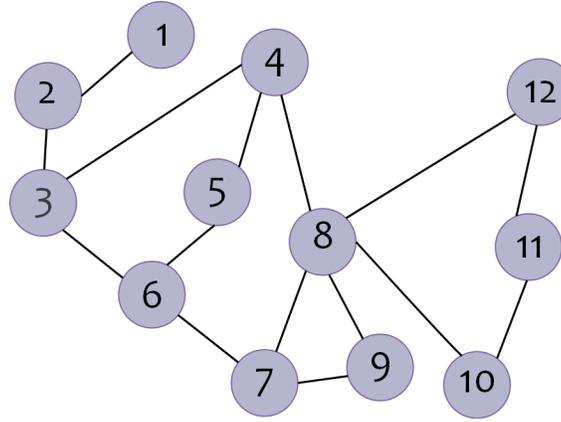


Figura 4.14: Rede de difusão

que, no pior caso, seriam necessárias $K - 1$ iterações. Por essa razão, a configuração dos passos de adaptação em cada nó é atualizada apenas a cada $K - 1$ iterações.

A partir da estimativa da potência do ruído em cada nó dada em (4.55), o passo de adaptação em cada um dos nós é dado por:

$$\widehat{\text{harm}}(\sigma_\eta^2) = \frac{1}{\frac{1}{K} \sum_{k=0}^{K-1} \frac{1}{\hat{\sigma}_{\eta,k}^2(n)}}, \quad (4.56)$$

$$\mu_k(r : r + K - 2) = \min \left\{ \mu_{\max}, \bar{\mu} \frac{\widehat{\text{harm}}(\sigma_\eta^2)}{\hat{\sigma}_{\eta,k}^2(n)} \right\}, \quad (4.57)$$

em que $\mu_k(r : r + K - 2)$ representa o passo de adaptação relacionado ao nó k da iteração r a $r + K - 2$ e μ_{\max} é o passo máximo que deve ser atribuído a cada nó a fim de evitar divergência do algoritmo.

O sistema de atribuição de passos adaptativos é testado para a rede mostrada na Fig. 4.14. A potência de ruído, assim como o passo ótimo relativo a cada nó da rede, estão mostrados na Fig. 4.15. Na Fig. 4.16 está apresentada a estimativa da média harmônica das potências de ruído dos nós da rede. De fato, pode-se observar que a estimativa converge para o valor real representado na linha pontilhada na Fig. 4.16. Na Fig. 4.17, é mostrado o passo de adaptação para cada um dos nós na rede. A partir desse resultado, é possível concluir que os passos de adaptação calculados a partir das estimativas da potência de ruído convergem para o valor ótimo, como esperado.

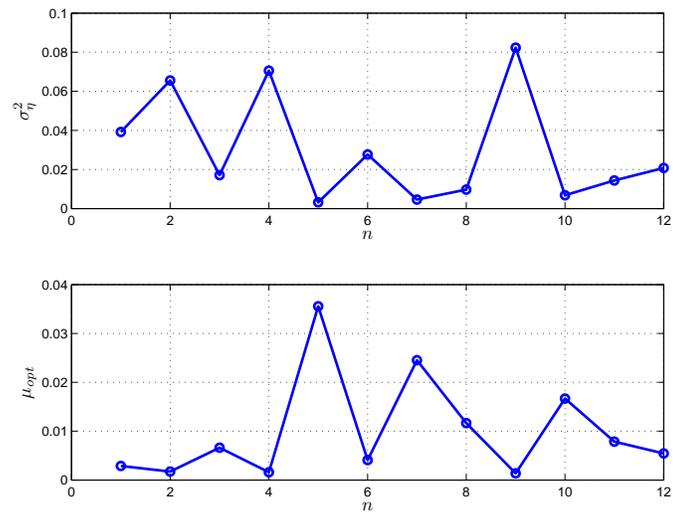


Figura 4.15: Potência de ruído e passo de adaptação ótimo associado a cada nó da rede mostrada na Fig. 4.14.

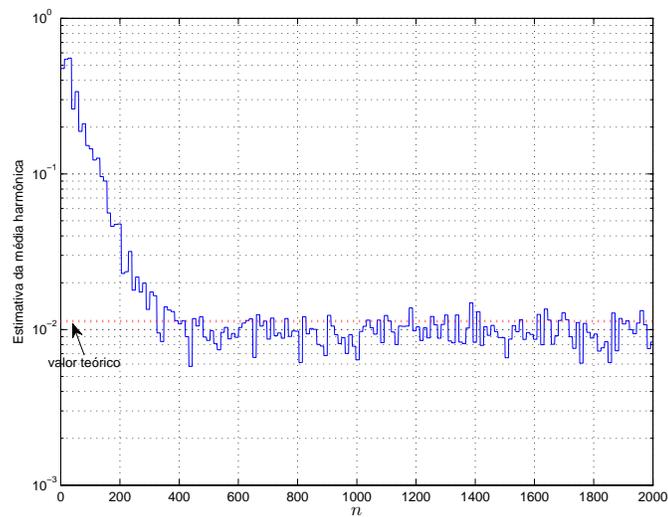


Figura 4.16: Média harmônica estimada.

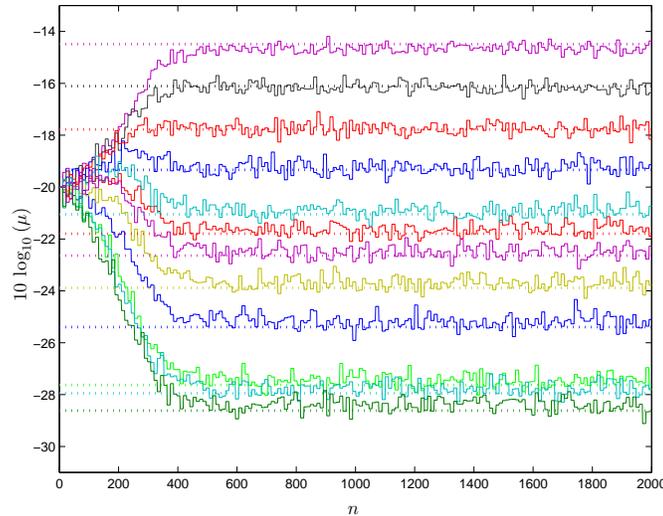


Figura 4.17: Passo adaptado considerando a rede mostrada na Fig. 4.14 com potência de ruído mostrada na Fig. 4.15.

4.2.4 Resultados de simulação

Nessa Seção, nós corroboraremos os resultados teóricos apresentados nas seções anteriores com alguns resultados de simulação.

Primeiramente, será considerada a rede de $K = 12$ nós mostrada na Fig. 4.14. No primeiro cenário de simulação, é considerado que a entrada é branca de potência unitária. O vetor a ser estimado apresenta $M = 5$ coeficientes iguais e potência unitária, *i.e.*, $\mathbf{w}_{opt} = [1, 1, \dots, 1] / \sqrt{M}$.

A potência de ruído em cada um dos nós está indicada na Fig. 4.15. O passo de adaptação médio foi fixado em $\bar{\mu} = 0,01$ e cada passo individual ótimo foi calculado de acordo com (4.51) e mostrado na Fig. 4.15.

Para esse valor de passo de adaptação médio, o desempenho das versões ATC e CTA do algoritmo é aproximadamente o mesmo como mostrado na Fig. 4.18. Esse resultado mostra que, para esse cenário de conectividade e passo de adaptação médio, não há diferença de desempenho significativo entre as duas versões do algoritmo. Assim, usaremos apenas o ATC nas demais simulações, mas pode-se esperar que resultados similares sejam encontrados para o algoritmo CTA.

A Fig. 4.19 mostra o MSD para esse cenário de simulação. O MSD foi estimado a partir da média de 100 realizações independentes de medidas instantâneas de MSD,

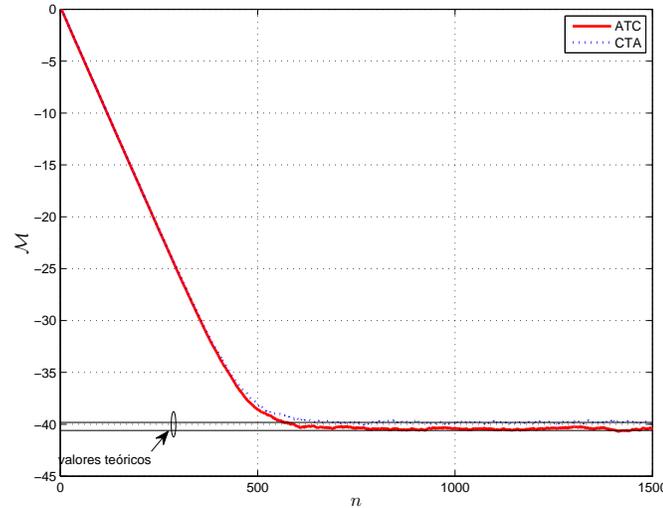


Figura 4.18: Comparação de desempenho entre os algoritmos ATC e CTA para a rede mostrada na Fig. 4.14 e configurações de passo de adaptação e potência de ruído dadas na Fig. 4.15.

i. e.:

$$\text{MSD}(n) = \frac{1}{K} \sum_{k=1}^K \|\psi_k(n) - \mathbf{w}_{opt}\|^2. \quad (4.58)$$

Cada uma das várias curvas de MSD apresentadas na Fig. 4.19 corresponde a uma configuração de passo de adaptação diferente, sorteada a partir de uma distribuição uniforme e normalizada para garantir um passo de adaptação médio de $\bar{\mu} = 0,01$. O primeiro ponto importante a ser observado é que a taxa de convergência, de fato, se mantém constante, mesmo que a rede não seja completamente conectada. Pode-se também notar que a configuração de passo de adaptação que leva ao menor valor de MSD é aquela mostrada na Fig. 4.15. Além disso, é importante enfatizar que o valor mínimo de MSD dado por (4.50) é uma boa aproximação para o valor mínimo de MSD atingido tanto para o caso em que a rede é completamente conectada quanto para o caso da rede apresentada na Fig. 4.14. Um outro fator importante a ser analisado é que o sistema com passo adaptativo mostrado na Seção 4.2.3 apresenta desempenho muito próximo ao encontrado com a configuração ótima do passo de adaptação.

Além do caso com entrada branca, também foram feitas simulações para o caso com entrada colorida. Nesse cenário, considerou-se que o vetor regressor de entrada

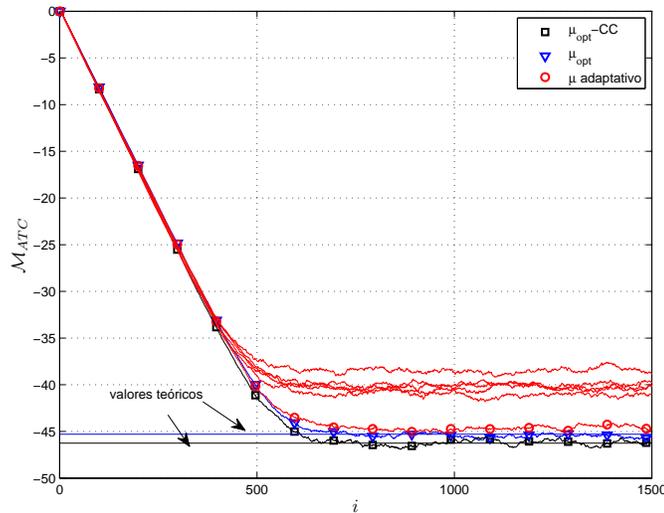


Figura 4.19: MSD para diferentes configurações de passo fixando $\bar{\mu} = 0,01$, entrada branca e rede mostrada na Fig. 4.14.

u_k foi modelado como um processo autoregressivo com função de autocorrelação dada por $r(n) = 0,7^{|n|}$.

Os resultados estão mostrados na Fig. 4.20, a partir dos quais pode-se notar que a taxa de convergência é também constante para o caso de entrada colorida, mesmo que a rede não seja completamente conectada. Além disso, nota-se que o valor mínimo de MSD é alcançado quando o valor do passo de adaptação é também escolhido de acordo com (4.51).

A seguir, será analisado o impacto do número de nós na rede a partir das simulações de uma rede com 24 nós mostrada na Fig.4.21. A potência de ruído para cada nó é apresentada na Fig.4.22 que também mostra os passos de adaptação ótimos dados por (4.51) para cada nó da rede.

Os resultados mostrados na Fig. 4.23 são relativos ao caso em que o vetor regressor é branco. Analisando os resultados, nota-se que a velocidade de convergência do algoritmo é aproximadamente a mesma, dado que a média dos passos é constante. Observa-se também a validade da expressão dada por (4.50) para a aproximação da MSD em regime. Note que a degradação em termos de MSD, no caso em que a rede é a apresentada na Fig. 4.21 em relação ao caso completamente conectado, é maior do que no caso da rede dada na Fig. 4.14. Conclusões semelhantes podem

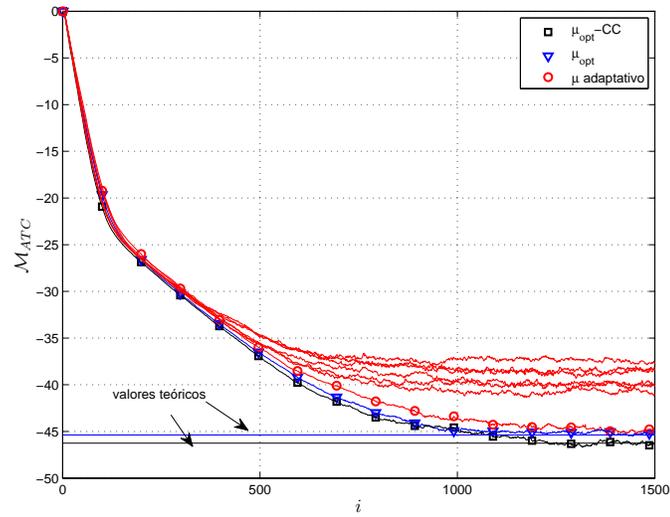


Figura 4.20: MSD para diferentes configurações de passo fixando $\bar{\mu} = 0,01$, entrada colorida e rede mostrada na Fig. 4.14.

ser extraídas da análise dos resultados relativos à entrada colorida e mostrados na Fig. 4.24.

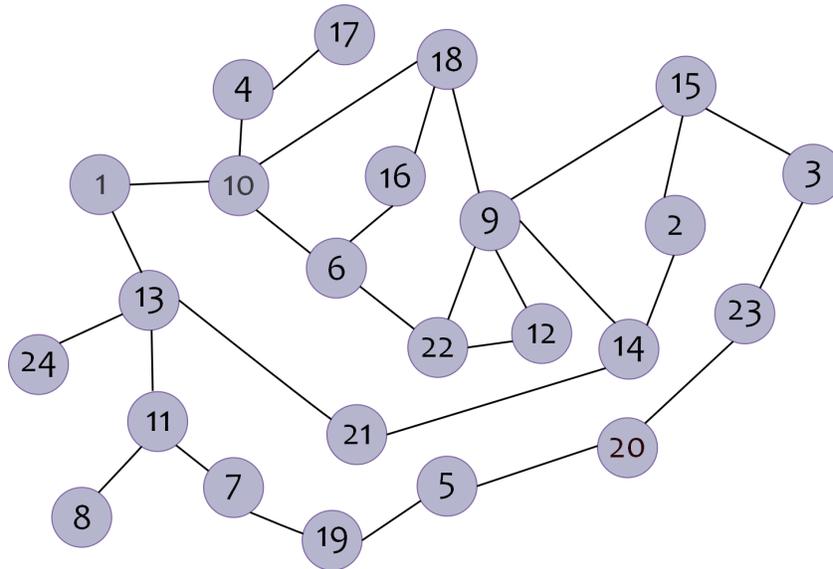


Figura 4.21: Rede de difusão com 24 nós.

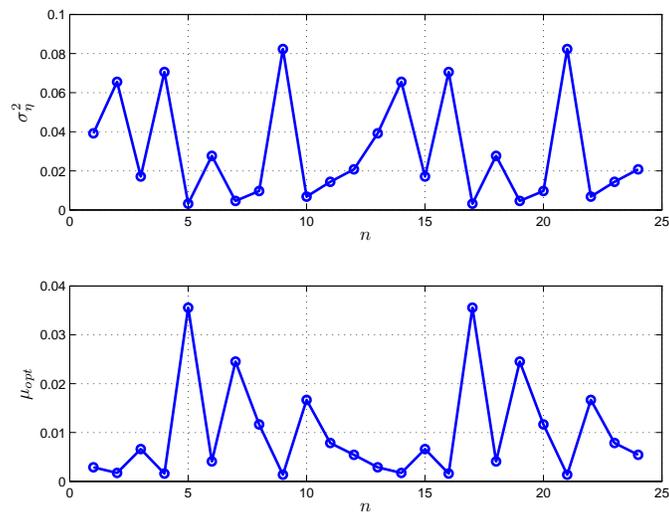


Figura 4.22: Potência de ruído e passo de adaptação ótimo associado a cada nó da rede mostrada na Fig. 4.21.

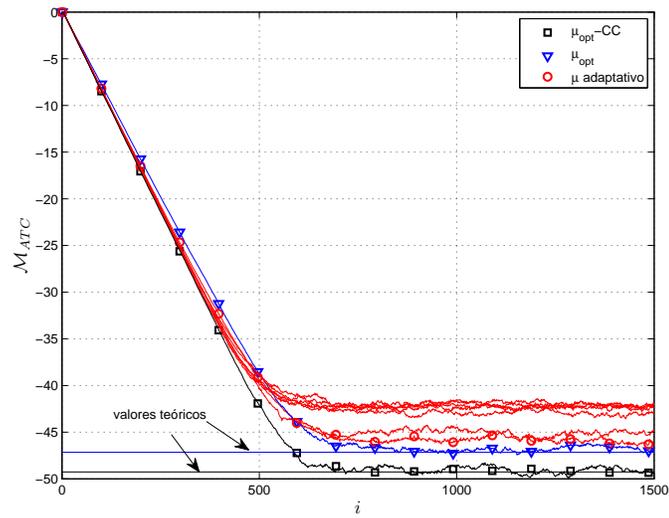


Figura 4.23: MSD para diferentes configurações de passo fixando $\bar{\mu} = 0,01$, entrada branca e rede mostrada na Fig. 4.21.

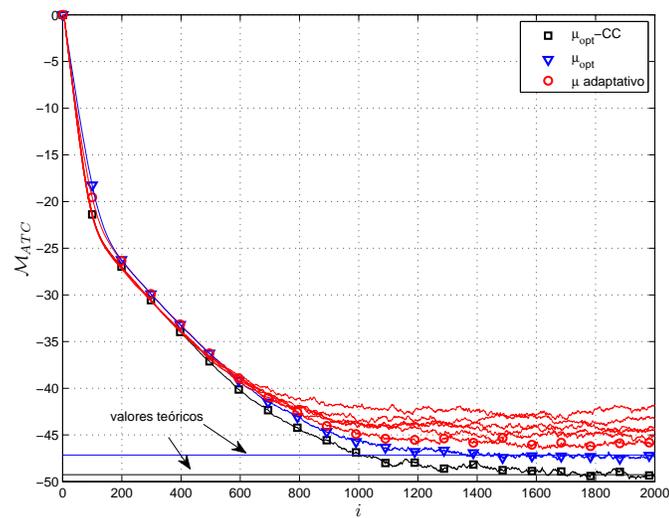


Figura 4.24: MSD para diferentes configurações de passo fixando $\bar{\mu} = 0,01$, entrada colorida e rede mostrada na Fig. 4.21.

4.3 *Set-membership* distribuído

Na Seção anterior, apresentamos como determinar adaptativamente os parâmetros ótimos de um algoritmo do tipo LMS distribuído em situações nas quais cada nó da rede encontra-se em diferentes condições de relação sinal-ruído. Tal estimação, acarreta um aumento da complexidade computacional. Nesta Seção, mostraremos que uma outra classe de algoritmos naturalmente combate a influência de nós com alta potência de ruído na rede sem apresentar um acréscimo no custo computacional. Trata-se dos algoritmos do tipo *set-membership* (SM) [63].

A característica principal dos algoritmos do tipo SM diz respeito ao fato de que a solução fornecida pelo o algoritmo não ser única. Todos os elementos do chamado conjunto viável (do inglês, *feasible set*) é admitida como uma possível solução do algoritmo. Dessa forma, a taxa de atualização do algoritmo é menor do que a de um algoritmo adaptativo clássico. Essa é uma estratégia especialmente interessante no caso em que o consumo de energia é um parâmetro crítico do sistema, assim como ocorre no caso em redes de sensores.

Aplicando a estratégia SM no processamento local do sistema, a adaptação ocorre sempre que o erro de estimação no nó for maior que um determinado limiar, $\bar{\gamma}$. Dessa forma, pode-se analisar o algoritmo SM como um algoritmo de passo variável:

$$\mu_k(n) = \begin{cases} 0, & |e_k(n)| > \bar{\gamma}_k \\ \mu_0 \left(1 - \frac{\bar{\gamma}}{|e_k(n)|}\right), & c.c. \end{cases}, \quad (4.59)$$

em que $\bar{\gamma}$ é uma constante que define o valor de erro mínimo necessário para a adaptação do algoritmo. Em geral, a constante $\bar{\gamma}$ é expressa na forma $\bar{\gamma} = \sqrt{\tau \sigma_{\eta,k}^2}$, em que τ é uma constante positiva. Se a constante $\bar{\gamma}$ for definida dessa forma, a probabilidade de atualização do algoritmo é independente da potência de ruído. No contexto de estimação distribuída em que os nós das redes apresentam diferentes potências de ruído, é interessante que os nós mais ruidosos apresentem uma menor probabilidade de atualização. A seguir, será mostrado que o algoritmo apresenta essa característica se a constante $\bar{\gamma}$ for definida como $\bar{\gamma} = \tau_k \sigma_{\eta,k}^2$.

Assim, note que o algoritmo será atualizado com a seguinte probabilidade no

instante n e no nó k [64]:

$$P_{up}(k, n) = Pr \{|e_k(n)| > \bar{\gamma}_k\}. \quad (4.60)$$

Como, nesta Seção, estamos assumindo que não há qualquer conhecimento *a priori* das potências de ruído nos demais nós da rede, iremos assumir que as constantes τ_k são as mesmas para todos os nós da rede *i.e.*, $\bar{\tau}_k = \tau, \forall k$.

Dessa forma, assumindo que o erro de estimação em cada nó apresenta distribuição de probabilidade aproximadamente gaussiana, as probabilidades de atualização serão dadas por:

$$P_{up}(k, n) = Pr \{|e_k(n)| > \bar{\gamma}_k\} = 2Pr \{e_k(n) > \tau\sigma_{\eta,k}^2\} = 2Q\left(\frac{\tau\sigma_{\eta,k}^2}{\sigma_{e(n),k}}\right), \quad (4.61)$$

em que $\sigma_{e(n),k}^2$ corresponde à variância do erro no nó k e no instante n e $Q(x)$ é a função erro definida como:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{t^2}{2}\right) dt. \quad (4.62)$$

Em regime, podemos aproximar $\sigma_{e(n),k}^2$ por $\sigma_{\eta,k}^2$, a seguinte expressão aproximada para a probabilidade de atualização é obtida:

$$P_{up}(k) \approx 2Q(\tau\sigma_{\eta,k}). \quad (4.63)$$

Como a função $Q(x)$ é monotonicamente decrescente, os nós sujeitos à maior potência de ruído apresentarão uma taxa de atualização mais baixa. Portanto, em média, contribuirão menos para a estimativa geral da rede.

Como exemplo, consideremos mais uma vez a rede mostrada na Fig. 4.14. Analisaremos os algoritmos LMS e sua versão SM (SM-LMS) considerando que a potência em cada nó da rede é dada por:

$$\sigma_{\eta,k}^2 = \bar{\sigma}_\eta^2 \exp(-\alpha k) K \frac{1 - \exp(-\alpha)}{1 - \exp(-K\alpha)} \quad 0 \leq k \leq K - 1, \quad (4.64)$$

em que K é o número de nós da rede, α é uma constante real positiva e $\bar{\sigma}_\eta^2$ é a potência média de ruído nos nós da rede.

Note que se $\alpha = 0$, todos os nós da rede apresentam a mesma potência de ruído

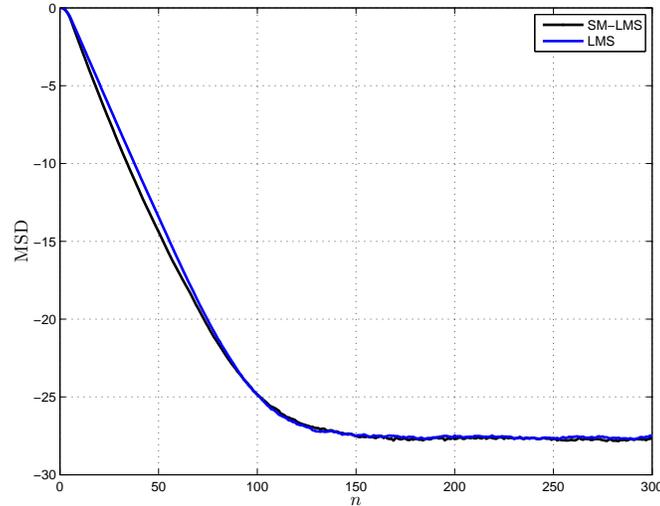


Figura 4.25: Comparação de desempenho entre os algoritmos LMS e SM-LMS, considerando que a potência de ruído em cada nó da rede é a mesma.

$\sigma_{\eta,k}^2 = \overline{\sigma_{\eta}^2} \forall k$. Quanto maior o valor de α , mais desigual é a distribuição de potência de ruído na rede.

O passo de adaptação do algoritmo LMS foi fixado em $\mu = 0,034$ para todos os nós da rede. Já para o algoritmo SM-LMS, o parâmetro μ_0 foi fixado em $\mu_0 = 0,05$ e o parâmetro τ foi fixado em $\tau_k = \tau = 2$. Esses parâmetros foram escolhidos de modo que o MSD em regime fosse o mesmo para os dois algoritmos, para o caso em que a potência de ruído é a mesma em todos os nós da rede, *i.e.*, $\alpha = 0$. Os resultados estão mostrados na Fig. 4.25, onde se pode notar que o desempenho dos dois algoritmos é praticamente equivalente.

No caso em que os nós apresentam diferentes potências de ruído, o algoritmo SM-LMS mostra uma clara vantagem em relação ao LMS como indicado nos resultados apresentados na Fig 4.26. O ganho de desempenho é crescente em relação ao parâmetro α , o que mostra que o algoritmo SM-LMS realmente é uma boa alternativa ao LMS no contexto em que a distribuição de potência é desigual.

A Fig. 4.27 mostra a probabilidade de atualização em função da potência de ruído em cada nó da rede. Note que a estimativa de probabilidade de detecção obtida por simulação coincide com a aproximação dada por (4.63).

Além do ganho de desempenho, deve-se destacar a menor complexidade computacional envolvida na implementação do algoritmo SM-LMS em comparação com o

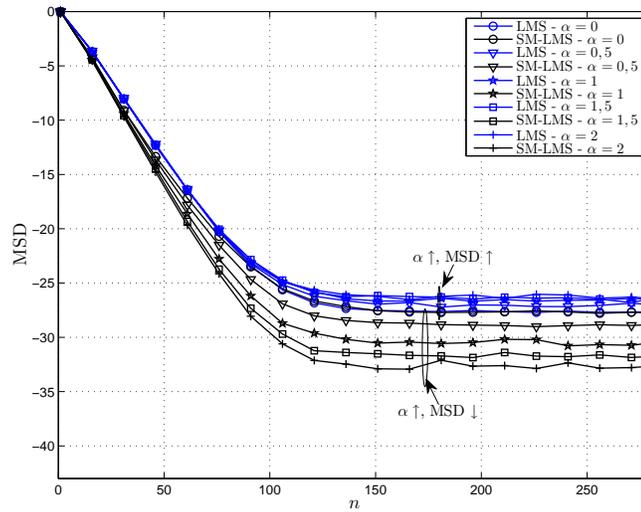


Figura 4.26: Comparação de desempenho entre os algoritmos LMS e SM-LMS, considerando que a potência de ruído em cada nó da rede é a dada por (4.64).

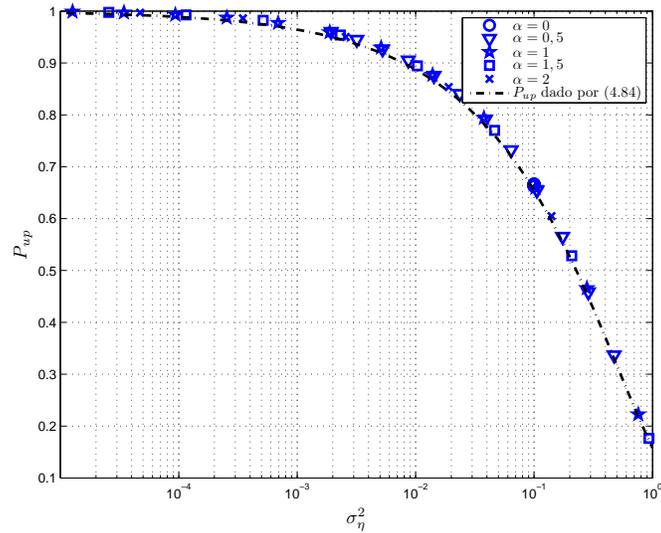


Figura 4.27: Probabilidade de atualização em função da potência de ruído considerando que a potência de ruído em cada nó da rede é a dada por (4.64).

LMS [55], já que o algoritmo atualiza apenas em algumas iterações.

4.3.1 *Set-membership* distribuído com redução de *feedback*

A ideia que fundamenta os algoritmos do tipo *set-membership* também pode ser aplicada a fim de reduzir a troca de informações na rede, de forma semelhante ao feito em [65] no algoritmo SM-LMS com *feedback* reduzido, o qual será denotado por RF-SM-LMS. Em [65], antes de enviar a estimativa aos nós vizinhos, cada nó realiza um teste de modo a avaliar se a estimativa será relevante para o seu vizinho. Para tanto, se define o conjunto:

$$\mathcal{H}_l(n) = \{ \boldsymbol{\phi} \mid |d_l(n) - \mathbf{u}_l(n)^T \boldsymbol{\phi}| < \gamma_{fb,l} \}. \quad (4.65)$$

em que $\gamma_{fb,l}$ é o limiar que determina se é necessário enviar a informação. Se $\boldsymbol{\phi}_k(n) \in \mathcal{H}_l(n)$, significa que a estimativa $\boldsymbol{\phi}_k(n)$ não é relevante para o nó l e, por essa razão, ela não deve ser transmitida ao nó l .

Observe que essa abordagem realmente diminui a taxa de ocupação de canal, todavia requer o conhecimento dos pares $\{\mathbf{u}_l(n), d_l(n)\} \forall l \in \mathcal{N}_k$. Neste trabalho, entretanto, assumimos que as únicas informações trocadas entre os nós são as estimativas. Por essa razão, se faz necessário definir um outro critério para decidir se a estimativa deve ser enviada aos nós vizinhos. Nesse contexto, definimos um critério local baseado apenas no par $\{\mathbf{u}_k(n), d_k(n)\}$. O nó k transmitirá sua estimativa ao vizinho se o erro de estimação for acima de um limiar local $\gamma_{fb,k}^{loc}$.

A Fig. 4.28 foi gerada considerando que a potência de ruído em cada nó da rede é dada por (4.64) com o parâmetro $\alpha = 1$. O passo de adaptação foi fixado em $\mu = 0,045$ para o algoritmo LMS e $\mu = 0,075$ para os algoritmos SM-LMS e RF-SM-LMS. O limiar local foi definido como $\gamma_{fb,k}^{loc} = \sqrt{\tau_{fb,k}^{loc} \sigma_\eta}$, com $\tau_{fb,k}^{loc} = 0,2$. A análise dos resultados mostra que o desempenho do algoritmo RF-SM-LMS é levemente inferior em relação ao do algoritmo SM-LMS. A grande vantagem do uso do algoritmo RF-SM-LMS é que houve troca de informações entre os nós da rede em apenas 72,7% das iterações. Todavia, é importante enfatizar que ambas as versões do tipo *set-membership* apresentam desempenho superior em relação ao LMS em

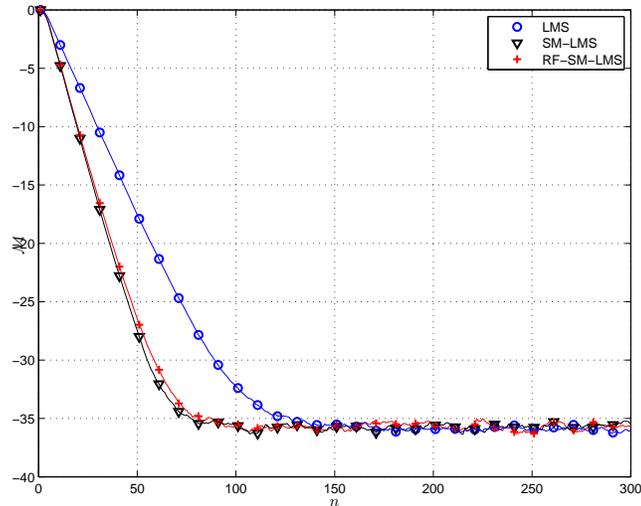


Figura 4.28: Comparação de desempenho entre os algoritmos LMS, SM-LMS e sua versão com *feedback* reduzido RF-SM-LMS.

termos de velocidade de convergência e custo computacional.

4.4 Conclusões

Neste capítulo, foram estudadas técnicas de estimação distribuída fundamentadas em algoritmos adaptativos clássicos. Focando no algoritmo LMS distribuído, foi mostrado que a taxa de convergência do algoritmo é determinada pela média aritmética dos passos de adaptação do algoritmo aplicado em cada nó da rede. Foi mostrado também, para uma taxa de convergência fixada, como determinar o passo de cada nó do algoritmo a fim de minimizar a MSD.

Também foi analisada a aplicação de algoritmos do tipo *set membership* no contexto distribuído e mostrou-se que essa abordagem é especialmente adequada no caso de redes heterogêneas como mostrado em [11].

Capítulo 5

Detecção Distribuída

No Capítulo anterior, abordou-se o problema de estimação distribuída aplicando algoritmos clássicos de filtragem adaptativa. No contexto de rádio cognitivo, mais especificamente no problema de sensoriamento espectral, algoritmos de detecção são mais relevantes do que algoritmos de estimação. Este Capítulo será dedicado a esse último problema, o qual será abordado em três cenários distintos. No primeiro caso, abordado na Seção 5.1, consideraremos que há conhecimento completo do sinal enviado pelo usuário primário e do canal entre o usuário primário e o usuário secundário. No segundo caso, discutido na Seção 5.2, consideraremos que há conhecimento apenas do sinal enviado pelo usuário primário. Finalmente, consideraremos o caso mais geral no qual o usuário secundário não dispõe de nenhuma informação do sinal do usuário primário.

5.1 Técnicas baseadas em filtro casado

Em [52], são aplicados algoritmos supervisionados distribuídos, mais especificamente, variantes do LMS distribuído, no contexto de detecção distribuída. Com o objetivo de aplicar tais técnicas de estimação por difusão no contexto de rádio cognitivo [66], considera-se o mesmo modelo já estudado na descrição do algoritmo LMS distribuído no Capítulo 4:

$$d_k(n) = \mathbf{u}_k(n)^T \mathbf{w}_{opt}(n) + \mathbf{z}_k(n), \quad (5.1)$$

em que $\mathbf{z}_k(n) \sim \mathcal{N}(0, \sigma_{z,k}^2)$ corresponde à parcela de ruído e $\mathbf{w}_{opt}(n)$ pode assumir um dos seguintes valores:

$$\mathbf{w}_{opt}(n) = \begin{cases} \mathbf{0}, & \mathcal{H}_0 \\ \mathbf{w}_s(n), & \mathcal{H}_1 \end{cases}. \quad (5.2)$$

Após a aplicação do algoritmo LMS distribuído, o vetor $\mathbf{w}_k(n)$ é estimado na n -ésima iteração pelo nó k e a seguinte estatística de teste é gerada [14]:

$$T = \text{Re} \{ \mathbf{w}_s^H \mathbf{w}_k(n) \}. \quad (5.3)$$

Com tal estatística, o teste de hipóteses é aplicado e, através do limiar λ , é tomada a decisão entre as hipóteses \mathcal{H}_0 e \mathcal{H}_1 .

Ainda em [66], a detecção distribuída é aplicada no contexto de rádio cognitivo. Para tanto, considerando que o sinal transmitido pelo usuário primário é dado por $s(n)$, o sinal recebido pelo k -ésimo usuário secundário pode ser modelado pelo seguinte teste de hipóteses:

$$\begin{cases} \mathcal{H}_0 : \mathbf{y}_k(n) = \boldsymbol{\eta}_k(n), \\ \mathcal{H}_1 : \mathbf{y}_k(n) = \mathbf{r}_k(n) + \boldsymbol{\eta}_k(n). \end{cases} \quad (5.4)$$

em que $\mathbf{r}_k(n)$ é escrito em termos da matriz de convolução do canal:

$$\mathbf{r}_k(n) = \mathbf{H}_k \mathbf{s}(n), \quad (5.5)$$

em que \mathbf{H}_k , para um canal de comprimento L , é dada por:

$$\mathbf{H}_k = \begin{bmatrix} h_k(0) & h_k(1) & \cdots & h_k(L-1) & \cdots & 0 \\ 0 & h_k(0) & \cdots & h_k(L-1) & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & h_k(0) & \cdots & h_k(L-1) \end{bmatrix}. \quad (5.6)$$

A estatística de teste é obtida através de detecção coerente em cada nó. Com o objetivo de tornar esse problema idêntico ao problema mostrado em (5.1) e (5.2),

[66] define a seguinte mudança de variáveis:

$$d_k(n) = \frac{\mathbf{s}^H(n) \mathbf{H}_k^H \mathbf{y}(n)}{\|\mathbf{H}_k \mathbf{s}(n)\|}, \quad (5.7)$$

$$\mathbf{u}_k(n) = \frac{\mathbf{s}^H(n) \mathbf{H}_k^H \mathbf{H}_k}{\|\mathbf{H}_k \mathbf{s}(n)\|}, \quad (5.8)$$

$$\mathbf{w}_{opt} = \mathbf{s}(n), \quad (5.9)$$

$$\mathbf{z}_k(n) = \frac{\mathbf{s}(n)^H \mathbf{H}_k^H \boldsymbol{\eta}_k(n)}{\|\mathbf{H}_k \mathbf{s}(n)\|}. \quad (5.10)$$

Uma limitação dessa abordagem é que o vetor de sinal \mathbf{w}_{opt} não pode depender do tempo. Em [66], argumenta-se que essa condição é encontrada em sistemas em que \mathbf{w}_{opt} corresponde a uma sequência de símbolos pilotos repetidos periodicamente, por exemplo. Esta dificuldade pode ser contornada se o teste de hipóteses for interpretado da seguinte forma:

$$\mathbf{u}_k(n) = \mathbf{H}_k \mathbf{s}(n), \quad (5.11)$$

$$d_k(n) = \sum_{q=n-M+1}^n y_k(q) = \mathbf{u}_k(n)^T \mathbf{w}_{opt} + \eta(n), \quad (5.12)$$

em que:

$$\begin{cases} \mathcal{H}_0 : \mathbf{w}_{opt} = \mathbf{0} \\ \mathcal{H}_1 : \mathbf{w}_{opt} = \mathbf{1} \end{cases}. \quad (5.13)$$

Observe que utilizando as transformações mostradas em (5.11), (5.12) e (5.13) não é necessário que o vetor de sinal apresente qualquer tipo de periodicidade.

Com o objetivo de aplicar os resultados do Capítulo 4 e calcular qual deveria ser a distribuição ótima de passo de adaptação da rede em termos de MSD, nesse ponto será feita a restrição de que o canal apresenta um único percurso, *i.e.*, $L = 1$.

No caso em que a entrada é branca, a matriz de convolução do canal assume a forma:

$$\mathbf{H}_k = h_k \mathbf{I}_M. \quad (5.14)$$

Logo, a matriz de autocorrelação do sinal de entrada, $\mathbf{u}_k(n)$ será dada por:

$$\mathbf{R}_{\mathbf{u}_k} = |h_k|^2 \sigma_s^2 \mathbf{I}_M. \quad (5.15)$$

Seguindo o mesmo desenvolvimento abordado no Capítulo 4, pode-se verificar

que a análise de média leva à conclusão de que, se os passos de adaptação em cada rádio cognitivo satisfazem a condição:

$$\sum_{k=1}^K |h_k|^2 \mu_k = \bar{\mu}, \quad (5.16)$$

em que $\bar{\mu}$ é uma constante, a taxa de convergência do algoritmo é a mesma para quaisquer que sejam os passos em cada rádio cognitivo.

Além disso, de modo semelhante ao mostrado no Capítulo 4, pode-se verificar que no caso de detecção com algoritmo ATC, o MSD será dado por:

$$\mathcal{M} \approx S \frac{1}{K} \sum_{k=1}^K |h_k|^2 \mu_k^2 \sigma_{\eta,k}^2, \quad (5.17)$$

em que S é um termo constante que depende apenas da média dos passos de adaptação:

$$S = \sum_{k=1}^K \sum_{m=1}^M \frac{\lambda_u(m) \lambda_c^2(n)}{1 - [\lambda_c(n)(1 - \bar{\mu} \lambda_u(m))]^2}. \quad (5.18)$$

Portanto, dada uma taxa de convergência constante, a configuração de passo que minimiza o MSD será dado pela solução do seguinte problema de otimização:

$$\begin{aligned} & \underset{\boldsymbol{\mu} = [\mu_1 \ \mu_1 \ \dots \ \mu_K]}{\text{minimizar}} && \sum_{k=1}^K |h_k|^2 \mu_k^2 \sigma_{\eta,k}^2 \\ & \text{sujeito a} && \sum_{k=1}^K \mu_k |h_k|^2 = \bar{\mu} \end{aligned}$$

Este problema pode ser resolvido por multiplicadores de Lagrange, de modo semelhante ao já detalhado no Capítulo 4, tendo como solução:

$$\mu_{opt}(k) = \frac{\bar{\mu}}{\sigma_{\eta,k}^2 \sum_{k=1}^K \frac{|h_k|^2}{\sigma_{\eta,k}^2}}. \quad (5.19)$$

Como exemplo, consideraremos a rede com $K = 6$ rádios cognitivos mostrada na Fig. 5.1. A SNR em cada rádio cognitivo está mostrada na Tabela 5.1.

CR	1	2	3	4	5	6
SNR (dB)	1	-4	-10	-5	2	-8

Tabela 5.1: SNR em cada rádio cognitivo na rede da Fig. 5.1.

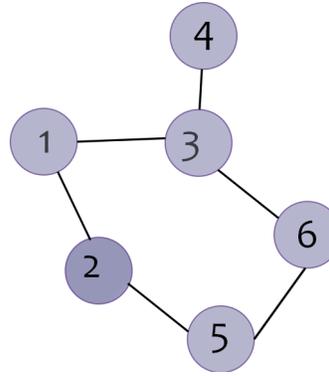


Figura 5.1: Rede com $K = 6$ rádios cognitivos

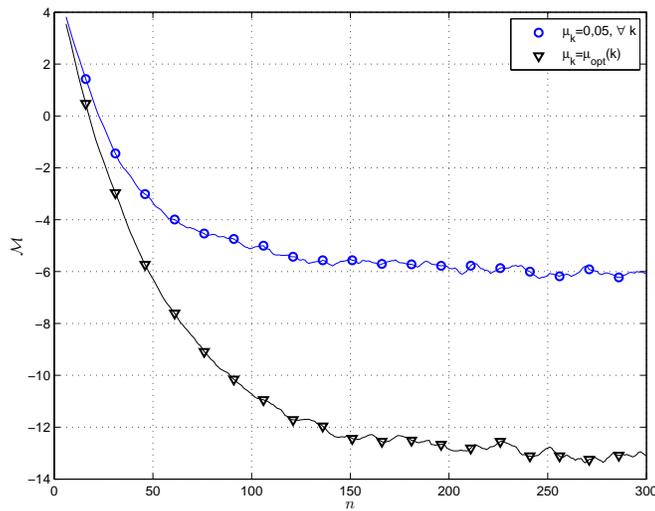


Figura 5.2: Comparação de desempenho em termos de MSD entre o sistema com passos de adaptação iguais em todos os nós e o sistema com passo otimizado

Considerando que o canal entre cada rádio cognitivo é do tipo Rayleigh com $L = 1$ trajeto, pode-se estimar o passo de adaptação ótimo de acordo com (5.19). A Fig. 5.2 mostra a comparação entre o MSD médio da rede considerando dois diferentes contextos considerando a matriz de combinação definida a partir da Regra Metropolis. No primeiro caso, o passo de adaptação é o mesmo em todos os nós. Já no segundo, o passo está otimizado de acordo com (5.19). Observamos que há uma diferença de desempenho de cerca de 7,5dB entre os dois modos de atribuição de passo de adaptação.

A Fig. 5.3 mostra como a diferença de desempenho em termos de MSD se reflete no desempenho do detector. Observamos que há redução tanto na probabilidade de

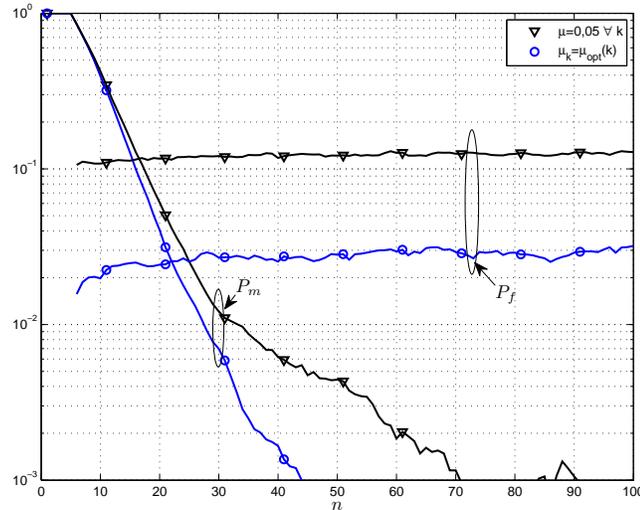


Figura 5.3: Comparação de desempenho de detecção entre o sistema com passos de adaptação iguais em todos os nós e o sistema com passo otimizado

perda de detecção quanto na probabilidade de falso alarme quando o passo otimizado é utilizado. No caso de canais com multipercuso, esse detector também poderia ser utilizado. Entretanto, a estratégia para a determinação do passo ótimo seria válida apenas se os canais entre os rádios cognitivos e o usuário primário apresentasse a mesma função de autocorrelação.

Uma limitação na aplicação de tal técnica é que é necessário realizar uma detecção coerente no teste de hipótese, utilizando a técnica do filtro casado mostrada na Seção 2.2. Esse fato implica em um conhecimento *a priori* do sinal do usuário primário, bem como do canal.

Na próxima Seção, trataremos do caso em que o sinal transmitido pelo usuário primário é conhecido, entretanto o canal não o é. Esse contexto pode ser encontrado no caso que os usuários primário e secundário compartilham uma mesma sequência de treinamento.

5.2 Detecção sem conhecimento perfeito de canal

Como dito na Seção anterior, o maior problema relativo às estratégias de detecção apresentadas em [66] é o fato de as técnicas serem fundamentadas em detecção do tipo filtro casado, sendo necessário conhecer tanto o canal entre os usuários primário

e secundário quanto a sequência transmitida pelo usuário primário [14].

Há contextos, entretanto, em que a sequência enviada é uma sequência de treinamento conhecida, mas não há conhecimento do canal. Analisaremos, a seguir, esse caso de detecção com conhecimento incompleto do sinal.

Segmentaremos a análise em duas etapas. Primeiramente, analisaremos o desempenho do detector considerando apenas um rádio cognitivo. Denominaremos essa etapa de sensoriamento local. Em seguida, será analisado o desempenho do sistema de sensoriamento espectral cooperativo.

5.2.1 Sensoriamento local

Em cada rádio cognitivo, o sinal recebido pode ser modelado da seguinte forma:

$$d(n) = \mathbf{u}(n)^T \mathbf{w}_{opt}(n) + \eta(n), \quad (5.20)$$

em que \mathbf{w}_{opt} pode assumir um dos seguintes valores:

$$\mathbf{w}_{opt}(n) = \begin{cases} 0, & \mathcal{H}_0 \\ \mathbf{h}, & \mathcal{H}_1 \end{cases}, \quad (5.21)$$

$\mathbf{u}(n)$ é o sinal transmitido pelo usuário primário, uma sequência de treinamento conhecida, e \mathbf{h} é o canal entre o usuário secundário e o usuário primário, que é desconhecido.

Escrevendo o problema sob a hipótese de sinal, em formato matricial, temos que o sinal recebido pelo usuário secundário no instante de tempo N é dado por:

$$\underbrace{\begin{bmatrix} d(1) \\ d(2) \\ \vdots \\ d(N) \end{bmatrix}}_{\mathbf{d}} = \underbrace{\begin{bmatrix} u(1) & 0 & \cdots & 0 \\ u(2) & u(1) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ u(L) & \cdots & u(2) & u(1) \\ \vdots & \vdots & \vdots & \vdots \\ u(N) & \cdots & u(N-1) & u(N-L+1) \end{bmatrix}}_{\mathbf{U}} \underbrace{\begin{bmatrix} h(1) \\ h(2) \\ \vdots \\ h(L) \end{bmatrix}}_{\mathbf{h}}. \quad (5.22)$$

Em [14], é mostrado que a estatística de teste ótima é dada por:

$$T(\mathbf{d}) = \hat{\mathbf{h}}^H \mathbf{U}^H \mathbf{U} \hat{\mathbf{h}}, \quad (5.23)$$

em que $\hat{\mathbf{h}}$ é o estimador de máxima verossimilhança (MLE) de \mathbf{h} , dadas as observações $d(n)$. Observe que, considerando que o sinal transmitido pelo usuário primário seja branco de média nula, a seguinte aproximação pode ser feita:

$$\mathbf{U}^H \mathbf{U} \approx N \mathbf{I}, \quad (5.24)$$

assim, substituindo (5.24) em (5.23), temos que uma estatística de teste sub-ótima é dada por:

$$\tilde{T}(\mathbf{d}) = \hat{\mathbf{h}}^H \hat{\mathbf{h}}. \quad (5.25)$$

Para obter o estimador de máxima verossimilhança, $\hat{\mathbf{h}}$, é necessário realizar operações custosas, tal como inversão e multiplicação de matrizes de ordem elevada. Uma alternativa para contornar esse problema seria utilizar um estimador sub-ótimo para \mathbf{h} , aplicando um algoritmo do tipo LMS. Denotando a estimativa no LMS no instante n por $\mathbf{w}(n)$, o rádio cognitivo decidirá que o usuário primário está ativo se a energia do sinal estimado $\mathbf{w}(n)$ estiver acima de um limiar pré estabelecido:

$$\|\mathbf{w}(n)\|^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda. \quad (5.26)$$

Dessa forma, a probabilidade de falso alarme em cada nó será dada por:

$$P_f = Pr \{ \|\mathbf{w}(n)\|^2 \geq \lambda | \mathcal{H}_0 \}, \quad (5.27)$$

e a probabilidade de perda de detecção será dada por:

$$P_m = 1 - Pr \{ \|\mathbf{w}(n)\|^2 \geq \lambda | \mathcal{H}_1 \}. \quad (5.28)$$

O desempenho do detector será determinado a partir da função densidade de probabilidade da variável aleatória $X = \|\mathbf{w}(n)\|^2$.

Observe que é necessário que o rádio cognitivo conheça o sinal $\mathbf{u}(n)$, entretanto não é necessário que o canal $\mathbf{h}(n)$ seja perfeitamente conhecido.

A seguir, apresentaremos o desempenho do detector com o LMS em regime, e, posteriormente, apresentaremos o desempenho do detector na fase de transitório do LMS.

Desempenho do detector em regime estacionário

As densidades de probabilidades da variável aleatória $X = \|\mathbf{w}(n)\|^2$ sob cada uma das hipóteses estão apresentadas nas duas proposições a seguir.

Proposition 5.2.1 (Modelagem estatística da variável aleatória $X = \|\mathbf{w}(n)\|^2$ sob a hipótese nula). *A variável aleatória $X = \|\mathbf{w}(n)\|^2$, em regime de estacionaridade e sob a hipótese nula, pode ser modelada como uma variável aleatória gama. Dessa forma, sua densidade de probabilidade é aproximada por:*

$$f_X(x) \approx \frac{1}{\Gamma(\kappa)\theta^\kappa} x^{\kappa-1} e^{-x/\theta}. \quad (5.29)$$

Já a função distribuição é aproximada por:

$$F_X(x) \approx \frac{\gamma(\kappa, x/\theta)}{\Gamma(\kappa)}, \quad (5.30)$$

em que θ e κ são, respectivamente, os parâmetros de escala e forma da distribuição. Esses parâmetros se relacionam com os momentos de primeira e segunda ordem da variável gama da seguinte forma:

$$\theta = \frac{\text{var}\{X\}}{\mathbb{E}\{X\}}, \quad (5.31)$$

$$\kappa = \frac{\mathbb{E}\{X\}}{\theta}. \quad (5.32)$$

Esses parâmetros podem ser aproximados por:

$$\theta_0 = \frac{\mu\sigma_\eta^2}{1 - \mu/2}, \quad (5.33)$$

$$\kappa_0 = \frac{M}{2}, \quad (5.34)$$

em que M é o número de coeficientes do filtro.

Demonstração. Sob a hipótese nula, $\mathbf{h} = \mathbf{0}$, de modo que $X = \|\mathbf{w}(n)\|^2$ é equivalente a $X = \|\mathbf{w}(n) - \mathbf{h}\|^2 = \|\tilde{\mathbf{w}}(n)\|^2$. Dessa forma,

$$X = \sum_{k=0}^{M-1} |\tilde{w}(n-k)|^2. \quad (5.35)$$

Neste ponto, assumimos que $X_k = \tilde{w}(n-k)$ apresenta distribuição aproximadamente gaussiana com média nula e variância dada por $\sigma_{X_k}^2 = \frac{\mathcal{M}}{M}$, em que \mathcal{M} é o MSD associado ao filtro. Dessa forma, podemos afirmar que X tem distribuição estatística equivalente à variável aleatória \overline{X} definida como:

$$\overline{X} = \sum_{k=0}^{M-1} \left(\sqrt{\frac{\mathcal{M}}{M}} X_k \right)^2 = \frac{\mathcal{M}}{M} \sum_{k=0}^{M-1} X_k^2, \quad (5.36)$$

em que $\overline{X}_k \sim \mathcal{N}(0, 1)$. Portanto, temos que:

$$\sum_{k=0}^{M-1} \overline{X}_k^2 \sim \chi^2(M), \quad (5.37)$$

em que $\chi^2(M)$ é uma variável aleatória do tipo χ^2 com M graus de liberdade. Concluimos que $X = \|\tilde{\mathbf{w}}(n)\|^2$ pode ser expressa como αY em que α é uma constante dada por $\alpha = \frac{\mathcal{M}}{M}$ e Y é uma variável aleatória do tipo χ^2 com M graus de liberdade.

O MSD é aproximado por [53]:

$$\mathcal{M} \approx M \frac{\mu \sigma_\eta^2}{2 - \mu}, \quad (5.38)$$

de modo que:

$$\alpha = \frac{\mu \sigma_\eta^2}{2 - \mu}. \quad (5.39)$$

Ainda, a função densidade de probabilidade de X é dada por:

$$f_X(x) = \frac{2 - \mu}{\mu \sigma_\eta^2} f_Y \left(\frac{x(2 - \mu)}{\mu \sigma_\eta^2} \right), \quad (5.40)$$

em que $f_Y(x)$ é a função densidade de probabilidade da variável aleatória Y [14]:

$$f_Y(x) = \frac{1}{2^{M/2} \Gamma\left(\frac{M}{2}\right)} x^{M/2-1} \exp\left(-\frac{x}{2}\right). \quad (5.41)$$

Logo:

$$f_X(x) = \frac{2 - \mu}{\mu \sigma_\eta^2} \frac{1}{2^{M/2} \Gamma\left(\frac{M}{2}\right)} \left(\frac{x(2 - \mu)}{\mu \sigma_\eta^2} \right)^{M/2-1} \exp\left(-\frac{x(2 - \mu)}{2\mu \sigma_\eta^2}\right). \quad (5.42)$$

Reorganizando os termos, reescrevemos $f_X(x)$ como:

$$f_X(x) = \frac{1}{\Gamma\left(\frac{M}{2}\right) \left(\frac{\mu\sigma_\eta^2}{1-\mu/2}\right)^{M/2}} x^{M/2-1} \exp\left(-\frac{x(1-\mu/2)}{\mu\sigma_\eta^2}\right). \quad (5.43)$$

Comparando (5.43) com (5.29), concluímos que X de fato apresenta distribuição gama com parâmetros dados por (5.33) e (5.34).

□

Proposition 5.2.2 (Modelagem estatística da variável aleatória $X = \|\mathbf{w}(n)\|^2$ sob a hipótese de sinal). *Sob a hipótese de sinal, a função densidade de probabilidade da variável aleatória X pode ser aproximada por:*

$$f_X(x) \approx \frac{1-\mu/2}{\mu\sigma_\eta^2} \exp\left(-\frac{1-\mu/2}{\mu\sigma_\eta^2}(x + \|\mathbf{h}\|^2)\right) \left(\frac{x}{\|\mathbf{h}\|^2}\right)^{\frac{M}{4}-\frac{1}{2}} I_{M/2-1}\left(\|\mathbf{h}\| \frac{2-\mu}{\mu\sigma_\eta^2} \sqrt{x}\right). \quad (5.44)$$

Demonstração. Sob a hipótese de sinal, a variável aleatória $X = \|\mathbf{w}(n)\|^2$ pode ser escrita como:

$$X = \|\mathbf{w}(n)\|^2 = \sum_{k=0}^{M-1} |\tilde{w}_k(n) + h_k|^2. \quad (5.45)$$

A variável aleatória $X_k = \tilde{w}_k(n) + h_k$ pode ser aproximada por uma variável aleatória gaussiana de média h_k e variância $\sigma_{X_k}^2 = \frac{MSD}{M} = \frac{\mu\sigma_\eta^2}{2-\mu}$, $\forall k$. Portanto, a variável aleatória \bar{X} :

$$\bar{X} = \sum_{k=0}^{M-1} \left| \sqrt{\frac{2-\mu}{\mu\sigma_\eta^2}} (\tilde{w}_k(n) + h_k) \right|^2, \quad (5.46)$$

corresponde à soma dos quadrados de M variáveis aleatórias gaussianas de média $\mathbb{E}\{X_k\} = h_k \sqrt{\frac{2-\mu}{\mu\sigma_\eta^2}}$ e variância unitária. Logo, podemos concluir que \bar{X} apresenta distribuição de probabilidade do tipo χ^2 descentralizada com M graus de liberdade e parâmetro de descentralização dado por [14]:

$$\bar{\lambda} = \sum_{k=0}^{M-1} \mathbb{E}\{X_k\}^2 = \sum_{k=0}^{M-1} \left(h_k \sqrt{\frac{2-\mu}{\mu\sigma_\eta^2}} \right)^2 = \frac{2-\mu}{\mu\sigma_\eta^2} \|\mathbf{h}\|^2. \quad (5.47)$$

Sendo assim, a densidade de probabilidade de \bar{X} é aproximada por:

$$f_{\bar{X}} \approx \frac{1}{2} \exp \left(-\frac{1}{2} \left(x + \frac{2-\mu}{\mu\sigma_\eta^2} \|\mathbf{h}\|^2 \right) \right) \left(\frac{x\mu\sigma_\eta^2}{\|\mathbf{h}\|^2(2-\mu)} \right)^{M/4-1/2} I_{M/2-1} \left(\|\mathbf{h}\| \sqrt{x \frac{2-\mu}{\mu\sigma_\eta^2}} \right). \quad (5.48)$$

Como $X = \frac{\mu\sigma_\eta^2}{2-\mu} \bar{X}$, temos que $f_X(x) = \frac{2-\mu}{\mu\sigma_\eta^2} f_{\bar{X}} \left(\frac{2-\mu}{\mu\sigma_\eta^2} x \right)$, o que resulta em (5.44). \square

A validade da Proposição 5.2.1 pode ser verificada nas Figs. 5.4, 5.5 e 5.6, em que a densidade de probabilidade teórica descrita na Proposição é comparada com o histograma gerado a partir de $N_r = 10^5$ realizações da variável aleatória X . Para cada configuração de passo de adaptação, número de coeficientes e potência de ruído testadas, estão apresentadas duas curvas de densidade de probabilidade teóricas. Na curva em vermelho, considera-se como parâmetros da densidade de probabilidade gama as aproximações dadas por (5.33) e (5.34). Já na curva em preto, os parâmetros são obtidos estimando os momentos de X por simulação e aplicando (5.31) e (5.32).

Os resultados mostrados na Fig. 5.4 foram gerados fixando a potência de ruído em $\sigma_\eta^2 = 1$, o número de coeficientes do filtro em $M = 1$ e os seguintes valores de passo de adaptação: $\mu = 0,05, 0,1$ e $0,2$. Observe que, para todas as configurações testadas, as densidades de probabilidade teóricas, tanto para os parâmetros teóricos dados por (5.31) e (5.32), quanto para os parâmetros obtidos por simulação Monte-Carlo coincidem com o histograma. Esse fato reforça a validade da Proposição para os diferentes valores de passo. A Proposição também parece ser válida para diferentes valores de potência de ruído, como indicado na Fig. 5.5. Já a Fig. 5.6, mostra que, para valores mais altos de M , o histograma difere das densidades teóricas geradas com os parâmetros dados por (5.33) e (5.34). Essa fato pode ser creditado à adoção da hipótese de que os coeficientes $\tilde{w}(n-k)$ são independentes entre si para garantir a validade de (5.38), o que não acontece na prática. Todavia, as densidades de probabilidade teóricas obtidas com os parâmetros simulados apresentam um comportamento próximo àquele observado no histograma. Dessa forma, concluímos que a variável aleatória X realmente apresenta distribuição do tipo gama, mas os seus parâmetros não correspondem a (5.33) e (5.34).

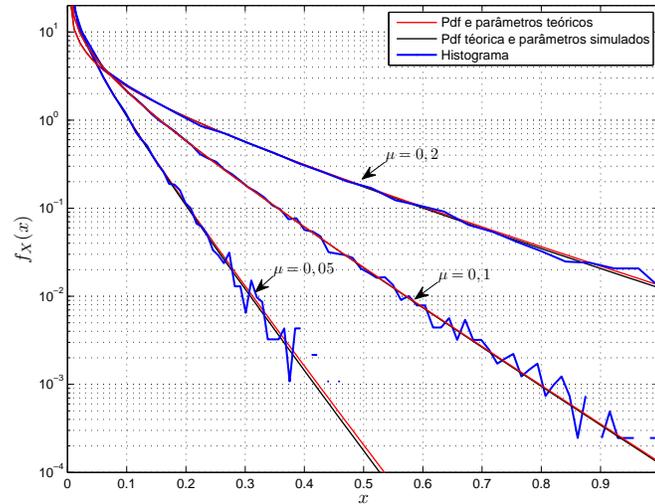


Figura 5.4: Comparação entre a densidade de probabilidade teórica dada na Proposição 5.2.1, considerando $\sigma_\eta^2 = 1$, $M = 1$ e vários valores de μ , e o histograma gerado a partir de $N_r = 10^5$ realizações da variável aleatória X sob a hipótese nula.

Verificou-se também a validade da Proposição sob a hipótese \mathcal{H}_1 e os resultados estão mostrados na Figs. 5.7, 5.8 e 5.9. Esses resultados mostram a comparação entre as densidades de probabilidade teóricas e o histograma gerado a partir de $N_r = 10^5$ realizações da variável aleatória X_1 . Considerou-se que o vetor \mathbf{h} é um vetor com todas as componentes iguais a 1. Nota-se que assim como no caso da hipótese nula, a conjectura é bem robusta aos valores de passo de adaptação e potência de ruído. Entretanto, para valores mais elevados de M , os parâmetros da distribuição não parecem ser dados como apresentados na Proposição 5.2.2.

Assumindo a validade das proposições 5.2.1 e 5.2.2, as probabilidades de falso alarme e perda de detecção são dadas por:

$$P_f(\lambda) = Pr\{X > \lambda; \mathcal{H}_0\} = 1 - Pr\{X < \lambda; \mathcal{H}_0\} = 1 - \frac{\gamma(\lambda/\theta_0, \kappa_0)}{\Gamma(\kappa_0)}, \quad (5.49)$$

$$\begin{aligned} P_m(\lambda) &= Pr\{X < \lambda; \mathcal{H}_1\} = \frac{\gamma(\lambda/\theta_1, \kappa_1)}{\Gamma(\kappa_1)} P_m(\lambda) = Pr\{X < \lambda; \mathcal{H}_1\} = \\ &= 1 - Q_{M/2} \left(\sqrt{\|\mathbf{h}\|^2 \frac{2-\mu}{\mu\sigma_\eta^2}}, \frac{\lambda(2-\mu)}{\mu\sigma_\eta^2} \right). \end{aligned} \quad (5.50)$$

As Figs. 5.10 e 5.11 mostram as probabilidades de falso alarme e perda de

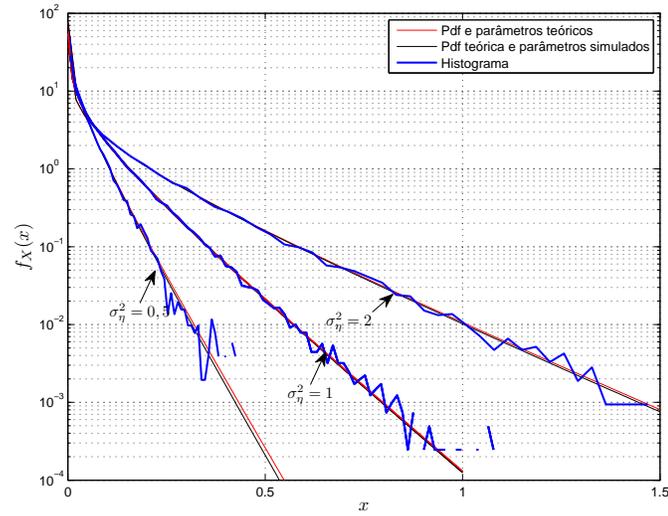


Figura 5.5: Comparação entre a densidade de probabilidade teórica dada na Proposição 5.2.1, considerando $M = 1$, $\mu = 0,1$ e vários valores de σ_η^2 , e o histograma gerado a partir de $N_r = 10^5$ realizações da variável aleatória X sob a hipótese nula.

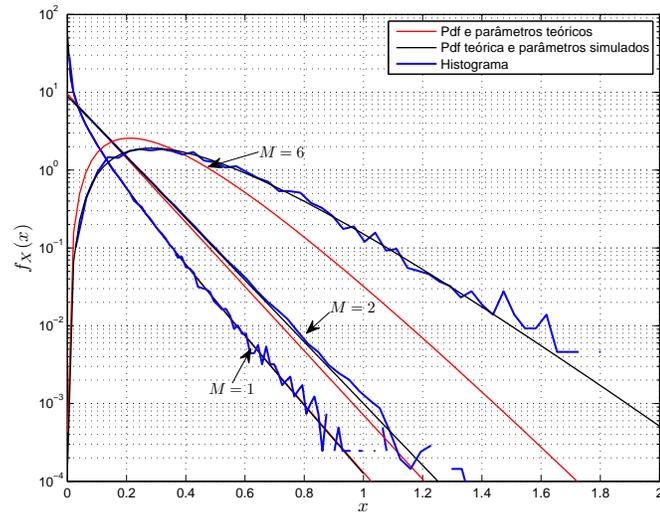


Figura 5.6: Comparação entre a densidade de probabilidade teórica dada na Proposição 5.2.1, considerando $\sigma_\eta^2 = 1$, $\mu = 0,1$, e vários valores de M e o histograma gerado a partir de $N_r = 10^5$ realizações da variável aleatória X sob a hipótese nula.

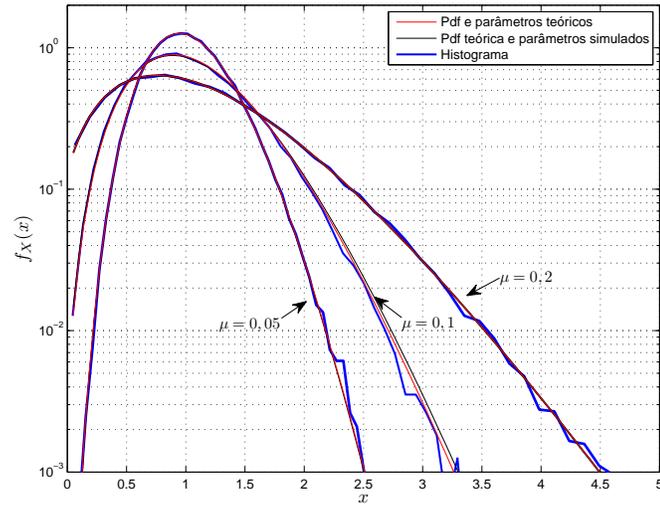


Figura 5.7: Comparação entre a densidade de probabilidade teórica dada na Proposição 5.2.2, considerando $\sigma_\eta^2 = 1$, $M = 1$, e vários valores de μ , e o histograma gerado a partir de $N_r = 10^5$ realizações da variável aleatória X sob a hipótese de sinal.

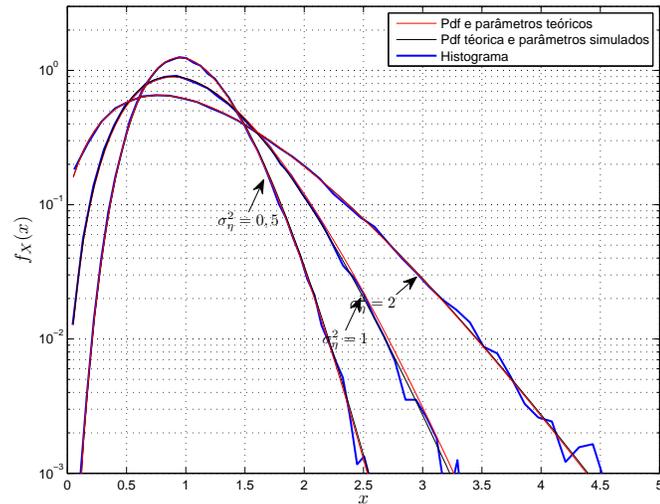


Figura 5.8: Comparação entre a densidade de probabilidade teórica dada na Proposição 5.2.2, considerando $M = 1$, $\mu = 0,1$ e vários valores de σ_η^2 , e o histograma gerado a partir de $N_r = 10^5$ realizações da variável aleatória X sob a hipótese de sinal.

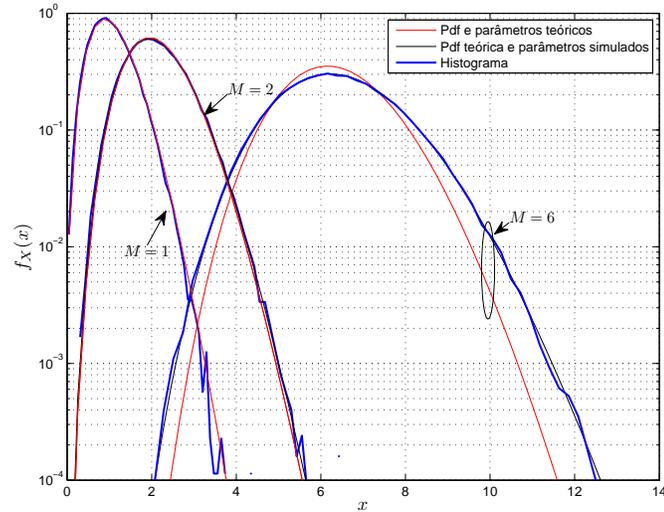


Figura 5.9: Comparação entre a densidade de probabilidade teórica dada na dada na Proposição 5.2.2, considerando $\sigma_\eta^2 = 1$, $\mu = 0,1$ e vários valores de M , e o histograma gerado a partir de $N_r = 10^5$ realizações da variável aleatória X sob a hipótese de sinal.

detecção em função do limiar λ . Percebe-se que a expressão teórica para P_f é, de fato, bastante precisa mesmo para valores baixos de P_f (*i.e.*, na cauda da distribuição). Já no caso da probabilidade de perda de detecção, nota-se que a expressão teórica é mais precisa para valores mais elevados de P_m , conforme esperado. Essas curvas foram geradas calculando as probabilidades de perda de detecção e falso alarme após a convergência do algoritmo. A Fig. 5.12 mostra as curvas de convergência para P_f e P_m considerando $\lambda = 0,4$ e passo de adaptação $\mu = 0,1$. Pode-se observar que o algoritmo convergiu após aproximadamente $N = 40$ iterações.

Desempenho do detector em regime transitório

As Proposições 5.2.1 e 5.2.2 são válidas assumindo que o sistema encontra-se em regime. Porém, também seria interessante avaliar o comportamento do sistema em regime transitório.

Nesse ponto, iremos considerar que o sinal transmitido pelo usuário primário é do tipo BPSK em que $u(n) = \pm 1$ com igual probabilidade.

Seja inicialmente um filtro de apenas um coeficiente. Tem-se que a seguinte

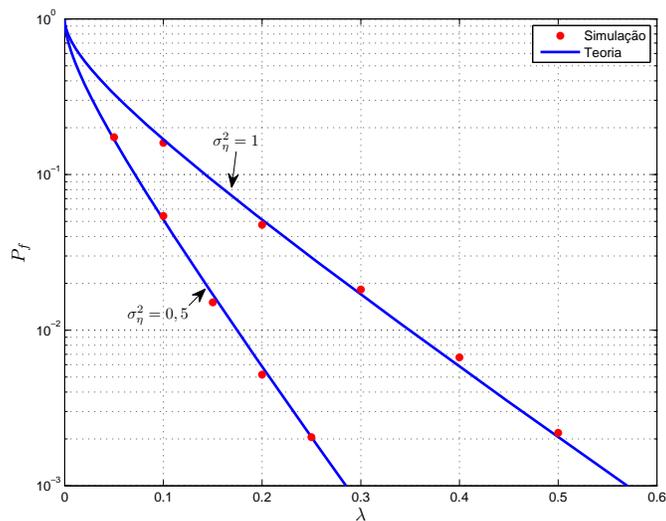


Figura 5.10: Comparação entre as probabilidade de falso alarme teóricas e simuladas, considerando $\mu = 0 = 0,1$, $M = 1$ e duas configurações de potência de ruído diferentes: $\sigma_\eta^2 = 1$ e $\sigma_\eta^2 = 0,5$

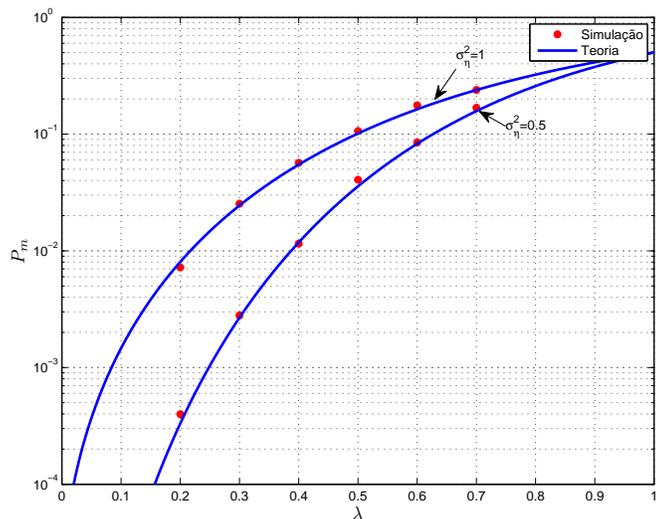


Figura 5.11: Comparação entre as probabilidade de perda de detecção alarme teóricas e simuladas, considerando $\mu = 0 = 0,1$, $M = 1$ e duas configurações de potência de ruído diferentes: $\sigma_\eta^2 = 1$ e $\sigma_\eta^2 = 0,5$

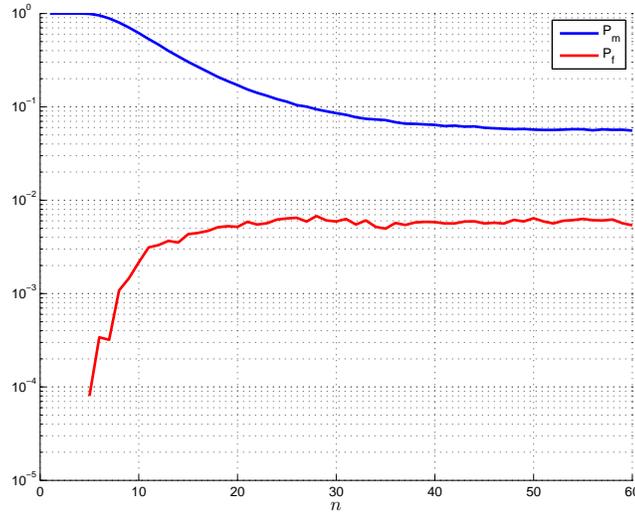


Figura 5.12: Convergência de P_f e P_m , considerando $\sigma_\eta^2 = 1$ e $\lambda = 0,4$

equação de atualização de coeficientes deve ser obedecida:

$$w(n+1) = w(n) + \mu [u(n)(h - w(n)) + \eta(n)] u(n). \quad (5.51)$$

Como $u(n)^2 = 1$, temos que:

$$w(n+1) = (1 - \mu) w(n) + \mu [h + u(n)\eta(n)]. \quad (5.52)$$

É interessante notar que $u(n)\eta(n)$ também é uma variável aleatória gaussiana e de mesma distribuição de probabilidade de $\eta(n)$. Essa conclusão vem do fato de que para os dois possíveis valores de $u(n)$, $u(n) = 1$ ou $u(n) = -1$, não há mudança nem da média $\eta(n)$, já que $\eta(n)$ tem média nula, nem de sua variância.

Portanto, denotando a variável aleatória $\mu [u(n)\eta(n) + h]$ por $\zeta(n)$, a equação de atualização de coeficientes pode ser escrita como:

$$w(n+1) = (1 - \mu) w(n) + \zeta(n), \quad (5.53)$$

em que $\zeta(n) \sim \mathcal{N}(\mu h, \mu^2 \sigma_\eta^2)$.

Dessa forma, podemos reconhecer (5.53) como a equação de um processo autorregressivo de ordem 1 excitado por uma entrada gaussiana de média não-nula.

Considerando que o sistema parte da condição inicial $w(0) = 0$, temos que a

variável aleatória $w(n) = 0$ é gaussiana com média dada por:

$$\mathbb{E}\{w(n)\} = \mu h \sum_{k=0}^{n-1} (1-\mu)^k = \mu h \frac{1 - (1-\mu)^n}{1 - (1-\mu)} = h(1 - (1-\mu)^n). \quad (5.54)$$

Note que sob a hipótese nula, $h = 0$, o que implica em $\mathbb{E}\{w(n)\} = 0$. Já sob a hipótese de sinal, temos que $\mathbb{E}\{w(n)\} \rightarrow h$ para valores crescentes de n .

Já a variância de $w_k(n)$ é dada por:

$$\sigma_{w(n)}^2 = \mu^2 \sigma_\eta^2 \sum_{k=0}^{n-1} (1-\mu)^{2k} = \mu^2 \sigma_\eta^2 \frac{1 - (1-\mu)^{2n}}{1 - (1-\mu)^2} = \mu \sigma_\eta^2 \frac{1 - (1-\mu)^{2n}}{2 - \mu}. \quad (5.55)$$

Note também que para valores crescentes de n , $\sigma_{w(n)}^2$ tende à expressão de MSD relacionada a um filtro de um coeficiente, conforme esperado.

A partir dessa análise, podemos estender esse resultado para um canal com M coeficientes e aplicar o mesmo tratamento usado nas proposições 5.2.1 e 5.2.2, com a exceção de que as expressões de média e variância de cada parcela X_k passam a ser dadas por (5.54) e (5.55). Ajustando as expressões das probabilidades de falso alarme e perda de detecção dadas em (5.49) e (5.50), temos que essas probabilidades no transiente do filtro são expressas por:

$$P_f(\lambda) = 1 - \gamma \left(\frac{\lambda}{\mu \sigma_\eta^2} \frac{(1 - \frac{\mu}{2})}{1 - (1-\mu)^{2n}}, \frac{M}{2} \right) \quad (5.56)$$

e

$$P_m(\lambda) = 1 - Q_{M/2} \left(\sqrt{\|\mathbf{h}\|^2 \left[\frac{2-\mu}{\mu \sigma_\eta^2} \right] \frac{(1 - (1-\mu)^n)^2}{1 - (1-\mu)^{2n}}}, \frac{\lambda(2-\mu)}{\mu \sigma_\eta^2 (1 - (1-\mu)^{2n})} \right). \quad (5.57)$$

Denotando $(1 - (1-\mu)^n)^2$ por $A(\mu)$ e $1 - (1-\mu)^{2n}$ por $B(\mu)$, podemos reescrever as expressões acima como:

$$P_f(\lambda) = 1 - \gamma \left(\frac{\lambda}{\mu \sigma_\eta^2} \frac{(1 - \frac{\mu}{2})}{B(\mu)}, \frac{M}{2} \right) \quad (5.58)$$

e

$$P_m(\lambda) = 1 - Q_{M/2} \left(\sqrt{\|\mathbf{h}\|^2 \left[\frac{2-\mu}{\mu \sigma_\eta^2} \right] \frac{A(\mu)}{B(\mu)}}, \frac{\lambda(2-\mu)}{\mu \sigma_\eta^2 B(\mu)} \right). \quad (5.59)$$

Considerando que o passo de adaptação obedece à condição $\mu \ll 1$, e utilizando

a aproximação de Taylor em primeira ordem da função $f(x) = (1 - x)^a$, tem-se que:

$$f(x) = (1 - x)^a = 1 - ax, \quad (5.60)$$

o que permite aproximar a função $A(\mu)$ por:

$$A(\mu) = (1 - (1 - \mu)^n)^2 \approx n^2 \mu^2 \quad (5.61)$$

e, para $B(\mu)$, de forma semelhante:

$$B(\mu) = 1 - (1 - \mu)^{2n} \approx 2n\mu. \quad (5.62)$$

Substituindo as aproximações acima em (5.58) e (5.59), temos que as probabilidades de falso alarme e perda de detecção são dadas pelas seguintes expressões:

$$P_f(\lambda) = 1 - \gamma \left(\frac{\lambda}{\mu\sigma_\eta^2} \frac{(1 - \frac{\mu}{2})}{2n\mu}, \frac{M}{2} \right) \quad (5.63)$$

e

$$P_m(\lambda) = 1 - Q_{M/2} \left(\sqrt{\|\mathbf{h}\|^2 \left[\frac{2 - \mu}{\mu\sigma_\eta^2} \right] \frac{n\mu}{2}}, \frac{\lambda(2 - \mu)}{\mu\sigma_\eta^2(2n\mu)} \right), \quad (5.64)$$

ou ainda, considerando que $2 - \mu \approx 2$, temos que:

$$P_f(\lambda) = 1 - \gamma \left(\frac{\lambda}{\mu^2\sigma_\eta^2} \frac{1}{2n}, \frac{M}{2} \right) \quad (5.65)$$

e

$$P_m(\lambda) = 1 - Q_{M/2} \left(\sqrt{\frac{n\|\mathbf{h}\|^2}{\sigma_\eta^2}}, \frac{\lambda}{\mu^2\sigma_\eta^2 n} \right). \quad (5.66)$$

A Fig. 5.13 mostra as curvas de probabilidade de falso alarme e perda de detecção, por iteração, para detecção utilizando o algoritmo LMS, considerando o passo $\mu = 0,05$, potência de ruído dada por $\sigma_\eta^2 = 2$ e o seguinte canal entre o usuário primário e o usuário secundário:

$$\mathbf{h}_k = [1 \quad 0,2 \quad 0,4]^T. \quad (5.67)$$

Observa-se que as curvas de probabilidade obtidas por Monte Carlo são próximas das expressões teóricas dadas por (5.58) e (5.59), seja para a probabilidade de falso alarme, seja para a probabilidade de perda de detecção. Todavia, as curvas de

probabilidades obtidas utilizando a aproximação de passo pequeno não são próximas das obtidas por simulação para esse valor de passo de adaptação.

A Fig. 5.14 mostra as curvas de probabilidade de falso alarme e perda de detecção, considerando um passo de adaptação menor, $\mu = 0,005$. Com exceção do passo de adaptação, o contexto de simulação é o mesmo dos resultados mostrados na Fig. 5.13. Para esse valor de passo, as curvas obtidas com as expressões (5.65) e (5.66) estão bem mais próximas das curvas simuladas do que no caso da Fig. 5.13, mas ainda há uma diferença não desprezível entre os dois resultados. Já para o caso mostrado na Fig. 5.15, a aproximação para pequenos passos equivale às expressões dadas em (5.58) e (5.59). Nesse caso, o valor do passo $\mu = 0,0001$ é suficientemente pequeno para garantir a validade de (5.65) e de (5.66). É interessante observar que, para o caso da probabilidade de perda de detecção, as curvas simuladas e teóricas não são equivalentes, mesmo para passos muito pequenos. Essa discrepância pode ser atribuída à modelagem da variável aleatória X como uma χ^2 descentralizada e ponderada. Nesse caso, assume-se tacitamente que as componentes de $\mathbf{w}(n)$ são independentes, o que não é verdade. Esse fato pode ser confirmado analisando um filtro de um único coeficiente. Nesse caso, as curvas de perda de detecção teóricas e simuladas coincidiriam. Para corroborar essa afirmação, consideraremos a detecção do sinal primário assumindo que o canal é AWGN com ganho $h = 1$ e potência de ruído $\sigma_\eta^2 = 2$. Analisando o resultado mostrado na Fig. 5.16, concluímos que as curvas teóricas, de fato, são bem próximas das curvas de simulação.

Analisando as expressões teóricas de probabilidade de falso alarme e perda de detecção dadas por (5.65) e (5.66), notamos que, definindo o limiar de decisão como $\lambda' = \frac{\lambda}{\mu^2 \sigma_\eta^2}$, o teste passa a ser dado por:

$$T(\mathbf{d}_k) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda' = \frac{\lambda}{\mu^2 \sigma_\eta^2}. \quad (5.68)$$

Utilizando essa estatística de teste normalizada, temos que o desempenho do detector será dado por:

$$P_f(\lambda') = 1 - \gamma\left(\lambda' \frac{1}{2n}, \frac{M}{2}\right) \quad (5.69)$$

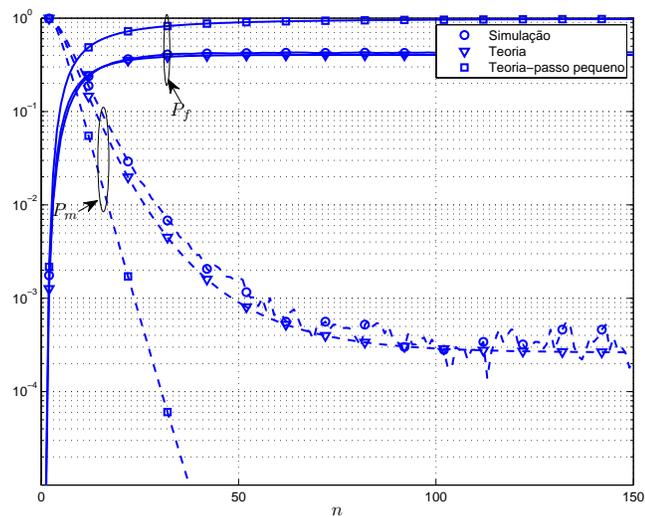


Figura 5.13: Desempenho do detector LMS considerando o canal dado por (5.67), limiar $\lambda = 0,15$, passo de adaptação $\mu = 0,05$ e $\sigma_{\eta}^2=2$.

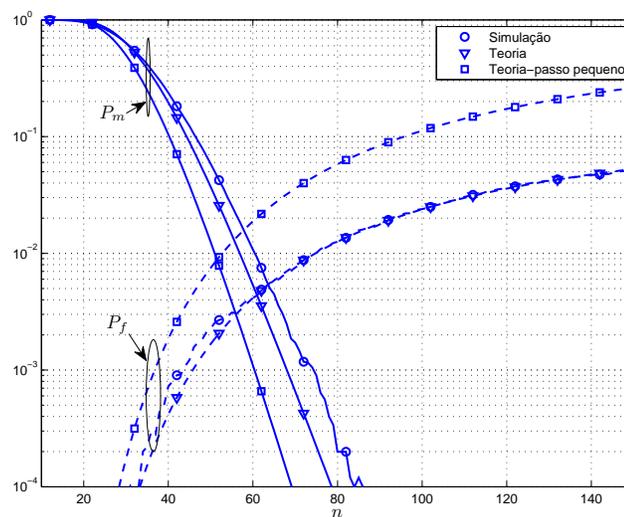


Figura 5.14: Desempenho do detector LMS considerando o canal dado por (5.67), limiar $\lambda = 0,03$, passo de adaptação $\mu = 0,005$ e $\sigma_{\eta}^2=2$.

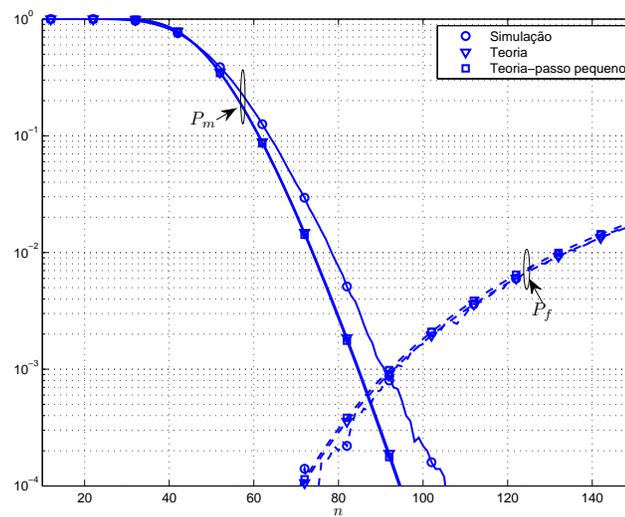


Figura 5.15: Desempenho do detector LMS considerando o canal dado por (5.67), limiar $\lambda = 0,0003$, passo de adaptação $\mu = 0,0001$ e $\sigma_\eta^2=2$.

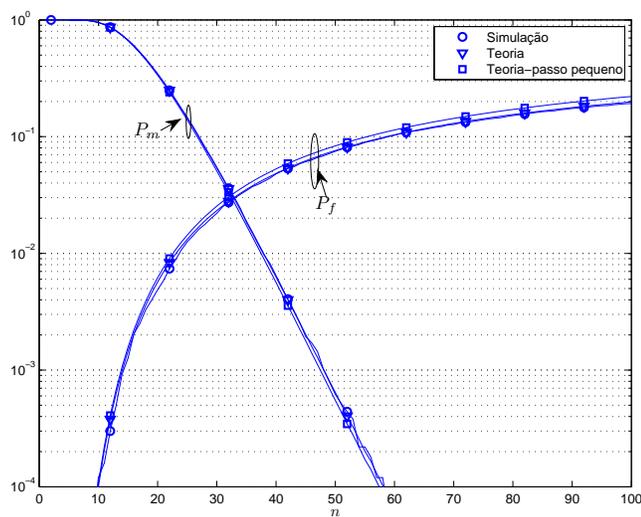


Figura 5.16: Desempenho do detector LMS considerando o canal dado por $h = 1$, limiar $\lambda = 0,0003$, passo de adaptação $\mu = 0,001$ e $\sigma_\eta^2=2$.

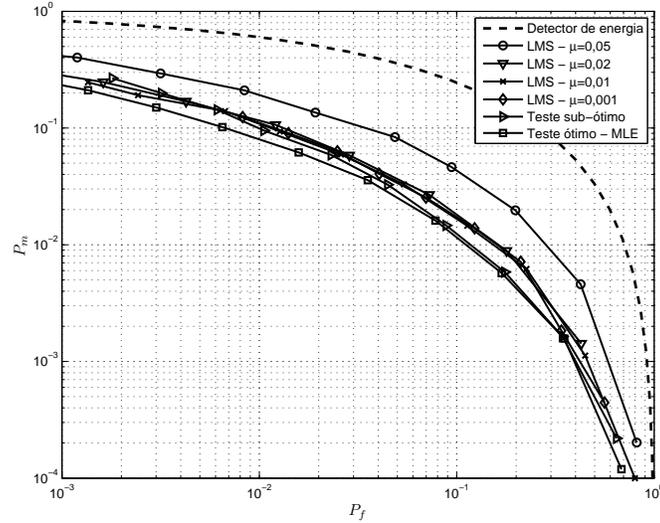


Figura 5.17: Comparação de desempenho entre os sistemas de detecção usando o algoritmo LMS configurado com diferentes passos de adaptação, o detector ótimo ML, o detector sub-ótimo dado em (5.25) e o detector de energia.

e

$$P_m(\lambda') = 1 - Q_{M/2} \left(\sqrt{\frac{n \|\mathbf{h}\|^2}{\sigma_\eta^2}}, \frac{\lambda'}{n} \right). \quad (5.70)$$

Analisando (5.69) e (5.70), observamos que considerando passos de adaptação pequenos, o desempenho do teste passa a ser assintoticamente independente do passo de adaptação adotado.

Para comparar o desempenho dos sistemas em termos de ROC, consideraremos mais um vez o canal dado em (5.67), potência de ruído $\sigma_\eta^2 = 2$ e vários valores de passo de adaptação. As curvas da Fig. 5.17 mostram a ROC do sistema de detecção com o algoritmo LMS comparada com os testes ótimo ML e subótimo dados por (5.25) e com o detector de energia. Perceba que, se o passo de adaptação não for muito elevado, o desempenho do teste através do algoritmo LMS é muito próximo do teste ML. Além disso, é interessante observar que, nesse caso, o desempenho do LMS com passo $\mu = 0,02$ já é equivalente ao desempenho do algoritmo LMS com passo $\mu = 0,01$. Reduzindo ainda mais o passo, considerando $\mu = 0,001$, nota-se que não há mudança de desempenho do algoritmo. Isso está de acordo com a análise das eqs. (5.69) e (5.70), onde percebemos que o desempenho do detector é invariante ao passo de adaptação, desde que o passo seja pequeno.

5.2.2 Detector cooperativo

Em um sistema com K rádios cognitivos, os usuários podem trocar informações entre si, de modo a melhorar a precisão do sensoriamento. Como foi analisado em Capítulos passados, há várias possibilidades para enviar e combinar as estimativas recebidas em cada nó. Uma das possibilidades seria o envio de um bit de decisão para os nós vizinhos da rede e, em cada nó, combinar as informações de uma das maneiras discutidas no Cap. 3, *i.e.*, aplicar uma das regras do tipo n em K . Essa estratégia é especialmente interessante por apresentar baixa complexidade, e por não precisar de nenhum tipo de conhecimento *a priori* do canal entre o usuário primário e o k -ésimo usuário secundário denotado por \mathbf{h}_k .

Apenas a título de ilustração, iremos considerar aqui um exemplo em que cada rádio envia um bit de decisão para os seus vizinhos e, localmente, será aplicada uma regra do tipo OU para que os rádios obtenham a decisão final.

Ainda, para que os rádios tenham acesso à informação de todos os nós da rede, a troca de decisões será efetuada de forma *com memória*, *i.e.*, para cada nova medida de sinal, deve-se esperar a troca de $K - 1$ decisões ao longo da rede. Dessa forma, garante-se que, dado que a rede é completamente conectada, todos os nós terão acesso aos bits de decisão em toda rede.

A Fig. 5.18 mostra a curva de ROC do sistema cooperativo, considerando um sistema com $K = 5$ rádios cognitivos. Além do resultado obtido por cooperação, também são mostradas as curvas de ROC obtidas a partir das medidas locais em cada um dos rádios. O desempenho do k -ésimo rádio está denotado por R_k . Os coeficientes do canal entre cada um deles e o usuário primário estão mostrados na Tabela 5.2. Pelos resultados mostrados na Fig. 5.18, percebemos que há um ganho considerável do sistema cooperativo em relação aos sistemas sem cooperação. Note ainda que o ganho de desempenho se estende para todos os rádios presentes no sistema.

	h_1	h_2	h_3	h_4	h_5
$n = 1$	0,661	0,297	0,728	0,963	0,561
$n = 2$	0,080	0,016	0,924	0,551	0,515
$n = 3$	0,614	0,289	0,771	0,537	0,580

Tabela 5.2: Coeficientes dos canais entre os rádios cognitivos e o usuário primário no exemplo da Fig. 5.18.

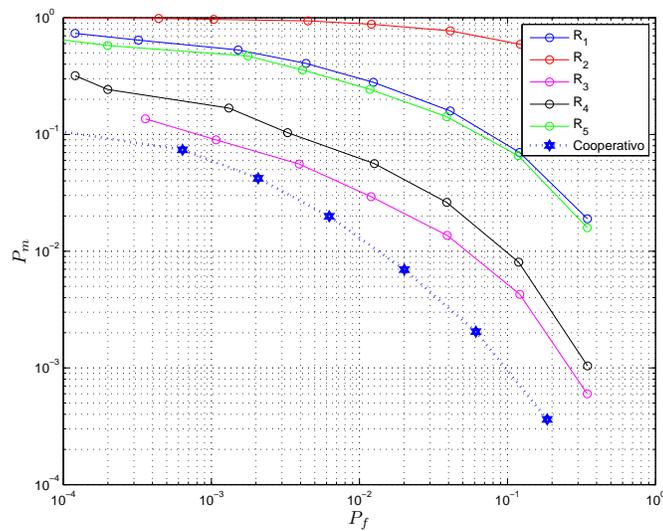


Figura 5.18: Curvas de ROC para o sistema cooperativo em comparação com os rádios cognitivos operando individualmente.

5.3 Detecção de energia usando o método do consenso

A aplicação de algoritmos de consenso em cenário de sensoriamento espectral se deu primeiramente em [67]. Nessa proposta, cada usuário secundário emprega um detector de energia em sua estimação espectral local e se utiliza de um algoritmo de consenso para estimar a medida de energia média na rede de rádio cognitivos. Nesse trabalho, é atribuído o mesmo peso às estimativas de energia locais, independentemente da SNR em cada rádio cognitivo. Em [68], é apresentado um algoritmo em que cada estimativa de energia é ponderada pela SNR em cada um dos rádios cognitivos da rede. Com essa estratégia, os resultados apresentados em [68] superam àqueles mostrados em [67], *i.e.*, para uma dada probabilidade de falso alarme, a probabilidade de perda de detecção é menor se a estratégia proposta em [68] é utilizada. Já em [69] e [70], algoritmos adaptativos são utilizados para estimar os pesos de ponderação de cada estimativa de energia dos nós da rede.

Um problema inerente aos algoritmos de consenso é baixa velocidade de convergência. Esse problema pode ser especialmente crítico no caso do rádio cognitivo, em que o ambiente é altamente não-estacionário e decisões precisam ser tomadas rapidamente.

Nesta Seção, usando um algoritmo de predição associado ao consenso, apresentaremos resultados de sensoriamento espectral distribuído aplicando algoritmos de consenso com velocidade de convergência acelerada.

5.3.1 O algoritmo de consenso

Na Seção 3.2, mostrou-se que, no caso de cooperação centralizada, a estatística de teste ótima é aproximada pelo receptor MRC:

$$T_{MRC}(\mathbf{y}) = \sum_{k=0}^{M-1} \gamma_k y_k, \quad (5.71)$$

em que y_k é a medida de energia fornecida por cada nó da rede e γ_k é a SNR medida em cada nó.

No caso que a rede não conta com um centro de fusão, os nós precisam trocar informação entre si, com o objetivo de estimar $T_{MRC}(\mathbf{y})$. Essa troca de informação pode se dar através do algoritmo do consenso [71].

No algoritmo do consenso, considera-se uma rede com K nós e vetor estado inicial, composto pela medida em cada nó da rede no instante $n = 0$:

$$\mathbf{x}(0) = [x_0(0) \ x_1(0) \ \cdots \ x_{K-1}(0)]^T. \quad (5.72)$$

Em cada instante de tempo n , o k -ésimo nó da rede envia sua medida $x_k(n)$ aos nós em sua vizinhança. A combinação das medidas recebidas pode ser modelada através do seguinte sistema linear:

$$x_k(n+1) = \sum_{m \in \mathcal{N}_k} c_{k,m} x_m(n) \quad (5.73)$$

que pode ser representado de forma matricial como:

$$\mathbf{x}(n+1) = \mathbf{C}\mathbf{x}(n), \quad (5.74)$$

em que \mathbf{x} é o chamado vetor de estados e \mathbf{C} é a matriz de combinação com elementos $c_{k,m}$. Alguns exemplos de possíveis matrizes de combinação são aqueles mostrados no Capítulo 4. Pode-se mostrar que se a matriz \mathbf{C} é uma matriz duplamente estocástica, *i.e.*, $\mathbf{C}\mathbf{1} = \mathbf{1}$ e $\mathbf{1}^T\mathbf{C} = \mathbf{1}^T$ e se a seguinte condição sobre o raio espectral de $\mathbf{C} - \frac{1}{K}\mathbf{1}\mathbf{1}^T$ é satisfeita:

$$\rho\left(\mathbf{C} - \frac{1}{K}\mathbf{1}\mathbf{1}^T\right) < 1 \quad (5.75)$$

as medidas em todos nós convergem para:

$$x_k(\infty) = \frac{1}{K} \sum_{m=0}^{K-1} x_m(0). \quad (5.76)$$

Portanto, se aplicarmos o algoritmo do consenso considerando $x_k(0) = \gamma_k y_k$, após a convergência do algoritmo, a medida em cada nó tenderá a uma versão escalonada da estatística de teste sub-ótimo apresentado no Capítulo 3:

$$x_k(\infty) = \frac{1}{K} \sum_{m=0}^{K-1} \gamma_k y_k. \quad (5.77)$$

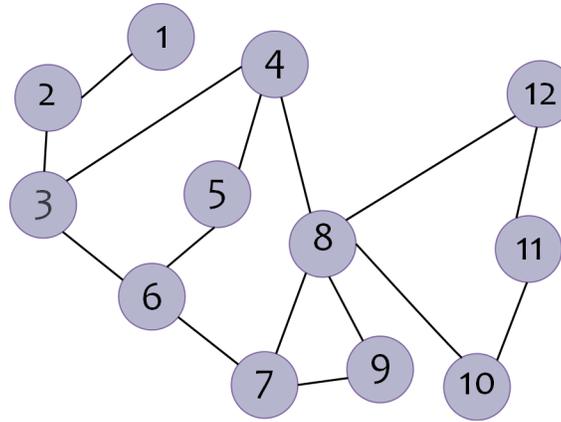


Figura 5.19: Rede de rádios cognitivos

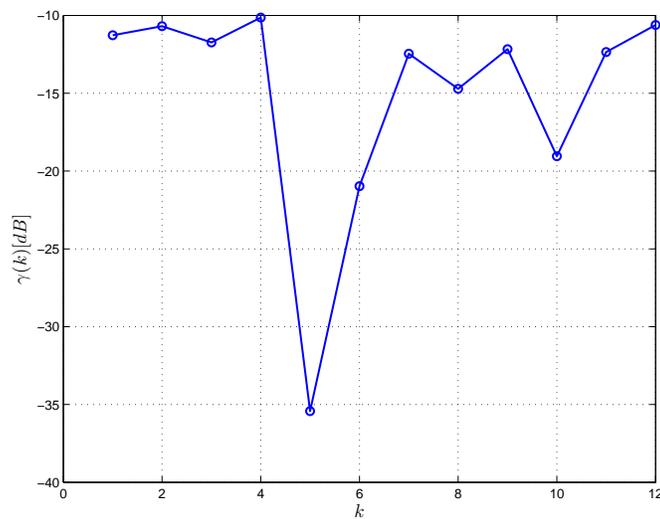


Figura 5.20: SNR em cada um dos nós da rede apresentada na Fig. 5.19

Como exemplo, consideraremos novamente a rede de $K = 12$ nós apresentada na Fig. 5.19, a qual é a mesma utilizando em exemplos do Capítulo 4 e repetida aqui por conveniência, com cada nó da rede apresentando a SNR mostrada na Fig. 5.20. O algoritmo de consenso foi aplicado definindo a matriz \mathbf{C} de acordo com a regra Metrópolis apresentada no Capítulo 4, e a estimativa do teste sub-ótimo por iteração está mostrada na Fig. 5.21. Analisando o resultado, notamos que o algoritmo converge apenas após cerca de 35 iterações.

A baixa velocidade de convergência do algoritmo pode ser contornada utilizando a técnica descrita em [13], onde se adiciona um termo de predição à estimativa atual

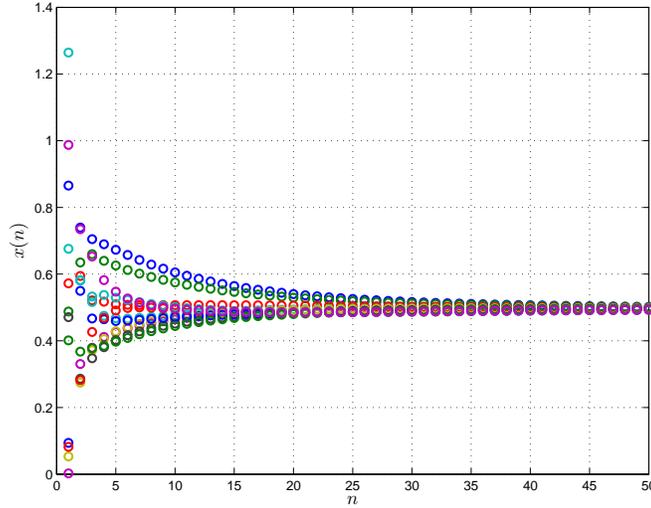


Figura 5.21: Estimativa ponderada de energia do teste sub-ótimo em cada nó da rede apresentada na Fig. 5.19. As estimativa em cada um dos nós está representada com uma cor diferente na figura.

derivada do algoritmo do consenso:

$$x_k(n) = \alpha x_k^p(n) + (1 - \alpha) x_k^C(n), \quad (5.78)$$

em que α é um fator de mistura tal que $0 \leq \alpha < 1$, $x_k^p(n)$ é o termo obtido pela predição de $x_k(n)$ a partir das últimas M_p amostras de $x_k(n)$ e $x_k^C(n)$ é o termo obtido a partir dos estados dos nós da rede vizinhos a k no instante $n - 1$:

$$x_k^C(n) = c_{k,k} x_k(n-1) + \sum_{m \in \mathcal{N}_k} c_{k,m} x_m(n-1), \quad (5.79)$$

$$x_k^p(n) = \theta_{M_p} x_k^C(n) + \sum_{m=1}^{M_p-1} \theta_m x_k^C(n - M_p + m), \quad (5.80)$$

em que θ_m corresponde ao coeficiente de predição relativo ao atraso $M_p - m$. Em [13], é mostrado que o vetor de coeficientes Θ que minimiza o erro de predição em termos de erro quadrático é dado por:

$$\Theta = [\theta_1 \ \theta_2 \ \cdots \ \theta_{M_p}]^T = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{1}, \quad (5.81)$$

em que \mathbf{A} é uma matriz de dimensão $2xM_p$ expressa como:

$$\mathbf{A} = \begin{bmatrix} -M_p + 1 & -M_p + 2 & \cdots & -1 & 0 \\ 1 & 1 & \cdots & 1 & 1 \end{bmatrix}. \quad (5.82)$$

Resta estimar o fator de mistura α . Esse fator pode ser determinado lembrando que algoritmos de consenso têm sua velocidade de convergência associada a [72]:

$$r(\mathbf{C}) \triangleq \rho\left(\mathbf{C} - \frac{1}{K}\mathbf{1}\mathbf{1}^T\right), \quad (5.83)$$

em que $\rho(\cdot)$ é o raio espectral.

Em [13], é mostrado que, no caso em que o consenso é acelerado pela parcela de predição, é possível reescrever o conjunto de equações (5.78), (5.79) e (5.80) através da seguinte forma matricial:

$$\begin{bmatrix} \mathbf{x}(n) \\ \mathbf{x}(n-1) \\ \vdots \\ \mathbf{x}(n-M_p+1) \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{C}_\alpha^{(M_p)} & \alpha\theta_{M_p-2}\mathbf{I}_K & \cdots & \alpha\theta_2\mathbf{I}_K & \alpha\theta_1\mathbf{I}_K \\ \mathbf{I}_K & \mathbf{0}_K & \cdots & \mathbf{0}_K & \mathbf{0}_K \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{0}_K & \mathbf{0}_K & \cdots & \mathbf{I}_K & \mathbf{0}_K \end{bmatrix}}_{\mathbf{C}_\alpha^{est}} \begin{bmatrix} \mathbf{x}(n-1) \\ \mathbf{x}(n-2) \\ \vdots \\ \mathbf{x}(n-M_p+1) \end{bmatrix}, \quad (5.84)$$

em que $\mathbf{0}_K$ é a matriz nula de ordem K e $\mathbf{C}_\alpha^{(M_p)}$ é definida como:

$$\mathbf{C}_\alpha^{(M_p)} \triangleq (1 - \alpha + \alpha\theta_{M_p})\mathbf{C} + \alpha\theta_{M_p}\mathbf{I}_K. \quad (5.85)$$

Expressando os estados do sistema através da matriz estendida mostrada em (5.84), pode-se determinar o fator de mistura que minimiza a velocidade de convergência resolvendo o seguinte problema de otimização:

$$\alpha_{opt} = \underset{\alpha}{\operatorname{argmin}} \rho\left(\mathbf{C}_\alpha^{est} - \frac{1}{K(M_p-1)}\mathbf{1}\mathbf{1}^T\right), \quad (5.86)$$

Ainda em [13], é mostrado que para o caso em que $M_p = 2$, o fator de mistura que minimiza (5.86) é dado por:

$$\alpha_{opt} = \frac{\lambda(1) + \lambda(K-1)}{2 - \lambda(1) - \lambda(K-1)}, \quad (5.87)$$

em que $\lambda(1)$ é o menor autovalor da matriz \mathbf{C}_α^{est} e $\lambda(K-1)$ é o segundo maior autovalor de \mathbf{C}_α^{est} .

Para $M_p > 2$, [13] não obtém expressão fechada e nem fornece sequer esclarecimento sobre como proceder para determinar o fator de mistura α ótimo para esses casos. Assim, é proposta uma solução desta questão através de uma abordagem

semi-analítica que será mostrada a seguir.

Para tanto, a fim de investigar se há uma relação entre $\lambda(K - 1)$ e α_{opt} , fixamos um número de nós, geramos matrizes de adjacência aleatórias e, para cada matriz gerada, definimos a matriz de conectividade através da regra Metrópolis. Posteriormente, para cada matriz de conectividade, variamos o fator de mistura no intervalo $0 \leq \alpha < 1$, determinamos sua respectiva matriz \mathbf{C}_α^{est} e, finalmente, calculamos $r(\mathbf{C}_\alpha^{est})$ como mostrado em (5.83). Nesse primeiro exemplo, consideramos que o termo de predição tem ordem $M_p = 3$ e que a rede tem $K = 12$ nós.

O gráfico de dispersão apresentando a relação entre o segundo maior autovalor da matriz, $\lambda(K - 1)$, e o valor ótimo de α está mostrado na Fig. 5.22. Por esse resultado, notamos que há uma relação aproximadamente logarítmica entre $\lambda(K - 1)$ e α_{opt} , fato que pode ser confirmado na análise da curva de dispersão entre $\lambda(K - 1)$ e $\log(\alpha_{opt})$ também mostrada na Fig. 5.22. Por essa curva de dispersão, concluímos que para uma rede com $K = 12$ nós e aplicando um algoritmo de consenso acelerado com um preditor de ordem $M_p = 3$, uma boa escolha para o parâmetro de mistura é dada pelo ajuste linear também mostrado na Fig. 5.22.

A pergunta que naturalmente surge agora é se tal curva de dispersão depende do número de nós. A fim de responder a essa questão, a Fig. 5.23 mostra os gráficos de dispersão entre $\lambda(K - 1)$ e $\log(\alpha_{opt})$ considerando redes com diferentes números de nós. A partir desse resultado, notamos que o ajuste linear continua válido independentemente do número de nós da rede.

Por fim, resta verificar se a aproximação da curva de dispersão continua podendo ser aproximada da mesma maneira para $M_p > 3$. Para tanto, a Fig. 5.24 mostra as curvas de dispersão para valores mais elevados de M_p . Notamos que a relação entre $\lambda(K - 1)$ e $\log(\alpha_{opt})$ continua praticamente linear. Todavia os coeficientes da reta mudam de acordo com a ordem do filtro preditor.

Com o objetivo de testar a precisão do ajuste linear, adotamos a rede da Fig. 5.19 com matriz de combinação Metrópolis, e calculamos $r(\mathbf{C}_\alpha^{est})$ de acordo com (5.83). A Fig. 5.25 mostra $r(\mathbf{C}_\alpha^{est})$ em função de α para vários valores de M_p . No caso em que $M_p = 2$, verifica-se que o valor ótimo de α realmente coincide com

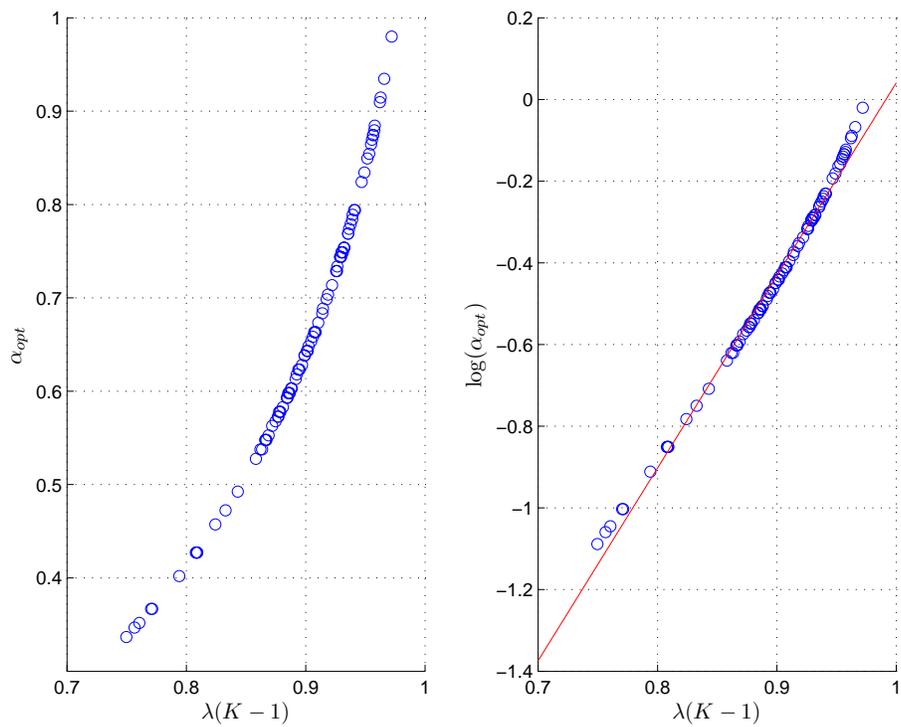


Figura 5.22: Gráfico de dispersão entre $\lambda(K - 1)$ e α_{opt} para $K = 12$. Curva de ajuste dada por: $\alpha_{opt} = \exp(4,7156 \lambda(K - 1) - 4,6749)$.

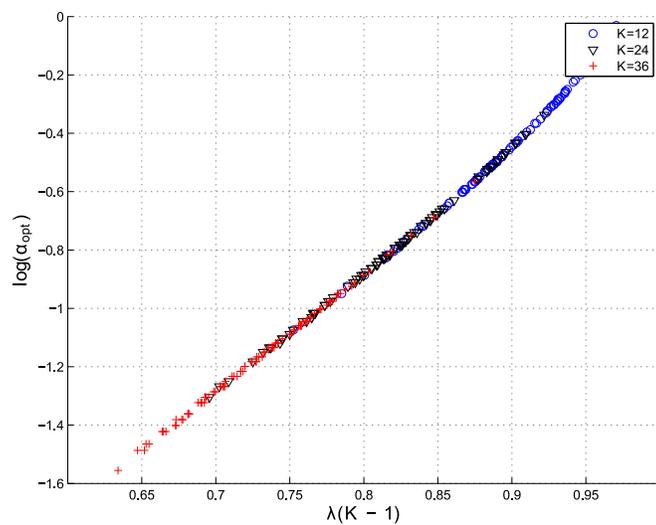


Figura 5.23: Gráfico de dispersão entre $\lambda(K - 1)$ e $\log(\alpha_{opt})$ para $K = 12$, $K = 24$ e $K = 36$.

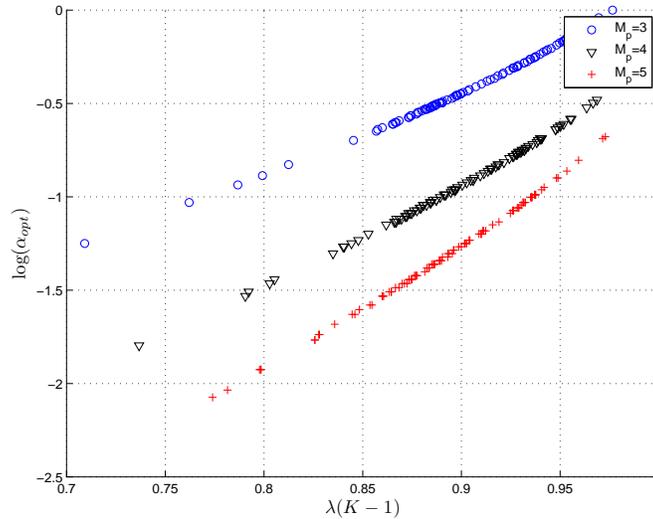


Figura 5.24: Curvas de dispersão entre $\lambda(K-1)$ e $\log(\alpha_{opt})$ para diferentes ordens do filtro de predição e redes com $K = 12$ nós.

o dado em (5.87). Para os demais valores de M_p , as curvas verticais mostram os valores de α_{opt} estimados pelo ajuste linear com coeficientes obtidos pelo método de mínimos quadrados. Percebe-se que as estimativas de α_{opt} são próximas dos valores obtidos experimentalmente.

A fim de observar como a redução do raio espectral de \mathbf{C}_α^{est} reflete na aceleração do algoritmo, voltamos ao cenário de SNR descrito na Fig. 5.20 e usamos o algoritmo para determinar a estatística de teste MRC. A Fig. 5.26 mostra as curvas de convergência considerando $\alpha = 0$, *i.e.*, sem usar o filtro de predição e também considerando um preditor com ordem $M_p = 3$ e parâmetro de mistura $\alpha = 0,8009$. Note que há um ganho de velocidade de convergência considerável, utilizando o preditor com o valor sub-ótimo obtido por ajuste de mínimos quadrados quando comparado com o consenso sem o uso do preditor. Para facilitar a visualização do ganho de desempenho com o algoritmo preditor, as curvas da Fig. 5.27 mostram o valor do estatística de teste MRC apenas no nó 1 da rede. Podemos notar que para a versão acelerada do algoritmo, a convergência se dá com aproximadamente $n = 18$ iterações. Já para a versão sem a predição, é necessário cerca de $n = 45$.

Uma outra importante análise que merece ser feita é como a abordagem de consenso auxiliado por predição impacta nas curvas de ROC do sistema de rádio

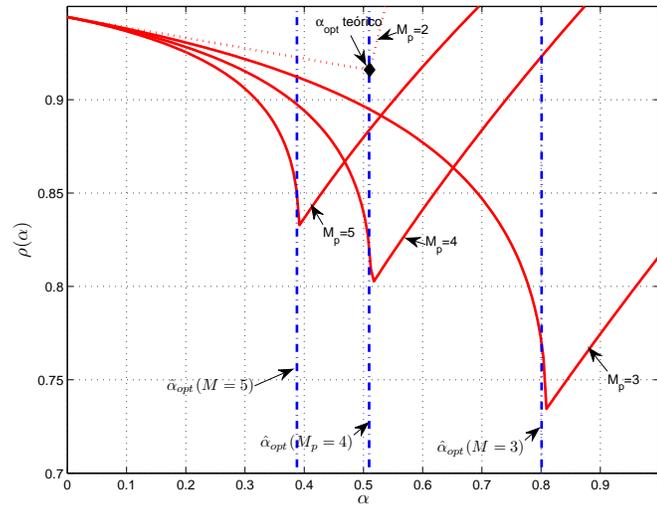


Figura 5.25: Comparação entre as estimativas de α_{opt} obtidas por ajuste linear e os valores ótimos obtidos experimentalmente para a rede apresentada na 5.19. Também é mostrado, na curva pontilhada, o raio espectral para o preditor de ordem $M_p = 2$.

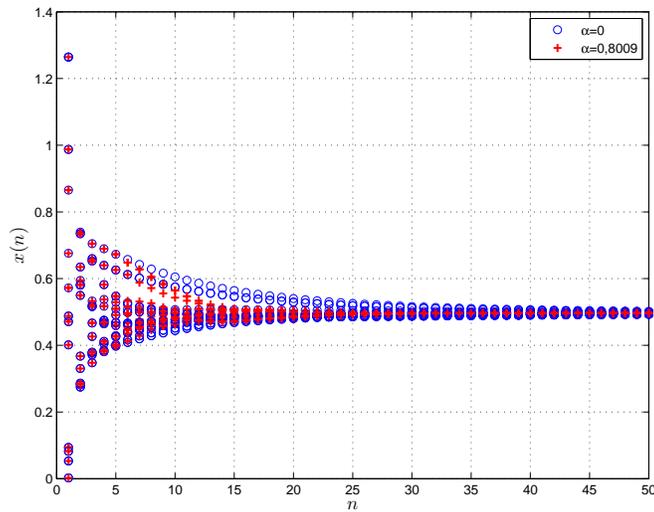


Figura 5.26: Comparação entre o desempenho do algoritmo de consenso e de sua versão com convergência acelerada na rede apresentada na 5.19.

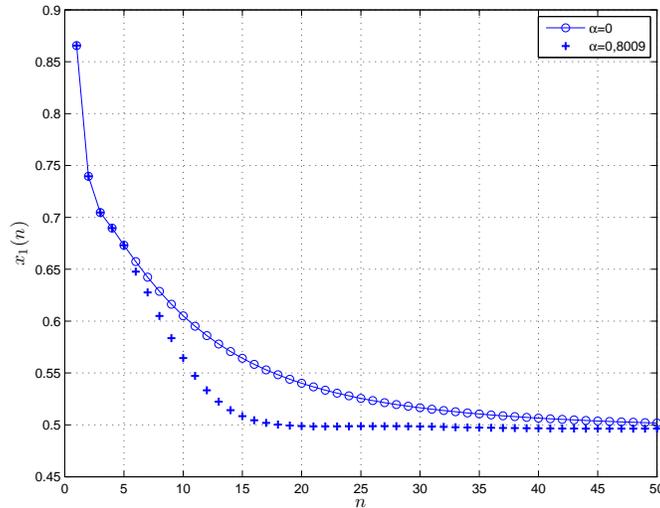


Figura 5.27: Convergência de estimativa de energia ponderada no nó 1 da rede da rede apresentada na Fig. 5.19.

cognitivo. A Fig. 5.28 mostra as curvas de ROC para um sistema de rádio cognitivo por detecção por energia com $M = 8$ amostras e com SNR de cada rádio definida na Fig. 5.20 cuja conectividade se dá como mostrado na Fig. 5.19. Para o algoritmo de consenso com convergência acelerada, utilizamos um preditor de ordem $M_p = 3$ e fator de mistura $\alpha = 0,8009$.

Note que o algoritmo com a velocidade de convergência acelerada apresenta desempenho equivalente ao MRC a partir de $n = 20$ amostras com as estimativas de energia. Já para a versão sem o fator de predição, é necessário esperar cerca de $n = 40$ iterações.

5.4 Conclusões

Neste capítulo, foram analisadas técnicas de detecção distribuída aplicadas em diferentes cenários. No primeiro caso, considerou-se que havia conhecimento de canal e da sequência transmitida pelo usuário primeiro, assim como feito em [66] .

No segundo caso, foi considerado que a estimação espectral é realizada a partir de uma sequência de treinamento enviada pelo usuário primário. Nesse contexto, há conhecimento da sequência enviada, porém não há conhecimento perfeito do canal. Esse problema foi abordado utilizando um filtro LMS como estimador de canal para

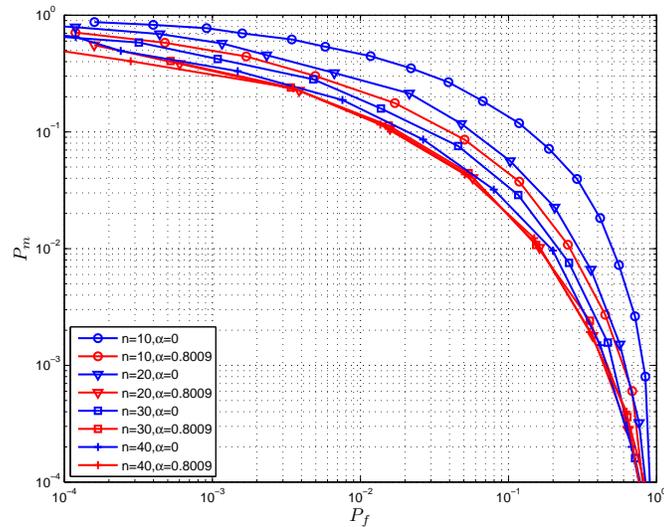


Figura 5.28: ROC do sistema de rádio cognitivo distribuído com rede apresentada na Fig. 5.19.

gerar a estatística de teste de detecção. Foi feita uma análise teórica e os resultados foram confrontados com simulação de Monte Carlo. Mostrou-se que os resultados se aproximam dos gerados pelo detector ML ótimo a um custo computacional menos elevado.

Já na seção 5.3 foi analisado o caso mais geral no qual não há qualquer conhecimento *a priori* do sinal enviado pelo usuário primário. Esse é o caso mais geral e que permite, portanto, uma maior flexibilidade de implementação. Nesse contexto, o resultado de [13] foi estendido para a estimação dos parâmetros de predição para ordens mais elevados. Dessa maneira, foi possível aumentar a velocidade de convergência do algoritmo de consenso.

Capítulo 6

Conclusões e Perspectivas

Ao longo do trabalho de Doutorado, vários aspectos da questão do sensoriamento espectral foram abordados. Foram inicialmente mostradas no Capítulo 2 as limitações de desempenho dos detectores utilizados para sensoriamento espectral e algumas alternativas para alcançar um desempenho de detecção mais confiável.

No Capítulo 3 foram analisados sistemas de sensoriamento espectral com decisão centralizada. Para o caso de troca de informação do tipo dura, foram definidos limitantes de desempenho ainda não mostrados pela literatura. Assumindo uma abordagem Bayesiana, foram analisados os desempenhos das diferentes regras de fusão de dados na presença de canal imperfeito entre os rádios cognitivos e o centro de fusão. Os resultados dessa análise foram publicados em [10]. Ainda nesse contexto, colocamos como uma possível extensão desse trabalho a determinação de limitantes também no caso em que se assume que o centro de fusão processa apenas as informações enviadas pelos usuários secundários e recebidas corretamente. Desse modo, diferentemente do caso já analisado em que se considera o canal entre os rádios cognitivos e o centro de fusão modelado como um BSC, o sistema passaria a ser modelado como um sistema de sensoriamento espectral cooperativo com um número variante de usuários secundários.

Um outro fator que merece destaque na Tese é a análise feita pra o caso em que os rádios cognitivos enviam informação suave ao centro de fusão. Essa análise difere da abordagem que assume gaussianidade no conjunto de dados recebidos no centro de fusão e chega à interessante conclusão que o detector ótimo no critério de máxima

de verossimilhança é aproximado pela técnica de fusão MRC.

O estudo da aplicação dos algoritmos de filtragem adaptativa para a questão de estimação distribuída também gerou algumas contribuições, como mostrado no Capítulo 4. Nesse capítulo, foi mostrado um método simples para a aplicação de técnicas já bem estabelecidas de filtragem adaptativa distribuída no contexto de redes heterogêneas. É importante salientar que os resultados apresentados nesse Capítulo não são necessariamente aplicados a redes de rádios de cognitivos no contexto de sensoriamento espectral. A análise se aplica à estimação de parâmetros em uma rede de sensores qualquer.

Já no Capítulo 5, foram apresentadas técnicas de detecção distribuída fundamentadas na questão de estimação distribuída focada no Capítulo 4. Em especial, na Seção 5.2, foi realizada uma análise estatística do algoritmo LMS e foram gerados resultados promissores no campo de sensoriamento espectral cooperativo no cenário em que não há conhecimento do canal entre os usuários primários e secundários. Dadas as características não-estacionárias dos sistemas de sensoriamento espectral, esse é um contexto importante de análise, uma vez que a estimação de canal nesse cenário pode gerar latências impeditivas para o sensoriamento espectral.

Um outro ponto que merece destaque é a questão do sensoriamento espectral no qual não há qualquer conhecimento *a priori* do sinal transmitido pelo usuário primário. Esse é o caso mais geral e, certamente, merece uma análise detalhada. As contribuições nesse contexto, discutidas na Seção 5.3, foram de caráter incremental e mostraram como estimar os parâmetros ótimos que permitem acelerar o consenso nessa situação.

Como perspectivas futuras colocamos inserir cenários de sensoriamento espectral mais realistas nos quais o sincronismo entre os usuários secundários e o centro de fusão não é perfeito [73] e ainda situações em que os sistemas cooperativos ainda contam com usuários maliciosos que degradam o desempenho do sistema [74]. Há ainda alguns resultados apresentados na Tese de Doutorado que merecem estudos mais aprofundados. Dentre eles, citamos a análise do sistema centralizado com troca de informação do tipo suave detalhada na Seção 3.2, a análise do caso de

estimação distribuída utilizando o algoritmo de difusão no contexto em que a rede não é completamente conectada, a aplicação do algoritmo LMS para sensoriamento espectral cooperativo sem o conhecimento de canal, a análise estatística apresentada na Seção 5.2 e, por fim, a estimação dos parâmetros ótimos do consenso acelerado para ordens de predição mais elevadas mostrada na Seção 5.3.

Referências Bibliográficas

- [1] S.W. Ellingson, “Spectral occupancy at VHF: implications for frequency-agile cognitive radios,” in *Vehicular Technology Conference, 2005. VTC-2005-Fall. 2005 IEEE 62nd*, sept., 2005, vol. 2, pp. 1379 – 1382.
- [2] Joseph Mitola III, *Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio*, Ph.D. thesis, Royal Institute of Technology (KTH), Stockholm, Sweden, May 2000.
- [3] III Mitola, J. and Jr. Maguire, G.Q., “Cognitive radio: making software radios more personal,” *Personal Communications, IEEE*, vol. 6, no. 4, pp. 13 –18, aug 1999.
- [4] K. Ben Letaief and Wei Zhang, “Cooperative communications for cognitive radio networks,” *Proceedings of the IEEE*, vol. 97, no. 5, pp. 878 –893, may 2009.
- [5] M. Gandetto and C. Regazzoni, “Spectrum sensing: A distributed approach for cognitive terminals,” *Selected Areas in Communications, IEEE Journal on*, vol. 25, no. 3, pp. 546 –557, april 2007.
- [6] A.G. Fragkiadakis, E.Z. Tragos, and I.G. Askoxylakis, “A survey on security threats and detection techniques in cognitive radio networks,” *Communications Surveys Tutorials, IEEE*, vol. 15, no. 1, pp. 428–445, First 2013.
- [7] C. Stevenson, G. Chouinard, Zhongding Lei, Wendong Hu, S.J. Shellhammer, and W. Caldwell, “IEEE 802.22: The first cognitive radio wireless regional area network standard,” *Communications Magazine, IEEE*, vol. 47, no. 1, pp. 130–138, January 2009.

-
- [8] S.H. Song, K. Hamdi, and K.B. Letaief, “Spectrum sensing with active cognitive systems,” *Wireless Communications, IEEE Transactions on*, vol. 9, no. 6, pp. 1849–1854, June 2010.
- [9] D. Cavalcanti and M. Ghosh, “Cognitive radio networks: Enabling new wireless broadband opportunities,” in *Cognitive Radio Oriented Wireless Networks and Communications, 2008. CrownCom 2008. 3rd International Conference on*, 15–17 2008, pp. 1–6.
- [10] A. de Paula and C. Panazio, “Cooperative spectrum sensing under unreliable reporting channels,” *Wireless Networks*, pp. 1–9, 2014.
- [11] A. de Paula and C. Panazio, “Analysis of distributed parameter estimation in wsn with unreliable nodes,” in *Wireless Communication Systems (ISWCS), 2012 International Symposium on*, Aug 2012, pp. 116–120.
- [12] A. de Paula and C. Panazio, “Diffusion LMS networks in the presence of noisy nodes: A convergence rate and msd analysis,” in *Machine Learning for Signal Processing (MLSP), 2013 IEEE International Workshop on*, Sept 2013, pp. 1–6.
- [13] T.C. Aysal, B.N. Oreshkin, and M.J. Coates, “Accelerated distributed average consensus via localized node state prediction,” *Signal Processing, IEEE Transactions on*, vol. 57, no. 4, pp. 1563–1576, April 2009.
- [14] S. M. Kay, *Fundamentals of Statical Signal Processing, Volume II: Detection Theory*, Prentice Hall, 1998.
- [15] R. Tandra and A. Sahai, “SNR walls for signal detection,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 2, no. 1, pp. 4–17, Feb 2008.
- [16] J. Proakis and M. Salehi, *Digital Communications*, Mc Graw-Hill, 5 edition, 2008.
- [17] Pramod K. Varshney, *Distributed Detection and Data Fusion*, Springer, 1996.

-
- [18] F. F. Digham, M.-S. Alouini, and M. K. Simon, “On the energy detection of unknown signals over fading channels,” *Communications, IEEE Transactions on*, vol. 55, no. 1, pp. 21–24, 2007.
- [19] F.F. Digham, M.-S. Alouini, and Marvin K. Simon, “On the energy detection of unknown signals over fading channels,” in *Communications, 2003. ICC '03. IEEE International Conference on*, May 2003, vol. 5, pp. 3575–3579 vol.5.
- [20] Zhi Quan, Shuguang Cui, H. Poor, and A. Sayed, “Collaborative wideband sensing for cognitive radios,” *Signal Processing Magazine, IEEE*, vol. 25, no. 6, pp. 60–73, november 2008.
- [21] T. Yucek and H. Arslan, “A survey of spectrum sensing algorithms for cognitive radio applications,” *Communications Surveys Tutorials, IEEE*, vol. 11, no. 1, pp. 116–130, quarter 2009.
- [22] S. Haykin, D.J. Thomson, and J.H. Reed, “Spectrum sensing for cognitive radio,” *Proceedings of the IEEE*, vol. 97, no. 5, pp. 849–877, may 2009.
- [23] D. Bhargavi and C.R. Murthy, “Performance comparison of energy, matched-filter and cyclostationarity-based spectrum sensing,” in *Signal Processing Advances in Wireless Communications (SPAWC), 2010 IEEE Eleventh International Workshop on*, june 2010, pp. 1–5.
- [24] J. Lunden, V. Koivunen, A. Huttunen, and H.V. Poor, “Collaborative cyclostationary spectrum sensing for cognitive radio systems,” *Signal Processing, IEEE Transactions on*, vol. 57, no. 11, pp. 4182–4195, nov. 2009.
- [25] Jarmo Lunden, Visa Koivunen, Anu Huttunen, and H. Vincent Poor, “Spectrum sensing in cognitive radios based on multiple cyclic frequencies,” in *Cognitive Radio Oriented Wireless Networks and Communications, 2007. CrownCom 2007. 2nd International Conference on*, aug. 2007, pp. 37–43.
- [26] Hou-Shin Chen, Wen Gao, and D.G. Daut, “Spectrum sensing using cyclostationary properties and application to IEEE 802.22 WRAN,” in *Global Te-*

- lecommunications Conference, 2007. GLOBECOM '07. IEEE*, Nov 2007, pp. 3133–3138.
- [27] W.A. Gardner, “Exploitation of spectral redundancy in cyclostationary signals,” *Signal Processing Magazine, IEEE*, vol. 8, no. 2, pp. 14–36, April 1991.
- [28] William A. Gardner, Antonio Napolitano, and Luigi Paura, “Cyclostationarity: Half a century of research,” *Signal Processing*, vol. 86, no. 4, pp. 639 – 697, 2006.
- [29] R.S. Roberts, W.A. Brown, and Jr. Loomis, H.H., “Computationally efficient algorithms for cyclic spectral analysis,” *Signal Processing Magazine, IEEE*, vol. 8, no. 2, pp. 38 –49, apr 1991.
- [30] Yonghong Zeng and Ying chang Liang, “Eigenvalue-based spectrum sensing algorithms for cognitive radio,” *Communications, IEEE Transactions on*, vol. 57, no. 6, pp. 1784 –1793, june 2009.
- [31] A. Kortun, T. Ratnarajah, M. Sellathurai, Caijun Zhong, and C.B. Papadias, “On the performance of eigenvalue-based cooperative spectrum sensing for cognitive radio,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 1, pp. 49 –55, feb. 2011.
- [32] F. Penna, R. Garello, and M. Spirito, “Cooperative spectrum sensing based on the limiting eigenvalue ratio distribution in wishart matrices,” *Communications Letters, IEEE*, vol. 13, no. 7, pp. 507 –509, july 2009.
- [33] P. Bianchi, M. Debbah, M. Maida, and J. Najim, “Performance of statistical tests for single-source detection using random matrix theory,” *Information Theory, IEEE Transactions on*, vol. 57, no. 4, pp. 2400 –2419, april 2011.
- [34] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*, Now Publisher Inc., 2004.
- [35] R. A. Fisher, “The sampling distribution of some statistics obtained from non-linear equations,” *Annals of Eugenics*, vol. 9, pp. 238–249, 1939.

- [36] P. L. Hsu, “On the distribution of roots of certain determinantal equations,” *Annals of Eugenics*, vol. 9, pp. 250–258, 1939.
- [37] S. N. Roy, “p-statistics or some generalizations in the analysis of variance appropriate to multivariate problems,” *Sankhya: Indian Journal of Statistics*, vol. 4, pp. 381–396, 1939.
- [38] Lu Wei and O. Tirkkonen, “Analysis of scaled largest eigenvalue based detection for spectrum sensing,” in *Communications (ICC), 2011 IEEE International Conference on*, june 2011, pp. 1–5.
- [39] Minho Kim and Chong-Ho Choi, “Hidden-node detection in IEEE 802.11n wireless LANs,” *Vehicular Technology, IEEE Transactions on*, vol. 62, no. 6, pp. 2724–2734, July 2013.
- [40] Wei Zhang, R.K. Mallik, and K. Ben Letaief, “Cooperative spectrum sensing optimization in cognitive radio networks,” in *Communications, 2008. ICC '08. IEEE International Conference on*, 19-23 2008, pp. 3411–3415.
- [41] Wei Zhang and Khaled Letaief, “Cooperative spectrum sensing with transmit and relay diversity in cognitive radio networks - [transaction letters],” *Wireless Communications, IEEE Transactions on*, vol. 7, no. 12, pp. 4761–4766, december 2008.
- [42] A. Ghasemi and E.S. Sousa, “Collaborative spectrum sensing for opportunistic access in fading environments,” in *New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005. 2005 First IEEE International Symposium on*, nov. 2005, pp. 131–136.
- [43] Amir Ghasemi and Elvino S. Sousa, “Opportunistic spectrum access in fading channels through collaborative sensing,” *Journal of Communications*, vol. 2, no. 2, pp. 71–82, 2007.

-
- [44] S. Chaudhari, J. Lunden, V. Koivunen, and H.V. Poor, “Cooperative sensing with imperfect reporting channels: Hard decisions or soft decisions?,” *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 18–28, Jan. 2012.
- [45] Xiangwei Zhou, J. Ma, G. Li, Young Kwon, and A.C.K. Soong, “Probability-based combination for cooperative spectrum sensing,” *Communications, IEEE Transactions on*, vol. 58, no. 2, pp. 463–466, february 2010.
- [46] Wei Zhang, R. Mallik, and K. Letaief, “Optimization of cooperative spectrum sensing with energy detection in cognitive radio networks,” *Wireless Communications, IEEE Transactions on*, vol. 8, no. 12, pp. 5761–5766, december 2009.
- [47] Jun Ma, Guodong Zhao, and Ye Li, “Soft combination and detection for cooperative spectrum sensing in cognitive radio networks,” *Wireless Communications, IEEE Transactions on*, vol. 7, no. 11, pp. 4502–4507, november 2008.
- [48] Zhi Quan, Shuguang Cui, and A.H. Sayed, “Optimal linear cooperation for spectrum sensing in cognitive radio networks,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 2, no. 1, pp. 28–40, feb. 2008.
- [49] R. Viswanathan and P.K. Varshney, “Distributed detection with multiple sensors i. fundamentals,” *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, jan 1997.
- [50] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*, Courier Dover Publications, April 2012.
- [51] C.G. Lopes and A.H. Sayed, “Diffusion least-mean squares over adaptive networks: Formulation and performance analysis,” *Signal Processing, IEEE Transactions on*, vol. 56, no. 7, pp. 3122–3136, july 2008.
- [52] F.S. Cattivelli and A.H. Sayed, “Diffusion LMS strategies for distributed estimation,” *Signal Processing, IEEE Transactions on*, vol. 58, no. 3, pp. 1035–1048, march 2010.

-
- [53] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 4 edition, 2002.
- [54] A. H. Sayed, *Fundamentals of Adaptive Filtering*, Jonh Wiley, 3 edition, 2003.
- [55] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, Springer, 2008.
- [56] I.D. Schizas, G. Mateos, and G.B. Giannakis, “Distributed LMS for consensus-based in-network adaptive processing,” *Signal Processing, IEEE Transactions on*, vol. 57, no. 6, pp. 2365–2382, June 2009.
- [57] G. Mateos, I.D. Schizas, and G.B. Giannakis, “Closed-form mse performance of the distributed LMS algorithm,” in *Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop, 2009. DSP/SPE 2009. IEEE 13th*, Jan 2009, pp. 66–71.
- [58] Sheng-Yuan Tu and A.H. Sayed, “Diffusion networks outperform consensus networks,” in *Statistical Signal Processing Workshop (SSP), 2012 IEEE*, Aug 2012, pp. 313–316.
- [59] Sheng-Yuan Tu and A.H. Sayed, “Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks,” *Signal Processing, IEEE Transactions on*, vol. 60, no. 12, pp. 6217–6234, Dec 2012.
- [60] C.G. Lopes and A.H. Sayed, “Diffusion least-mean squares over adaptive networks,” in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, april 2007, vol. 3, pp. III–917 –III–920.
- [61] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra Book*, SIAM, 2001.
- [62] A. Khalili, A. Rastegarnia, J. A. Chambers, and W. M. Bazzi, “An optimum step-size assignment for incremental LMS adaptive networks based on average convergence rate constraint,” *AEU - International Journal of Electronics and Communications*, vol. 67, no. 3, pp. 263 – 268, 2013.

- [63] S. Gollamudi, S. Nagaraj, S. Kapoor, and Yih-Fang Huang, “Set-membership filtering and a set-membership normalized lms algorithm with an adaptive step size,” *IEEE Signal Processing Letters*, vol. 5, no. 5, pp. 111–114, May 1998.
- [64] M.V.S. Lima and P.S.R. Diniz, “On the steady-state mse performance of the set-membership nlms algorithm,” in *7th International Symposium on Wireless Communication Systems (ISWCS), 2010*, Sept. 2010, pp. 389–393.
- [65] S. Werner, Yih-Fang Huang, M.L.R. de Campos, and V. Koivunen, “Distributed parameter estimation with selective cooperation,” in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. ICASSP 2009.*, April 2009, pp. 2849–2852.
- [66] F.S. Cattivelli and A.H. Sayed, “Distributed detection over adaptive networks using diffusion adaptation,” *Signal Processing, IEEE Transactions on*, vol. 59, no. 5, pp. 1917–1932, may 2011.
- [67] Zhiqiang Li, F.R. Yu, and Minyi Huang, “A distributed consensus-based cooperative spectrum-sensing scheme in cognitive radios,” *Vehicular Technology, IEEE Transactions on*, vol. 59, no. 1, pp. 383–393, 2010.
- [68] Wenlin Zhang, Zheng Wang, Yi Guo, Hongbo Liu, Yingying Chen, and J. Mitola, “Distributed cooperative spectrum sensing based on weighted average consensus,” in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, 2011, pp. 1–6.
- [69] F.C. Ribeiro, M. L R De Campos, and S. Werner, “Distributed cooperative spectrum sensing with adaptive combining,” in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, March 2012, pp. 3557–3560.
- [70] F.C. Ribeiro, M. L R De Campos, and S. Werner, “Distributed cooperative spectrum sensing with selective updating,” in *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, Aug 2012, pp. 474–478.

-
- [71] R. Olfati-Saber, J.A. Fax, and R.M. Murray, “Consensus and cooperation in networked multi-agent systems,” *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, 2007.
- [72] Lin Xiao, Stephen Boyd, and Seung jean Kim, “Distributed average consensus with least-mean-square deviation,” *Journal of Parallel and Distributed Computing*, vol. 67, pp. 33–46, 2005.
- [73] Chengqi Song and Qian Zhang, “Sliding-window algorithm for asynchronous cooperative sensing in wireless cognitive networks,” in *Communications, 2008. ICC '08. IEEE International Conference on*, may 2008, pp. 3432–3436.
- [74] P. Kaligineedi, M. Khabbajian, and V.K. Bhargava, “Malicious user detection in a cognitive radio cooperative sensing system,” *Wireless Communications, IEEE Transactions on*, vol. 9, no. 8, pp. 2488–2497, August 2010.

Apêndice A

Demonstração da expressão para n_{opt}

Considerando o Risco de Bayes como uma função do parâmetro n , tem-se:

$$\mathcal{R}(n) = Q_f(n) \frac{\alpha}{\alpha + \beta} C_f + Q_m(n) \frac{\beta}{\alpha + \beta} C_m. \quad (\text{A.1})$$

Daí, substituindo (3.1) e (3.2) em (A.1), obtemos:

$$\mathcal{R}(n) = \sum_{i=0}^{K-n} \binom{K}{i} \left\{ \frac{\alpha}{\alpha + \beta} C_f [(1 - P_f)(1 - P_e) + P_f P_e]^i [P_f(1 - P_e) + (1 - P_f)P_e]^{K-i} + \frac{\beta}{\alpha + \beta} C_m [P_m(1 - P_e) + (1 - P_m)P_e]^i [(1 - P_m)(1 - P_e) + P_m P_e]^{K-i} \right\}. \quad (\text{A.2})$$

Queremos encontrar n que minimize (A.1). Para tanto é necessário analisar a derivada de (A.1) em relação a n :

$$\frac{\partial}{\partial n} \mathcal{R}(n) \approx \mathcal{R}(n+1) - \mathcal{R}(n). \quad (\text{A.3})$$

Logo:

$$\mathcal{R}(n) = -\binom{K}{K-n} \left\{ [(1 - P_f)(1 - P_e) + P_f P_e]^{K-n} [P_f(1 - P_e) + (1 - P_f)P_e]^n + [P_m(1 - P_e) + (1 - P_m)P_e]^{K-n} [(1 - P_m)(1 - P_e) + P_m P_e]^n \right\}. \quad (\text{A.4})$$

Impondo $\frac{\partial}{\partial n} \mathcal{R}(n) = 0$, tem-se:

$$\frac{\alpha}{\alpha + \beta} C_f \left[\frac{(1 - P_f)(1 - P_e) + P_f P_e}{P_m(1 - P_e) + P_e(1 - P_m)} \right]^{K-n} = \frac{\beta}{\alpha + \beta} C_m \left[\frac{(1 - P_m)(1 - P_e) + P_m P_e}{P_f(1 - P_e) + P_e(1 - P_f)} \right]^n. \quad (\text{A.5})$$

Finalmente:

$$n = \left\lceil \frac{K - \phi}{1 + \theta} \right\rceil, \quad (\text{A.6})$$

com θ e ϕ dados por (3.32) e (3.33), respectivamente.

Apêndice B

Demonstração da aproximação para o MSD

B.1 Algoritmo LMS-CTA

Para obter a expressão teórica para o MSD, pode-se reescrever a equação de atualização do erro de estimação como:

$$\tilde{\boldsymbol{\psi}}(n) = (\mathbf{I}_{MK} - \mathbf{D}\mathbf{u}(n-1)\mathbf{u}^T(n-1)) \mathbf{G}\tilde{\boldsymbol{\psi}}(n-1) - \mathbf{D}\boldsymbol{\eta}(n-1)\mathbf{u}(n-1), \quad (\text{B.1})$$

em que $\boldsymbol{\eta}(n)$ é um vetor coluna contendo os vetores de ruído de cada nó da rede, definido como:

$$\boldsymbol{\eta}(n) = \left[\eta_1(n)\mathbf{1}_M^T \quad \eta_2(n)\mathbf{1}_M^T \quad \cdots \quad \eta_K(n)\mathbf{1}_M^T \right]^T \quad (\text{B.2})$$

e $\mathbf{u}(n)$ é um vetor coluna contendo os elementos do vetor regressor em cada nó da rede:

$$\mathbf{u}(n) = \left[\mathbf{u}_1^T(n) \quad \mathbf{u}_2^T(n) \quad \cdots \quad \mathbf{u}_K^T(n) \right]^T. \quad (\text{B.3})$$

Assumindo que o ruído tem média zero e é independente da entrada $\mathbf{u}(n)$, a seguinte equação é válida:

$$\mathbf{E} \left\{ \tilde{\boldsymbol{\psi}}(n)\tilde{\boldsymbol{\psi}}^T(n) \right\} = \mathbf{A}\mathbf{E} \left\{ \tilde{\boldsymbol{\psi}}(n-1)\tilde{\boldsymbol{\psi}}^T(n-1) \right\} \mathbf{A}^T + \mathbf{D}^2\mathbf{Q}\mathbf{R}_u, \quad (\text{B.4})$$

em que $\mathbf{A} = (\mathbf{I}_{MK} - \mathbf{D}\mathbf{R}_u)\mathbf{G}$, $\mathbf{Q} = \boldsymbol{\sigma}_\eta^2 \otimes \mathbf{I}_M$ e $\boldsymbol{\sigma}_\eta^2$ é um vetor coluna contendo as potências de ruído em cada nó:

$$\boldsymbol{\sigma}_\eta^2 = [\sigma_{\eta,1}^2 \quad \sigma_{\eta,2}^2 \quad \cdots \quad \sigma_{\eta,K}^2]^T. \quad (\text{B.5})$$

Em regime, pode-se afirmar que $\mathbf{E} \left\{ \tilde{\boldsymbol{\psi}}(n)\tilde{\boldsymbol{\psi}}^T(n) \right\} = \mathbf{E} \left\{ \tilde{\boldsymbol{\psi}}(n-1)\tilde{\boldsymbol{\psi}}^T(n-1) \right\}$. Definindo $\mathbf{X} = \mathbf{E} \left\{ \tilde{\boldsymbol{\psi}}(n)\tilde{\boldsymbol{\psi}}^T(n) \right\}$, conclui-se, portanto, que o MSD será dado pelo traço de \mathbf{X} , o qual é a solução da seguinte equação discreta de Lyapunov:

$$\mathbf{X} = \mathbf{A}\mathbf{X}\mathbf{A}^T + \mathbf{D}^2\mathbf{Q}\mathbf{R}_u. \quad (\text{B.6})$$

Aproximando a matriz \mathbf{A} por $(\mathbf{I}_{MK} - [\bar{\mu}\mathbf{I}_K \otimes \mathbf{R}])\mathbf{G}$, temos que \mathbf{A} é aproximada por

uma matriz positiva definida, de modo que é possível afirmar:

$$\mathbf{A} \approx \mathbf{U}\mathbf{\Lambda}_A\mathbf{U}^T, \quad (\text{B.7})$$

em que $\mathbf{\Lambda}_A$ é uma matriz diagonal cujos elementos são os valores singulares de \mathbf{A} . Substituindo a decomposição (B.7) em (B.6):

$$\mathbf{X} = \mathbf{U}\mathbf{\Lambda}_A\mathbf{U}^T\mathbf{X}\mathbf{U}\mathbf{\Lambda}_A\mathbf{U}^T + \mathbf{D}^2\mathbf{Q}\mathbf{R}_u. \quad (\text{B.8})$$

Ou ainda:

$$\mathbf{U}^H\mathbf{X}\mathbf{U} = \mathbf{\Lambda}_A\mathbf{U}^H\mathbf{X}\mathbf{U}\mathbf{\Lambda}_A + \mathbf{U}^H\mathbf{D}^2\mathbf{Q}\mathbf{R}_u\mathbf{U}. \quad (\text{B.9})$$

Nesse ponto, é introduzida a seguinte variável auxiliar:

$$\mathbf{Y} = \mathbf{U}^H\mathbf{X}\mathbf{U}. \quad (\text{B.10})$$

Observa-se que os traços das matrizes \mathbf{X} e \mathbf{Y} são iguais:

$$\text{Tr}\{\mathbf{Y}\} = \text{Tr}\{\mathbf{U}^H\mathbf{X}\mathbf{U}\} = \text{Tr}\{\mathbf{X}\mathbf{U}\mathbf{U}^H\} = \text{Tr}\{\mathbf{X}\}. \quad (\text{B.11})$$

Escrevendo (B.9) em termos da nova variável \mathbf{Y} , se tem:

$$\mathbf{Y} = \mathbf{\Lambda}_A\mathbf{Y}\mathbf{\Lambda}_A + \mathbf{U}^H\mathbf{D}^2\mathbf{Q}\mathbf{R}_u\mathbf{U}. \quad (\text{B.12})$$

Como $\mathbf{\Lambda}_A$ é uma matriz diagonal, os elementos diagonais de \mathbf{Y} são dados por:

$$Y_{k,k} = \frac{\mathbf{U}^H\mathbf{D}^2\mathbf{Q}\mathbf{R}_u\mathbf{U}|_{k,k}}{1 - \lambda_A(k)^2}, \quad (\text{B.13})$$

em que $\mathbf{U}^H\mathbf{D}^2\mathbf{Q}\mathbf{R}_u\mathbf{U}|_{k,k}$ é o k -ésimo elemento diagonal da matriz $\mathbf{D}^2\mathbf{Q}\mathbf{R}_u\mathbf{U}$ e $\lambda_A(k)$ é o k -ésimo autovalor da matriz \mathbf{A} . Para simplificar a notação, define-se $b(k) = \mathbf{U}^H\mathbf{D}^2\mathbf{Q}\mathbf{R}_u\mathbf{U}|_{k,k}$.

Como consequência direta:

$$\text{Tr}\{\mathbf{X}\} = \text{Tr}\{\mathbf{Y}\} = \sum_{k=0}^{KM-1} \frac{b(k)}{1 - \lambda_A^2(k)}. \quad (\text{B.14})$$

É interessante analisar os elementos $b(k)$. Para tanto, vamos determinar os autovetores de \mathbf{A} que compõem \mathbf{U} . Lembrando que assumimos a seguinte aproximação:

$$\mathbf{A} \approx (\mathbf{I}_{MK} - [\bar{\mu} \mathbf{I}_K \otimes \mathbf{R}]) \mathbf{G} = \mathbf{C} \otimes (\mathbf{I}_M - \bar{\mu} \mathbf{R}), \quad (\text{B.15})$$

e expressando as matrizes \mathbf{C} e \mathbf{R} através de suas respectivas decomposições em autovalores, temos:

$$\mathbf{C} = \mathbf{U}_C\mathbf{\Lambda}_C\mathbf{U}_C^T, \quad (\text{B.16})$$

$$\mathbf{R} = \mathbf{U}_R\mathbf{\Lambda}_R\mathbf{U}_R^T. \quad (\text{B.17})$$

Assim:

$$\mathbf{A} \approx (\mathbf{U}_C\mathbf{\Lambda}_C\mathbf{U}_C^T) \otimes (\mathbf{U}_R(\mathbf{I}_M - \bar{\mu}\mathbf{\Lambda}_R)\mathbf{U}_R^T) = (\mathbf{U}_C \otimes \mathbf{U}_R)(\mathbf{\Lambda}_C \otimes (\mathbf{I}_M - \bar{\mu}\mathbf{\Lambda}_R))(\mathbf{U}_C^T \otimes \mathbf{U}_R^T). \quad (\text{B.18})$$

Logo:

$$\mathbf{A} \approx \underbrace{(\mathbf{U}_C \otimes \mathbf{U}_R)}_{\mathbf{U}} \underbrace{(\boldsymbol{\Lambda}_C \otimes (\mathbf{I}_M - \bar{\mu}\boldsymbol{\Lambda}_R))}_{\boldsymbol{\Lambda}_A} \underbrace{(\mathbf{U}_C^T \otimes \mathbf{U}_R^T)}_{\mathbf{U}^T}. \quad (\text{B.19})$$

Portanto, os autovalores de \mathbf{A} podem ser aproximados por:

$$\lambda_a(n + Km) = \lambda_c(n) (1 - \bar{\mu}\lambda_u(m)), \quad 0 \leq m \leq M - 1, 0 \leq n \leq K - 1 \quad (\text{B.20})$$

e os elementos $b(k)$ são os elementos diagonais de:

$$\mathbf{U}^T \mathbf{D}^2 \mathbf{Q} \mathbf{R}_u \mathbf{U} = (\mathbf{U}_C^T \otimes \mathbf{U}_R^T) (\boldsymbol{\mu}^2 \boldsymbol{\sigma}_\eta^2 \otimes \mathbf{I}_M) (\mathbf{I}_K \otimes \mathbf{R}) (\mathbf{U}_C \otimes \mathbf{U}_R). \quad (\text{B.21})$$

Rearrmando os termos:

$$\mathbf{U}^T \mathbf{D}^2 \mathbf{Q} \mathbf{R}_u \mathbf{U} = (\mathbf{U}_C^T \boldsymbol{\mu}^2 \boldsymbol{\sigma}_\eta^2 \mathbf{U}_C) \otimes \underbrace{\mathbf{U}_R^T \mathbf{R} \mathbf{U}_R}_{\boldsymbol{\Lambda}_R}. \quad (\text{B.22})$$

Lembrando que o autovetor de \mathbf{C} associado ao autovalor dominante ($\lambda_c = 1$) é dado por $\mathbf{v} = \frac{1}{K} \mathbf{1}_K$, temos que a seguinte aproximação é também válida:

$$\mathbf{U}_C^T \boldsymbol{\mu}^2 \boldsymbol{\sigma}_\eta^2 \mathbf{U}_C \approx \left(\frac{1}{K} \sum_{k=1}^K \sigma_{\eta,k}^2 \mu_k \right) \mathbf{I}_K. \quad (\text{B.23})$$

Portanto, temos que:

$$\mathbf{U}^T \mathbf{D}^2 \mathbf{Q} \mathbf{R}_u \mathbf{U} \approx \left(\frac{1}{K} \sum_{k=1}^K \sigma_{\eta,k}^2 \mu_k \right) \mathbf{I}_K \otimes \boldsymbol{\Lambda}_R. \quad (\text{B.24})$$

Através dessa expressão para os elementos $b(k)$, (B.14) pode ser manipulada de modo que o MSD é aproximado por:

$$\mathcal{M}_{CTA} \approx S \frac{1}{K} \sum_{k=1}^K \mu_k^2 \sigma_{\eta,k}^2, \quad (\text{B.25})$$

em que S é um termo constante que depende apenas da média dos passos de adaptação:

$$S = \sum_{n=1}^K \sum_{m=1}^M \frac{\lambda_u(m)}{1 - (\lambda_c(n)(1 - \bar{\mu}\lambda_u(m)))^2}. \quad (\text{B.26})$$

B.2 Algoritmo LMS-ATC

A expressão do MSD para o algoritmo LMS-ATC pode ser obtida de forma similar ao desenvolvimento relativo ao algoritmo LMS-CTA. A diferença se dá na expressão da equação discreta de Lyapunov, dada por (B.6) no caso CTA. No caso ATC, a equação de Lyapunov é a seguinte [59]:

$$\mathbf{X} = \mathbf{A}_{ATC} \mathbf{X} \mathbf{A}_{ATC}^T + \mathbf{G} \mathbf{D}^2 \mathbf{Q} \mathbf{R}_u \mathbf{G}. \quad (\text{B.27})$$

Escrevendo \mathbf{A}_{ATC} em termos de seus valores singulares, temos:

$$\mathbf{A}_{ATC} = \mathbf{U}\mathbf{\Lambda}_A\mathbf{V}^T. \quad (\text{B.28})$$

Substituindo (B.28) em (B.27), podemos reescrever (B.27) em termos da variável intermediária $\mathbf{Y} = \mathbf{V}^T\mathbf{X}\mathbf{V}$ e, finalmente, pode-se reescrever (B.27) como:

$$\mathbf{Y} = \mathbf{A}_{ATC}\mathbf{X}\mathbf{A}_{ATC}^T + \mathbf{G}\mathbf{D}^2\mathbf{Q}\mathbf{R}_u\mathbf{G}. \quad (\text{B.29})$$

Observe que essa expressão é similar à (B.6), com a diferença de que a matriz \mathbf{G} pré e pós multiplica o termo $\mathbf{Q}\mathbf{R}_u$. Seguindo o mesmo desenvolvimento do caso CTA, conclui-se que:

$$\mathcal{M}_{ATC} \approx S \frac{1}{K} \sum_{n=1}^K \mu_k^2 \sigma_{\eta,k}^2, \quad (\text{B.30})$$

em que S é um termo constante que depende apenas da média dos passos de adaptação:

$$S = \sum_{n=1}^K \sum_{m=1}^M \frac{\lambda_u(m)\lambda_c^2(n)}{1 - (\lambda_c(n)(1 - \bar{\mu}\lambda_u(m)))^2}. \quad (\text{B.31})$$

Apêndice C

Publicações Relevantes

Os trabalhos em anexo foram publicados durante o período do Doutorado e estão diretamente relacionados à Tese:

1. A. de Paula and C. Panazio, “Cooperative spectrum sensing under unreliable reporting channels,” *Wireless Networks*, pp. 1–9, 2014.
2. A. de Paula and C. Panazio, “Analysis of distributed parameter estimation in WSN with unreliable nodes,” in *Wireless Communication Systems (ISWCS), 2012 International Symposium on*, Aug 2012, pp. 116–120.
3. A. de Paula and C. Panazio, “Diffusion LMS networks in the presence of noisy nodes: A convergence rate and msd analysis,” in *Machine Learning for Signal Processing (MLSP), 2013 IEEE International Workshop on*, Sept 2013, pp. 1–6.

Cooperative spectrum sensing under unreliable reporting channels

Amanda de Paula · Cristiano Panazio

© Springer Science+Business Media New York 2014

Abstract This article analyzes a cooperative spectrum sensing scheme using a distributed approach with a fusion center considering an unreliable reporting channel. The spectrum sensing is applied to a cognitive radio system, where each cognitive radio sends its decision to a fusion center through a reporting channel, in which an n -out-of- K rule is applied. We compare the performance of the decision rules, analyzing the impact of the errors introduced by the reporting channel, considering the Bayes risk criterion.

Keywords Cognitive radio · Spectrum sensing · Detection and estimation · Bayes risk

1 Introduction

The increasing demand for communication resources is leading to a scarcity in the spectral bands available for transmission. Such a scarcity is mainly due to the inflexible spectrum allocation policies, where the bands are statically assigned. As shown in [1], this scenario leads to an inefficient spectral occupancy.

Motivated by the necessity of implementing more efficient band allocation schemes, several papers have recently proposed systems based on cognitive radio (CR) [2–4]. In such systems, the CRs are the secondary users (SUs) that are allowed to occupy the band licensed to primary users

(PUs), if the PUs are not using it for the time being. Such systems are implemented through the so-called cognitive cycle [3], in which the CR is able to sense the environment and to adapt itself to its conditions.

Specifically in the sensing step of the cognitive cycle, the CRs should be able to determine whether the spectral band is vacant or not. This task is performed applying a spectrum sensing method [5]. However, its performance can be highly degraded in hostile environments where the hidden terminal problem arises [6]. An alternative to overcome this problem is to apply cooperative algorithms [6–8]. These cooperative algorithms provide a better spatial coverage and are more robust to fading and shadowing since they are able to combine the measurements provided by the various CRs in the system in order to generate a more reliable spectrum sensing.

The distributed cooperative spectrum sensing can be performed with or without a fusion center. In the former scheme, each CR sends its decision to a common fusion center. In the latter, the CRs exchange information among them using some message parsing strategy.

Restricting our attention to the approach with a fusion center, a problem that arises is how to merge the decisions provided by the different CRs in order to provide a more reliable sensing.

In [8, 9], it is pointed out that the OR decision rule is more suitable in many cases of practical interest. However, these analysis considered that the reporting channel between the CR and the fusion center was perfect. In [10–13] the cooperative spectrum sensing systems were analyzed when errors are introduced by the reporting channel and the receiver operating characteristic (ROC) curves obtained by the n -out-of- K rules were investigated and compared. It was observed that differently from the perfect reporting channel case, the optimum rule would be other than the OR rule.

A. de Paula (✉) · C. Panazio
Department of Control and Telecommunications Engineering,
Escola Politécnica of the University of São Paulo, São Paulo,
SP, Brazil
e-mail: amanda@lcs.poli.usp.br

C. Panazio
e-mail: cpanazio@lcs.poli.usp.br

Also, for some applications, instead of analyzing the ROC curves, it is more interesting to adopt a Bayesian optimization method [14]. The Bayesian approach is particularly interesting when prior probabilities can be assigned to each hypothesis, which is exactly the CR case, where these prior probabilities can be given in terms of a channel occupancy model.

In [15], the decision rule was optimized considering a Bayesian approach for the case of a perfect reporting channel. In [7], the CR system was analyzed in terms of a *total error* criterion, which is a simplification of the Bayesian approach. This analysis was also restricted to the perfect reporting channel scenario. In [16], the authors considered a Bayesian approach in the case of imperfect reporting channel. However, in [16] the authors are concerned mainly with the limitation due to the reporting channel and do not consider the effect of the imperfect local detection in the Bayes risk.

In this article, differently from [10–13], we adopt a Bayesian criterion and investigate the decision rules of the kind n -out-of- K , analyzing the impact of the errors introduced by the reporting channel. Also differently from [16], we analyze the impact of both imperfect reporting channel and imperfect local detectors. In this context, we provide a detailed analysis comparing the OR and the AND rules and we also extend the analytical expression that gives the n value that minimizes the Bayes risk to the case of imperfect reporting channel context. In addition, in the simulation results sections, we also show how the imperfect reporting channel impacts on the the optimum n value if the Neyman-Pearson (NP) criterion is adopted.

This article is organized as follows. In Sect. 2, the system model utilized throughout this paper is depicted. In Sect. 3, local and cooperative spectrum sensing are described. In Sect. 4, we present the expression for Bayes risk and calculate it lower bounds. In Sect. 5 we analyze how the reporting error impact on the comparison between the performance of two different decision rules. Section 6 presents the solution for the minimum Bayes risk problem given fixed local detectors. The simulation results are shown in Sect. 7 Finally, in Sect. 8, the conclusions of the paper are stated.

2 System model

In this article, we consider a cooperative CR system with K CRs. We assume that the i th CR receives the signal transmitted by the PU through a channel h_i and that the signal is corrupted by additive white Gaussian noise (AWGN). Each CR senses the spectrum and sends an one-bit quantized decision to the fusion center through the imperfect reporting channel modeled as a binary symmetric channel (BSC) with crossover probability P_e . Finally, the

fusion center decides that the spectrum is occupied by applying an n -out-of- K rule, which means that the PU is stated to be active if the fusion center receives at least n -out-of- K decisions reporting it. This system model is depicted in Fig. 1.

3 Problem formulation

3.1 Local sensing

The received signal in the i th CR can be expressed as one of the following hypotheses:

$$r_i(k) = \begin{cases} \eta_i(k), & \mathcal{H}_0 \\ h_i x(k) + \eta_i(k), & \mathcal{H}_1 \end{cases}, \quad 1 \leq k \leq M, \quad (1)$$

where h_i is the channel coefficient, which magnitude is assumed to be a Rayleigh random variable with variance $\sigma_{h_i}^2$ and static during the window of M samples, where each sample is presented by the index k , $x(k)$ is the signal transmitted by the PU and $\eta_i(k)$ is an AWGN with variance $\sigma_{\eta_i}^2$.

In this paper, we assume that each CR applies an energy detection rule in order to decide between these two hypotheses, but other detection rules are also possible without loss of generality. The energy decision rule consists in the comparison of the estimated signal energy to a given threshold λ . The estimated received signal energy, i.e., the decision statistic is given by:

$$T(r_i) = \frac{1}{\sigma_{\eta_i}^2} \sum_{k=1}^M |r_i(k)|^2. \quad (2)$$

The hypothesis test is then accomplished by:

$$T(r_i) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda_i, \quad (3)$$

which means that the i th CR states that the spectrum is occupied by the PU if the metric $T(r_i)$ is greater than λ_i .

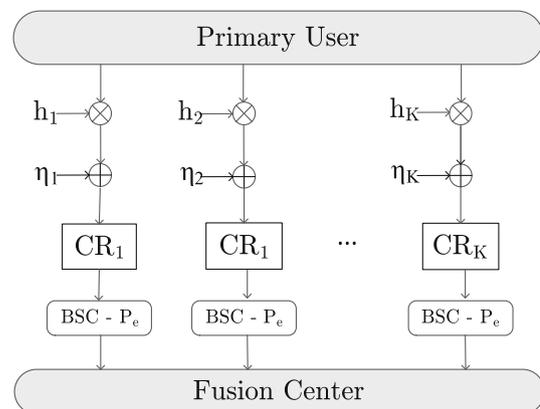


Fig. 1 System model

Without loss of generality, we assume in this work that the noise variance is known to all secondary users. This is a common assumption also adopted, for instance, in [6–8, 10]. In the case that the noise power is unknown, it can be estimated through the techniques presented in [17]. If the noise power is not well estimated the detector performance can be severely degraded as discussed in [18]. However, it is important to emphasize that, even in this case, the analyses provided herein are still valid since they are not attached to the local detector performance.

In the specification of spectrum sensing systems, we have to define the false alarm and miss detection probabilities (P_f and P_m , respectively). The P_f is defined as the probability of the CR incorrect declaring that the spectrum is occupied under \mathcal{H}_0 , i.e., $P_f = \Pr\{T(r) \geq \lambda | \mathcal{H}_0\}$. The P_m , is defined as the probability of the CR erroneously stating that the spectrum is vacant given that the PU is transmitting, i.e., $P_m = \Pr\{T(r) < \lambda | \mathcal{H}_1\}$.

A single CR using the energy detection rule, in the defined Rayleigh fading scenario, results in the following probabilities [19]:

$$P_f = \frac{\Gamma(M, \frac{\lambda}{2})}{\Gamma(M)}, \tag{4}$$

$$P_m = 1 - e^{-\frac{\lambda}{2}} \sum_{l=0}^{M-2} \frac{(\frac{\lambda}{2})^l}{l!} + \left(\frac{1 + 2M\gamma}{2M\gamma}\right)^{M-1} \left(e^{-\frac{\lambda}{2+4M\gamma}} - e^{-\frac{\lambda}{2}} \sum_{l=0}^{M-2} \frac{\left(\frac{2\lambda M\gamma}{2+4M\gamma}\right)^l}{l!} \right) \tag{5}$$

where $\Gamma(x)$ s the gamma function, $\Gamma(x, y)$ is the upper incomplete gamma function and γ is the average system signal-to-noise ratio (SNR) under \mathcal{H}_1 :

$$\gamma = \frac{E\left\{\sum_{k=1}^M |h_i x(k)|^2\right\}}{E\left\{\sum_{k=1}^M |\eta_i(k)|^2\right\}} \tag{6}$$

These results are valid considering that the signal is static during the observation time.

It is important to note that P_f and P_m are parameterized by the threshold λ . P_f is a decreasing function of λ , while P_m is an increasing function of λ . Therefore, in order to specify the threshold λ , one should analyze the compromise between low P_f and high P_m .

3.2 Cooperative sensing

Previously, we have analyzed the spectrum sensing performed in each CR. In this subsection, we deal with the data processing in the fusion center.

Without loss of generality, in this article, we consider that the statistical distribution parameters of channel coefficients (h_i) and the AWGN variance ($\sigma_{\eta_i}^2$) are the same for every CR. This make it possible to parametrize P_f and P_m as a function of only one λ for all CRs. In addition, we consider a particular case where each CR decision sent to the fusion center is corrupted with equal average error probability P_e . With this assumption, the mathematical expression of the overall false alarm and miss detection probabilities can be shown in a more compact form. The general case, where each CR decision is corrupted with different values of P_e , can be derived from this particular case, but it would result in large mathematical expressions without any additional behavior that can be seen from this particular case.

Applying the n -out-of- K rule, we have that the overall false alarm and overall miss detection probabilities after the decision provided by the fusion center are respectively given by [11]:

$$Q_f = \sum_{i=0}^{K-n} \binom{K}{i} [(1 - P_f)(1 - P_e) + P_f P_e]^i [P_f(1 - P_e) + (1 - P_f)P_e]^{K-i} \tag{7}$$

$$Q_m = \sum_{i=K-n+1}^K \binom{K}{i} [P_m(1 - P_e) + (1 - P_m)P_e]^i [(1 - P_m)(1 - P_e) + P_m P_e]^{K-i} \tag{8}$$

4 Bayes risk

In this article, we use the Bayes risk criterion that is a linear combination of $Q_f(n)$ and $Q_m(n)$. Assuming that correct decisions incur in no additional cost, the Bayes risk is defined as [14, 15]:

$$\mathcal{R}(n) = Q_f(n)Pr\{\mathcal{H}_0\}C_f + Q_m(n)Pr\{\mathcal{H}_1\}C_m, \tag{9}$$

where $Pr\{\mathcal{H}_0\}$ is the null-hypothesis a priori probability, i.e., the probability of the spectral hole event, C_f is the associated cost with a false alarm event, $Pr\{\mathcal{H}_1\}$ is the signal hypothesis a priori probability, i.e., the probability of the spectral occupancy event, and C_m is the associated cost to the miss detection event. C_f and C_m are positive constants that usually obey the relation $C_m > C_f$, once the CR system should always protect the PU and the damage caused by a miss detection event is greater than the damage caused by a false alarm occurrence.

The a priori probabilities are defined in terms of the channel occupancy model. In [20], the channel occupancy is modeled as a renew process alternating between the idle and busy states, where each one of these states is modeled as a Poisson random variable with parameters α and β respectively. It is important to note that the mathematical analysis developed in the following could be generalized

for any channel occupancy model, since its choice only interferes in the a priori probabilities calculation. Thus, assuming the Poisson model, the busy and idle mean values are given by $\frac{1}{\alpha}$ and $\frac{1}{\beta}$, which correspond to the following a priori probabilities $Pr\{\mathcal{H}_0\} = \frac{1/\beta}{1/\beta+1/\alpha} = \frac{\alpha}{\alpha+\beta}$ and $Pr\{\mathcal{H}_1\} = \frac{1/\alpha}{1/\beta+1/\alpha} = \frac{\beta}{\alpha+\beta}$.

Hence, the Bayes risk can be written as:

$$\mathcal{R}(n) = Q_f(n) \frac{\alpha}{\alpha + \beta} C_f + Q_m(n) \frac{\beta}{\alpha + \beta} C_m. \tag{10}$$

Before proceeding with the minimization problem on $\mathcal{R}(n)$, it is interesting to investigate the impact of P_e on the Bayes risk. Analyzing (7) and (8), one can note that if $P_f \ll P_e$, $Q_f(n)$ can be approximated as:

$$Q_f(n) \approx \overline{Q}_f(n) = \sum_{i=0}^{K-n} \binom{K}{i} (1 - P_e)^i P_e^{K-i}. \tag{11}$$

In a similar way, the condition $P_m \ll P_e$ leads to the following approximation of $Q_m(n)$:

$$Q_m(n) \approx \overline{Q}_m(n) = \sum_{i=K-n+1}^K \binom{K}{i} P_e^i (1 - P_e)^{K-i}. \tag{12}$$

If these two conditions are simultaneously met, $\mathcal{R}(n)$ can be satisfactorily approximated as:

$$\overline{\mathcal{R}}(n) = \overline{Q}_f(n) \frac{\alpha}{\alpha + \beta} C_f + \overline{Q}_m(n) \frac{\beta}{\alpha + \beta} C_m. \tag{13}$$

We know that due to the compromise between P_f and P_m , the two conditions $P_f \ll P_e$ and $P_m \ll P_e$ simultaneously hold only in high SNR regime and with a very effective detection method. Therefore, usually (13), is a conservative lower bound for $\mathcal{R}(n)$. It is worth to point out that this bound states that, even in a very favorable detection scenario, the reporting error can severely degrade the system performance.

However, it is interesting to note that (13) can be a tight lower bound even if one of these condition is not satisfied. An interesting situation happens for $n = 1$, i.e., the OR rule. If the condition $P_f \ll P_e$ is satisfied, then:

$$Q_f(1) \approx \overline{Q}_f(1) = 1 - (1 - P_e)^K, \tag{14}$$

and applying (8) for $n = 1$:

$$Q_m(1) = (P_m(1 - P_e) + P_e(1 - P_m))^K. \tag{15}$$

In this situation, we can relax the condition over P_m such that:

$$C(P_m) = \frac{\alpha C_f (1 - (1 - P_e)^K)}{\beta C_m (P_m(1 - P_e) + P_e(1 - P_m))^K} \gg 1, \tag{16}$$

In such a case, $\mathcal{R}(1) \approx \overline{\mathcal{R}}(1)$, as will be confirmed in Sect. 7 As n grows, to keep the bound $\overline{\mathcal{R}}(n)$ tight, P_m must gradually become much smaller than P_e , but the constraint over P_f can be increasingly relaxed, becoming similar to (16).

On the other hand, we can also provide upper bounds to the Bayes risk applying a worst case analysis, where we consider a local detector characterized by the condition $P_f + P_m = 1$ [14]. With that condition, one can note that applying (7) and (8) in (10) lead to:

$$A + (B - A)\overline{Q}_m(n) \leq \mathcal{R}(n) \leq B + (A - B)\overline{Q}_f(n) \tag{17}$$

where

$$A = \frac{\alpha C_f}{\alpha + \beta} \tag{18}$$

$$B = \frac{\beta C_m}{\alpha + \beta} \tag{19}$$

if $\beta C_m < \alpha C_f$ the inequalities directions should be switched.

The first bound, $A + (B - A)\overline{Q}_m(n)$, is attained for $P_m = 0$ and $P_f = 1$. The second bound, $B + (A - B)\overline{Q}_f(n)$, is reached in the case $P_m = 1$ and $P_f = 0$. Note that since $\overline{Q}_m(n)$ is an increasing function of n and $\overline{Q}_f(n)$ is a decreasing function of n , reporting errors affect more the second bound for lower values of n than the first one. The opposite statement is valid for higher n values. It is important to emphasize that these bounds are only reached in the case that the local detector are really poor and, for that reason, are of limited practical interest.

5 Analysis through the scalar field Δ_{n_1, n_2}

In this section we investigate the difference between the Bayes risk for two different decision rules with parameters n_1 and n_2 :

$$\Delta_{n_1, n_2} = \mathcal{R}(n_1) - \mathcal{R}(n_2), \tag{20}$$

One can note the (7) and (8) can be rewritten as:

$$Q_f = \sum_{i=0}^{K-n} \binom{K}{i} [1 - P_f^e]^i [P_f^e]^{K-i}, \tag{21}$$

$$Q_m = \sum_{i=K-n+1}^K \binom{K}{i} [P_m^e]^i [1 - P_m^e]^{K-i}, \tag{22}$$

where the new parameters P_f^e and P_m^e are given by:

$$P_f^e = P_f(1 - 2P_e) + P_e, \tag{23}$$

$$P_m^e = P_m(1 - 2P_e) + P_e. \tag{24}$$

It is important to point out that, given an error probability P_e , the new parameters' ranges are:

$$P_e \leq P_f^e, P_m^e \leq 1 - P_e. \tag{25}$$

Therefore, we can see the difference between the Bayes risk as a bivariate scalar field with variables P_f^e and P_m^e [21]:

$$\Delta_{n_1, n_2}(P_f^e, P_m^e) = \mathcal{R}_{n_1}(P_f^e, P_m^e) - \mathcal{R}_{n_2}(P_f^e, P_m^e). \tag{26}$$

If $\Delta_{n_1, n_2}(P_f^e, P_m^e) < 0$, the $n = n_1$ rule is more favorable than the $n = n_2$. Otherwise, it is more advantageous the $n = n_2$ rule.

Two of the more often used fusion rules are the OR and the AND rules, i.e., the $n = 1$ and $n = K$ cases. Considering these two fusion rules, the scalar field $\Delta_{n_1, n_2}(P_f^e, P_m^e)$ can be rewritten as:

$$\Delta_{1, K}(P_f^e, P_m^e) = \frac{1}{\alpha + \beta} \left[\alpha C_f \left(1 - (1 - P_f^e)^K - (P_f^e)^K \right) + \beta C_m \left(-1 + (1 - P_m^e)^K + (P_m^e)^K \right) \right]. \tag{27}$$

It can be shown that (27) presents only one critical point $(P_{f,s}^e, P_{m,s}^e) = (\frac{1}{2}, \frac{1}{2})$ and that the Hessian matrix associated to this critical point is given by:

$$\mathcal{H} = \begin{bmatrix} -2\frac{\alpha}{\alpha+\beta}C_fK(K-1)(\frac{1}{2})^{K-2} & 0 \\ 0 & 2\frac{\beta}{\alpha+\beta}C_mK(K-1)(\frac{1}{2})^{K-2} \end{bmatrix}. \tag{28}$$

The determinant of the Hessian matrix is negative regardless the K value and the Bayes risk parameters. Therefore, one can conclude that this critical point is a saddle point of the scalar field. This means that the surface curves up in one direction and curves down in the other direction. For instance, fixing $P_f^e = P_{f,s}^e = \frac{1}{2}$ and letting $P_m^e \rightarrow P_e$, we have that

$$\Delta_{1, K} \rightarrow \Delta_{1, K}^{\max} = \frac{\alpha C_f}{\alpha + \beta} \left(1 - \frac{1}{2^{K-1}} \right) + \frac{\beta C_m}{\alpha + \beta} (-1 + P_e^K + (1 - P_e)^K) \tag{29}$$

which is the maximum value $\Delta_{1, K}$.

On the other hand, fixing $P_m^e = P_{m,s}^e = \frac{1}{2}$ and letting $P_f^e \rightarrow P_e$, we have that

$$\Delta_{1, K} \rightarrow \Delta_{1, K}^{\min} = \frac{\alpha C_f}{\alpha + \beta} (1 - P_e^K - (1 - P_e)^K) + \frac{\beta C_m}{\alpha + \beta} \left(-1 + \frac{1}{2^{K-1}} \right) \tag{30}$$

which is the minimum value $\Delta_{1, K}$.

Considering that $1 - P_e^K - (1 - P_e)^K \approx 1 - (1 - P_e)^K$, we can show that the scalar field will be non-negative if the following conditions are satisfied:

$$a \leq \frac{1}{1 - \frac{1}{2^{K-1}}} \tag{31}$$

$$P_e \geq g(a) = 1 - \left[1 - a \left(1 - \frac{1}{2^{K-1}} \right) \right]^{\frac{1}{K}}, \tag{32}$$

with a given by $\frac{\beta C_m}{\alpha C_f}$.

Therefore, we can conclude that in these conditions, the AND rule will outperform the OR rule. Considering that $K \gg 1$, (31) can be simplified in $\beta C_m \leq \alpha C_f$. We can interpret this condition as following: a necessary condition for the AND rule to outperform the OR rule is that Q_f impacts more on the Bayes risk than Q_m . The other necessary condition expressed in (32) says that P_e must be larger than a threshold given by (32). Then, in the perfect reporting channel case, we can not state that the AND rule will always outperform the OR rule regardless the P_f and P_m values.

In addition, we can also show that the scalar field will always be non-positive if conditions (31) and (32) hold considering $a = \frac{\alpha C_f}{\beta C_m}$. In this case, a necessary condition to (32) hold is that Q_m impacts more on the Bayes risk than Q_f .

The function $g(a)$ is showed in Fig. 2 for different values of K . One can observe that if Q_f and Q_m impact on the Bayes risk is approximately the same, i.e., if $\beta C_m \approx \alpha C_f \leftrightarrow a \approx 1$, P_e should be prohibitively high to satisfy (32) ($P_e \approx 0.5$). Moreover, one can also note that for increasing K values, the P_e lower bound decreases.

It is interesting to note that in spite of this saddle point $(P_{f,s}^e, P_{m,s}^e)$ being related to the scalar field concerning P_f^e and P_m^e , it is also a saddle point of the scalar field related to P_f and P_m . In order to realize this, note that applying Eqs. (23) and (24), we have that the critical point $(\frac{1}{2}, \frac{1}{2})$ is mapped in:

$$P_f = P_m = \frac{\frac{1}{2} - P_e}{1 - 2P_e} = \frac{1}{2} \tag{33}$$

regardless of the P_e value, and thus $(P_{f,s}^e, P_{m,s}^e) = (\frac{1}{2}, \frac{1}{2})$ is also a saddle point of the field $\Delta_{1, n}(P_f, P_m)$ for all possible P_e values.

This can be noted in Fig. 3, where it is considered a system with $K = 8$ cognitive radios, parameters $\alpha = 0.2$, $\beta = 0.8$, $C_f = 0.3$ and $C_m = 0.7$ and $P_e = 0, 10^{-2}, 10^{-1}$. Observe that regardless the P_e value, the saddle point is fixed in $(\frac{1}{2}, \frac{1}{2})$. It is also interesting to note that for these parameters, we have $a = \frac{\alpha C_f}{\beta C_m} = 0.1071$ satisfying (31).

This value of a is mapped in $g(a) = 0.0139$. Hence, if $P_e > 0.0139$, the OR rule will outperform the AND rule regardless of the P_f and P_m values. In fact, in the $P_e = 10^{-1}$ case, the scalar field is non-positive as we expected.

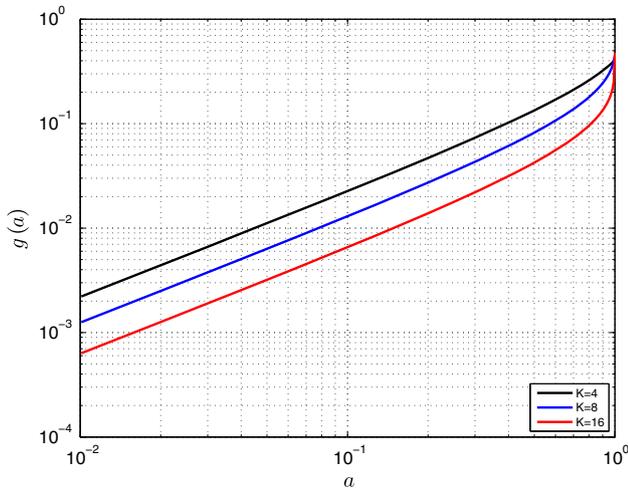


Fig. 2 Function $g(a)$ given by (32) which shows the minimum P_e for which the AND rule surpasses the OR rule when $a = \frac{\beta C_m}{\alpha C_f}$ satisfies (31)

6 The optimum Bayes risk

Another question that arises is how to choose the parameter n that would minimize the Bayes risk, given P_f, P_e and P_m . In [15], this minimization problem is solely solved for the error-free case. This result can be generalize for the case $P_e \neq 0$ as in [16]. This is accomplished using the fact that $\mathcal{R}(n)$ presents a single minimum [15]. Therefore, we can state that the optimum parameter n will be the smallest integer that satisfies:

$$\mathcal{R}(n + 1) - \mathcal{R}(n) \geq 0 \tag{34}$$

Some algebraic manipulations show that if we apply (34) in (9), we have that n should satisfy:

$$n \geq \frac{K - \phi}{1 + \theta} \tag{35}$$

where

$$\theta = \frac{\log\left(\frac{P_f(1-P_e) + (1-P_f)P_e}{(1-P_m)(1-P_e) + P_e P_m}\right)}{\log\left(\frac{P_m(1-P_e) + (1-P_m)P_e}{(1-P_f)(1-P_e) + P_f P_e}\right)}, \tag{36}$$

$$\phi = \frac{\log\left(\frac{\alpha C_f}{\beta C_m}\right)}{\log\left(\frac{P_m(1-P_e) + (1-P_m)P_e}{(1-P_f)(1-P_e) + P_f P_e}\right)}. \tag{37}$$

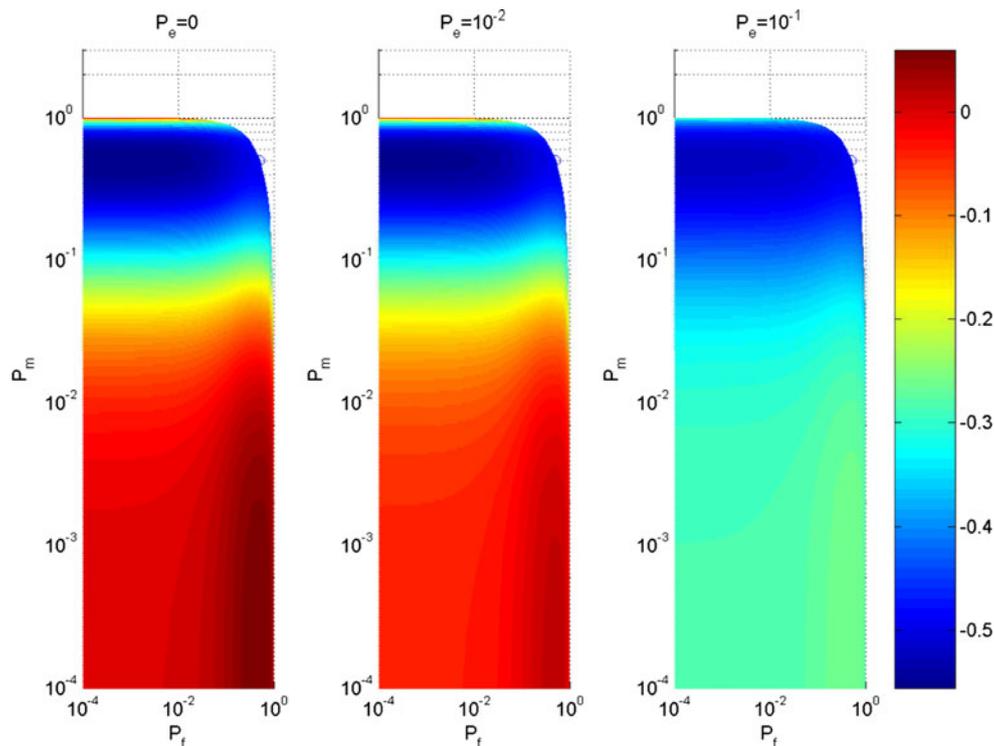


Fig. 3 Scalar field $\Delta_{1,K}$ considering $K = 8, \alpha = 0.2, \beta = 0.8, C_f = 0.3$ and $C_m = 0.7$ for various P_e values

Remembering that n_{opt} must remain in the interval $1 \leq n_{opt} \leq K$, we have that:

$$n_{opt} = \begin{cases} 1, & \left\lceil \frac{K-\phi}{1+\theta} \right\rceil < 1 \\ \left\lceil \frac{K-\phi}{1+\theta} \right\rceil, & 1 \leq \frac{K-\phi}{1+\theta} \leq K \\ K, & \frac{K-\phi}{1+\theta} > K \end{cases} \quad (38)$$

We can note that n_{opt} given by (38) is a function of the parameters θ and ϕ , which depend on the P_f and P_m . However, these probabilities are parameterized by the threshold λ . Therefore, for different values of λ , we will have different values for the n_{opt} parameter.

7 Simulation results

In this section, the theoretical results obtained in the previous sections are corroborated with simulation results. We consider a system with parameters $\alpha = 0.8$, $\beta = 0.2$, $C_f = 0.3$ and $C_m = 0.7$. In such context, the average idle period is longer than the average busy period and the Bayes risk is more affected by the miss detection events than by the false alarm events.

Figure 4 shows the Bayes risk as a function of the parameter λ in a system with $K = 4$ CRs. The channel between the CRs and the PU is Rayleigh with an average SNR equal to 3 dB. An energy detector is applied with $M = 16$. The system is analyzed under the hypotheses of an error-free report channel, i.e., $P_e = 0$, and an imperfect report channel, with $P_e = 10^{-2}$. From Fig. 4, one can note that the value of n corresponding to the minimum Bayes risk depends on the λ value as predicted from the theoretical analysis. It is important to observe that the variation between the Bayes risk with and without error depends on the chosen n parameter, where one can note that the impact is greater for the $n = 1$ case than for higher n values.

In order to explain such behavior for the $n = 1$ case, it is important to indicate that P_f and P_m in the minimum $\mathcal{R}(1)$ are $P_f = 2.0635 \cdot 10^{-4}$ and $P_m = 0.2532$. Note that only P_f satisfies the condition $P_f \ll P_e$, although P_m is such that $\mathcal{C}(P_m) = 15.2126 \gg 1$. Therefore, the lower bound for \mathcal{R} presented in (13) should be tight in this situation. In fact, $\mathcal{R}(1) = 0.0103$ versus $\overline{\mathcal{R}}(1) = 0.0095$ for $P_e = 10^{-2}$. Note also that the reporting errors severely degrade the system performance, since the minimum value of $\mathcal{R}(1)$ with $P_e = 0$ is $7.72 \cdot 10^{-4}$, which is much lower than the minimum achievable $\overline{\mathcal{R}}(1) = 0.0095$ with $P_e = 10^{-2}$. For the other n values, this situation does not occur and the performance degradation is more subtle. As a consequence, in spite of the global minimum of \mathcal{R} being reached with $n = 1$ for $P_e = 0$, for $P_e = 10^{-2}$ the $n = 2$ provides the minimum Bayes risk, which is by far lower than $\mathcal{R}(1)$.

In Fig. 5, we show the optimum n value obtained applying (38). The analytical expression for n_{opt} can be validated comparing Figs. 4 and 5. For instance, we can see that in the error-free case, for $\lambda > \lambda_1 = 57.9$, the fusion rule that leads to the lowest Bayes risk is given by $n = 1$. This is exactly the predicted value expressed in Fig. 5. Similarly, when $P_e = 10^{-2}$, $n = 1$ is the optimum fusion rule for $\lambda > \lambda_2 = 73.2$, which corresponds to the n_{opt} given in Fig. 5.

We can also corroborate the theoretical results showed in Sect. 5 comparing the OR and the AND rules, when $\alpha = 0.2$, $\beta = 0.8$, $C_f = 0.3$ and $C_m = 0.7$. These parameters are such that $a = \frac{\alpha C_f}{\beta C_m} = 0.1071$, applying (32), we can state that if $P_e \geq 0.0243$, the OR rule will always surpass the AND rule. From the results shown in Fig. 6, we can observe that considering the error-free case, the AND rule slightly surpasses the OR rule for $13 \leq \lambda \leq 30$. However if

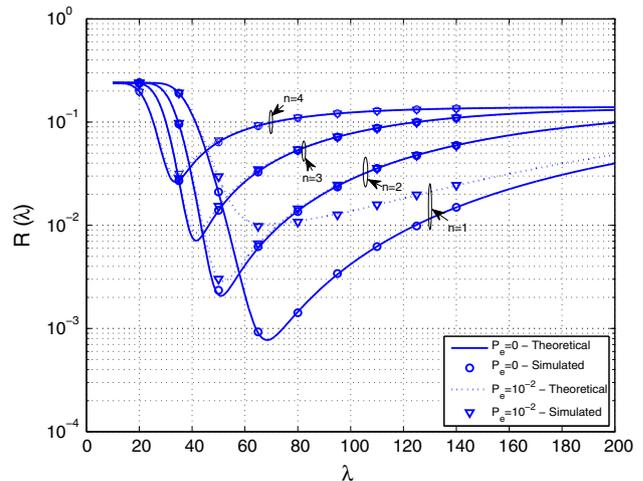


Fig. 4 Bayes risk for different values of n with $K = 4$

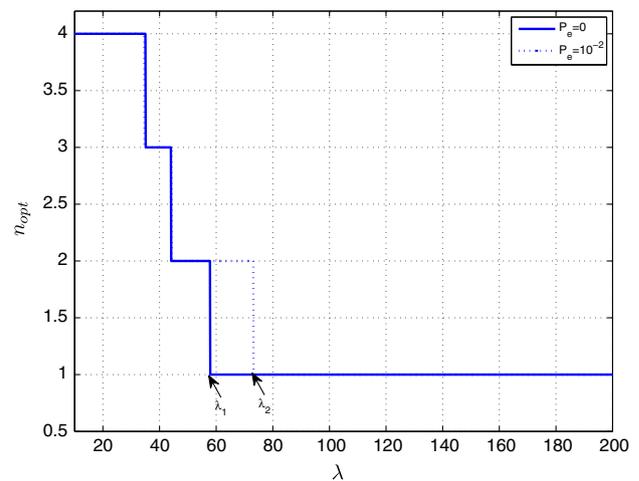


Fig. 5 Optimum n parameter given by (38)

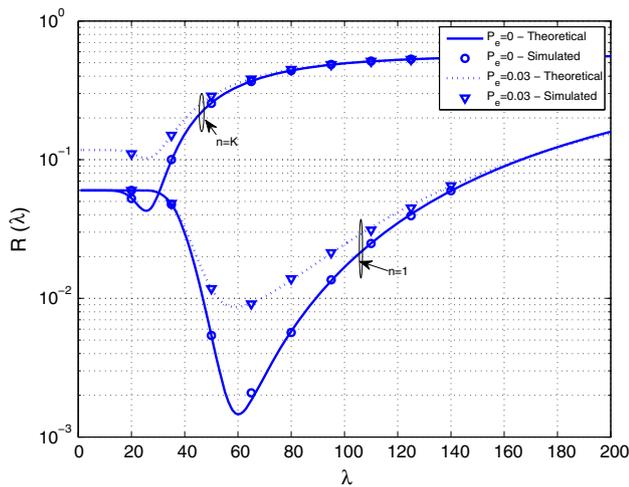


Fig. 6 Comparison between the OR and the AND rules with parameters $\alpha = 0.2$, $\beta = 0.8$, $C_f = 0.3$ and $C_m = 0.7$

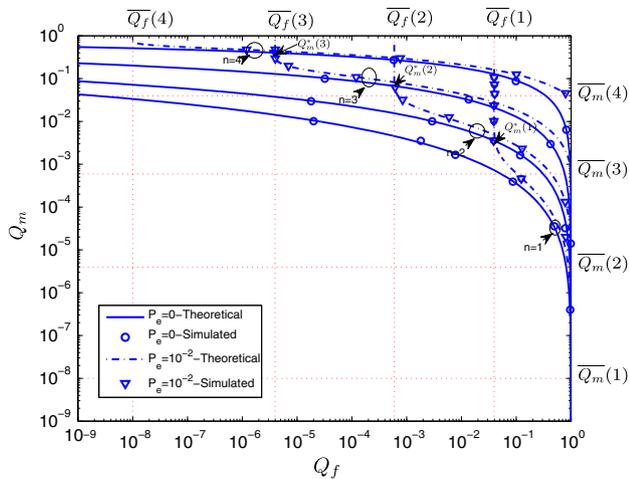


Fig. 7 ROC for different values of n with $K = 4$

$P_e = 0.03$, the OR rule surpasses the AND rule as we expected.

One can also analyze the detector performance under the NP criterion. In a such case, for a given miss detection probability, the optimum detector is the one that gives the minimum false alarm probability. In Fig. 7 we show the ROC for all n values. From this result, one can note that, concerning the perfect reporting channel case, the optimum n -value is $n = 1$ regardless the fixed target Q_m value. Considering a reporting channel with error probability $P_e = 10^{-2}$, the optimum n depends on the fixed target Q_m . Denoting this probability as Q_m^t , we have that:

$$n_{opt} = \begin{cases} 1, & Q_m^t \leq Q_m^*(1) \\ n + 1, & Q_m^*(n) < Q_m^t \leq Q_m^*(n + 1) \\ K, & Q_m^t > Q_m^*(K - 1) \end{cases} \quad (39)$$

where $Q_m^*(n)$ corresponds to the minimum miss-detection probability that leads to $\overline{Q}_f(n)$, as indicated in Fig. 7.

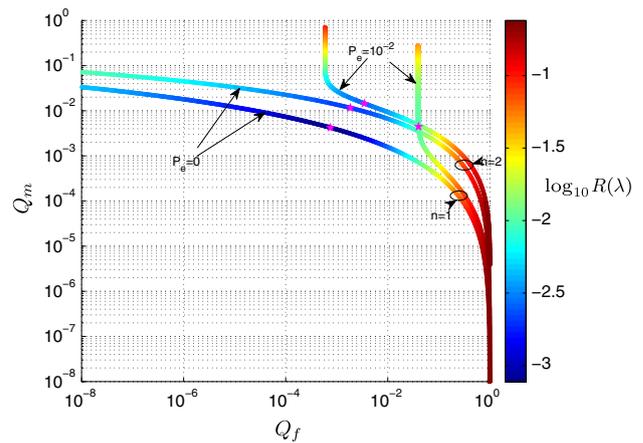


Fig. 8 Impact on the Bayes risk when the NP criterion is adopted. The asterisk indicates the points where the Bayes risk is minimum

Another interesting way to analyze the impact of the reporting error in the system is analyzing the Fig. 8, where the ROC colormap is influenced by the Bayes risk. It is observed that the ROC curves for the $n = 2$ case present almost the same color range for both $P_e = 0$ and $P_e = 10^{-2}$. On the other hand, for the $n = 1$ case, the ROC color range corresponding to $P_e = 10^{-2}$ is associated to larger values of the Bayes risk than the one related to $P_e = 0$, as already observed in Fig. 4.

These walls presented in the ROC curves shown in Figs. 7 and 8 are due to the phenomenon of the bit-error-probability (BEP) wall studied in [11, 12, 16]. In [11, 12] the BEP wall impact on the ROC was analyzed and it was discussed that the AND and OR rules are more sensitive to this degradation imposed by the reporting channel. It was also shown that this degradation increases with the number of cognitive radios.

8 Conclusion

In this paper we have analyzed cooperative spectrum sensing systems with fusion center in the context of imperfect reporting channel. We have analyzed the impact of the unreliable reporting channel condition on the optimum decision rule, concerning the minimum Bayes risk. This theoretical analysis was corroborated by simulation results. We have also provided a detailed comparison between the OR and the AND rules, presenting conditions on the error probability that lead to good scenarios for both decision rules. In the simulation result section, the decision rules performance were evaluated using both Bayesian and NP criteria. The theoretical and the simulation results have shown that, depending on the decision rule, the impact of imperfect reporting channel can lead to a significant performance degradation.

Acknowledgments This work was funded by CAPES (Brazil).

References

1. Ellingson, S. W. (2005). Spectral occupancy at VHF: Implications for frequency-agile cognitive radios. *IEEE Vehicular Technology Conference*, 2, 1379–1382.
2. Mitola, J. III, Maguire, G. Q. Jr. (1999). Cognitive radio: making software radios more personal. *IEEE Personal Communications*, 6(4), 13–18.
3. Haykin, S. (2005). Cognitive radio: Brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications*, 23(2), 201–220.
4. Marinho, J., & Monteiro, E. (2012). Cognitive radio: Survey on communication protocols, spectrum decision issues, and future research directions. *Wireless Networks*, 18(2), 147–164.
5. Haykin, S., Thomson, D. J., & Reed, J. H. (2009). Spectrum sensing for cognitive radio. *Proceedings of the IEEE*, 97(5), 849–877.
6. Quan, Z., Cui, S., Poor, H. V., & Sayed, A. (2008). Collaborative wideband sensing for cognitive radios. *IEEE Signal Processing Magazine*, 25(6), 60–73.
7. Zhang, W., Mallik, R., & Letaief, K. B. (2009). Optimization of cooperative spectrum sensing with energy detection in cognitive radio networks. *IEEE Transactions on Wireless Communications*, 8(12), 5761–5766.
8. Ghasemi, A., & Sousa, E. S. (2007). Opportunistic spectrum access in fading channels through collaborative sensing. *Journal of Communications*, 2(2), 71–82.
9. Letaief, K. B., & Zhang, W. (2009). Cooperative communications for cognitive radio networks. *Proceedings of the IEEE*, 97(5), 878–893.
10. Atapattu, S., Tellambura, C., & Jiang, H. (2011). Energy detection based cooperative spectrum sensing in cognitive radio networks. *IEEE Transactions on Wireless Communications*, 10(4), 1232–1241.
11. Chaudhari, S., Lunden, J., Koivunen, V., & Poor, H. V. (2012). Cooperative sensing with imperfect reporting channels: Hard decisions or soft decisions?. *IEEE Transactions on Signal Processing*, 60(1), 18–28.
12. Chaudhari, S., Lunden, J., Koivunen, V., & Vincent Poor, H. (2013). BEP walls for cooperative sensing in cognitive radios using k-out-of-n fusion rules. *Signal Processing*, 93(7), 1900–1908.
13. Birkan Yilmaz, H., Tugcu, T., & Fatih, A. (2012). Novel quantization-based spectrum sensing scheme under imperfect reporting channel and false reports. *International Journal of Communication Systems*. doi:10.1002/dac.2408.
14. Kay, S. M. (1998). *Fundamentals of statistical signal processing, volume II: Detection theory*. New Jersey: Prentice Hall.
15. Varshney, P. K. (1996). *Distributed detection and data fusion*. New York: Springer.
16. Chaudhari, S., Lunden, J., Koivunen, V. (2012). BEP walls for cooperative bayesian detection with reporting channel errors. In *2012 IEEE 23rd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pp. 2166–2172.
17. Mariani, A., Giorgetti, A., & Chiani, M. (2011). Effects of noise power estimation on energy detection for cognitive radio applications. *IEEE Transactions on Communications*, 59(12), 3410–3420.
18. Yin, W., Ren, P., Cai, J., & Su, Z. (2013). Performance of energy detector in the presence of noise uncertainty in cognitive radio networks. *Wireless Networks*, 19(5), 629–638.
19. Digham, F. F., Alouini, M. S., & Simon, M. K. (2007). On the energy detection of unknown signals over fading channels. *IEEE Transactions on Communications*, 55(1), 21–24.
20. Zhou, X., Ma, J., Li, G., Kwon, Y., & Soong, A. C. K. (2010). Probability-based combination for cooperative spectrum sensing. *IEEE Transactions on Communications*, 58(2), 463–466.
21. Stewart, J. (2000). *Multivariable calculus: Concepts and contexts*. Brooks/Cole.



Amanda de Paula received the B.Sc. degree from Federal University of Pernambuco (UFPE) in 2008 and the M.Sc. degree from Escola Politécnica of the University of São Paulo (EPUSP) in 2010, both in electrical engineering. She is currently pursuing the Ph.D. degree also in electrical engineering at Escola Politécnica of the University of São Paulo (EPUSP). Her research interests include multicarrier modulations, cognitive radio, detection and estimation theory.



Cristiano Panazio received a B.Sc. and an M.Sc. degree in electrical engineering from the State University of Campinas (UNICAMP), Brazil, in 1999 and 2001, respectively. He received his Ph.D. in 2005, also in electrical engineering, from the Conservatoire National des Arts Métiers (CNAM), Paris, France. In 2006, he became assistant professor at Escola Politécnica of the University of São Paulo. He is member of the Brazilian Telecommunications

Society (SBRT) since 1999. His research interests include equalization, multicarrier modulations, spread spectrum techniques and cognitive radio.

Analysis of distributed parameter estimation in WSN with unreliable nodes

Amanda de Paula, Cristiano Panazio
Escola Politécnica
University of São Paulo
São Paulo, Brazil
Email: amanda,cpanazio@lcs.poli.usp.br

Abstract—In this article we analyze the diffusion normalized least mean square (NLMS) and its set-membership version (SM-NLMS) diffusion algorithms in a scenario where sensor nodes are subjected to different noise variances. We show through simulation that the SM-NLMS is a more robust algorithm in such condition, in addition to the provided reduced energy consumption. We also show that, in such context, the reduced feedback SM-NLMS (RF-SM-NLMS) presents a similar performance when compared to the SM-NLMS with an additional energy saving and lower channel occupancy. Moreover, we propose an adaptive way to choose the SM-NLMS and RF-SM-NLMS parameters in order to provide further performance enhancement in the presence of nodes subjected to different noise variances.

I. INTRODUCTION

Estimation of an unknown deterministic parameter in a wireless sensor network (WSN) is an extensively exploited topic in the literature [1]–[3]. It can be applied in a variety of contexts such as agriculture environments, where parameters such as temperature, humidity or wind speed are needed to be estimated. As another example, WSN can be also applied in the healthy area, where patients physiological data must be monitored.

In the majority of WSN implementations a fusion center is required in order to perform the estimation. For instance, see the systems described in [2] and [4]. The main concern with these implementations requiring a fusion center is that they tend to demand a high energy consumption. In order to mitigate this problem, implementations that avoid the use of a fusion center have gained attention in the last years [5], [6].

An interesting proposal is the distributed adaptive filters [7], [8]. Most of the classical adaptive algorithms can be also applied in distributed contexts. For example, distributed version of the least mean square (LMS) algorithm can be found in [7], and of the recursive least squares (RLS) can be found in [9].

Due to the importance of energy saving in WSN, a natural solution to the distributed adaptive filtering problem is to employ algorithms based on the set membership (SM) idea [10]. The application of the SM algorithm in distributed context was exploited in [11], where it is shown how energy consumption can be reduced applying this algorithm when compared to the normalized least mean square (NLMS) algorithm.

*The authors would like to thank CAPES and FAPESP for financial support.

Another aspect that deserves attention in the distributed algorithms analysis is the case where the observation in each node in the network does not provide the same reliability regarding the estimation of the unknown parameter. In this situation, the estimate provided by the noisy nodes, *i.e.*, the nodes that provide the least reliable parameter estimation, tends to deteriorate the global estimate in the network. This heterogenous scenario in a WSN can arrive in many situations. For instance, the spectrum sensing in a cognitive radio network in which the channel between each radio and the common primary user presents a different signal-to-noise ratio is an example of such situation [12]. In this case, the noisy cognitive radios would deteriorate the network global estimate. A possible solution to overcome this problem is to provide an adaptive cooperative scheme as in [13] or to use a variable step size LMS as proposed in [14].

Here, we show through simulations that the SM in addition to be more advantageous in terms of energy consumption, it also shows a more robust behavior in a network where noisy nodes are present. We also show how to further improve the SM performance in a distributed network with noisy nodes, providing a variable SM threshold inspired on the variable step size algorithm shown in [14].

The rest of this paper is organized as follows. In Section II we show the network topology used throughout this paper, followed by Section III, where we describe the algorithms used in the local node processing. In Section IV, we show an approach to choose the local processing parameters, and in Section V, we show some simulations results. Finally, the conclusions are stated in Section VI.

II. DISTRIBUTED ADAPTIVE ESTIMATION

In this article, we consider that the cooperation between the nodes occur accordingly to the diffusion rule in [8]. In this mode of cooperation, each node can exchange its estimate with a set of nodes. The network considered throughout this paper presents the topology described in Fig. 1.

We consider that the nodes can reliably exchange information among the paths depicted in the network. In addition, we assume that the network is a bidirectional network, which means that if the node k is connected to node l , they can exchange information with each other. Still regarding the

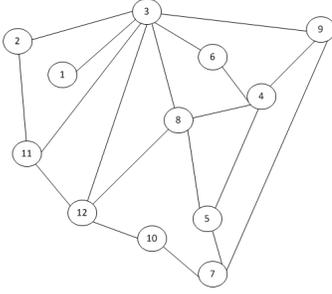


Fig. 1. Diffusion network

network topology, we define the neighborhood of a node k , \mathcal{N}_k , as the set of all nodes connected to it, including itself.

In a local node k , a linear model is assumed. Therefore, every node $k \in 1, \dots, N$ has access to the set of measurements $\{\mathbf{u}_k(i), d_k(i)\}$, that are related as follows:

$$d_k(i) = \mathbf{u}_k(i)^T \mathbf{w}_{opt} + \eta_k(i) \quad (1)$$

where \mathbf{w}_{opt} is the optimum deterministic unknown vector to be estimated and $\eta_k(i)$ is the Gaussian measurement noise with zero mean and variance $\sigma_{\eta,k}^2$.

In the distributed adaptive estimation algorithm, in addition to the pair $\{\mathbf{u}_k(i), d_k(i)\}$, each node in the network has access to the estimates generated in its neighborhood, defined as $\psi_l, l \in \mathcal{N}_k$. Then, an aggregate estimate ϕ_k is obtained by the fusion of the local estimates. Restricting this function to a linear combination, ϕ_k is given by:

$$\phi_k(i) = \sum_{l \in \mathcal{N}_k} c_{k,l} \psi_l(i) \quad (2)$$

The coefficients $c_{k,l}$ can be chosen in several ways. In [7], it is stated that the coefficients matrix \mathbf{C} should be such that the sum of the elements in its column must be equal to one. With this condition, the following relation must hold:

$$\sum_{l \in \mathcal{N}_k} c_{k,l} = 1 \quad (3)$$

A possible choice for these coefficients is given accordingly to the Metropolis rule [7]. Assuming this rule, we have that:

$$c_{k,l} = \begin{cases} \frac{1}{\max\{n_k, n_l\}}, & k \in \mathcal{N}_l \\ 0, & k \notin \mathcal{N}_l \\ 1 - \sum_{l \in \mathcal{N}_k, l \neq k} c_{k,l} & k = l \end{cases} \quad (4)$$

where n_k and n_l represent the number of nodes connected to the node k and to the node l , respectively.

III. LOCAL NODE PROCESSING

Locally, each node in the network can apply several algorithms in order to update its estimate. It is common that versions of the LMS algorithm are applied in order to keep a low local node processing complexity. Furthermore, in this article, two different variants of the LMS algorithm are considered: the NLMS algorithm and the set-membership normalized least mean square (SM-NLMS) algorithm.

A. NLMS

Considering the NLMS algorithm, the local estimate $\phi_k(i)$ is updated as described in Table 1.

Table 1

Algorithm 1 NLMS

$$\begin{aligned} \phi_k(i) &= \sum_{l \in \mathcal{N}_k} c_{k,l} \psi_l(i) \\ e(i) &= d(i) - \phi_k(i)^T \mathbf{u}_k(i) \\ \psi_k(i+1) &= \phi_k(i) + \mu_k \frac{e(i) \mathbf{u}_k(i)}{\mathbf{u}_k(i)^T \mathbf{u}_k(i) + \delta} \end{aligned}$$

The μ_k in Table 1 is the NLMS step-size for the k^{th} node and δ is a small constant that avoids division by zero in the filter coefficients update.

B. SM-NLMS

An alternative algorithm to the NLMS is given by the SM algorithm. By doing so, the local estimates are updated only if the estimation error is greater than a threshold. The SM-NLMS algorithm is described in Table 2 [15].

Table 2

Algorithm 2 SM-NLMS

$$\begin{aligned} \phi_k(i) &= \sum_{l \in \mathcal{N}_k} c_{k,l} \psi_l(i) \\ e(i) &= d(i) - \phi_k(i)^T \mathbf{u}_k(i) \end{aligned}$$

if $|e(i)| > \gamma_k$ **then**
 $\mu_k = 1 - \frac{\gamma_k}{|e(i)|}$
else
 $\mu_k = 0$
end if

$$\psi_k(i+1) = \phi_k(i) + \mu_k \frac{e(i) \mathbf{u}_k(i)}{\mathbf{u}_k(i)^T \mathbf{u}_k(i) + \delta}$$

In Table 2, γ_k is the set-membership threshold which is commonly chosen as $\gamma_k = \sqrt{\tau_k \sigma_{\eta,k}^2}$ with τ_k being an auxiliary constant that is commonly set to satisfy $\tau_k \in (1, 10)$ [16]. As will be clarified in the Section V, the auxiliary constant τ_k plays an important role in the convergence speed of the algorithm.

C. RF-SM-NLMS

In the previous section, we described the distributed SM-NLMS as presented in [11]. Also in [11], it is suggested to not only use the SM algorithm to update the local estimate, but to also utilize its idea to share only important information with the neighborhood. This is accomplished defining the constraint set:

$$\mathcal{H}_k(i) = \{\psi_l \mid |d_k(i) - \mathbf{u}_k(i)^T \psi_l| < \gamma_{fb,k}\} \quad (5)$$

where $\gamma_{fb,k}$ is a threshold parameter similar to γ_k in the SM-NLMS. We set $\gamma_{fb,k} = \sqrt{\tau_{fb,k} \sigma_{\eta,k}^2}$, where $\tau_{fb,k}$ is an auxiliary constant that impacts on the information exchange probability of the algorithm.

If the estimate $\psi_l(i) \in \mathcal{H}_k(i)$, it means that this estimate does not provide an important information to the node k . As a conclusion, the node l does not have to send this estimate to the node k . With such approach, we reduce the channel occupancy for message exchange among the nodes and more energy can be saved. The higher $\tau_{fb,k}$ is, the less restrictive is the constraint set and the less likely will be the information exchange among the nodes.

This approach is termed as reduced feedback set-membership (RF-SM-NLMS), which is described in Table 3 [11].

Table 3

Algorithm 3 RF-SM-NLMS

$$\begin{aligned} \mathcal{N}'_k &= \{l \in \mathcal{N}_k \mid \psi_l(i-1) \notin \mathcal{H}_k(i-1)\} \\ \phi_k(i) &= \sum_{l \in \mathcal{N}'_k} c_{k,l} \psi_l(i) \\ e(i) &= d(i) - \phi_k(i)^T \mathbf{u}_k(i) \end{aligned}$$

if $|e(i)| > \gamma_k$ **then**
 $\mu_k = 1 - \frac{\gamma_k(i)}{|e(i)|}$
else
 $\mu_k = 0$
end if

$$\psi_k(i+1) = \phi_k(i) + \mu_i \frac{e(i) \mathbf{u}_k(i)}{\mathbf{u}_k(i)^T \mathbf{u}_k(i) + \delta}$$

IV. NLMS STEP-SIZE AND SM-NLMS τ PARAMETER

In this section, we show how to choose the NLMS step-size and the SM-NLMS auxiliary parameter in the context where different noise variances are experienced in the nodes.

In [14] an approach to choose the distributed LMS step-size based on the noise average variance in each node is provided. The idea is that smaller step-sizes should be provided to the nodes with higher noise variances. Therefore, the interference of such noisy nodes in the network estimate is mitigated. With that approach, the LMS step-size is given as:

$$\mu_k = (\mu_{max} - \mu_{min}) e^{-a\sigma_{\eta,k}^2} + \mu_{min} \quad (6)$$

where a is a constant that controls the impact of $\sigma_{\eta,k}^2$ in the step-size value. Actually, accordingly to (6), the nodes from which the observation are more reliable, *i.e.*, with small $\sigma_{\eta,k}^2$, present a larger step-size, providing a faster convergence. On the other hand, since the noisy nodes are kept with small step-size, they do not compromise the quality of the estimate.

We propose then a similar approach in order to enhance the quality of the estimate provided by the SM-NLMS and the RF-SM-NLMS algorithm. Analyzing these algorithms, it is interesting to note that the SM-NLMS and the RF-SM-NLMS algorithm can be viewed as a variable step-size NLMS algorithm. In addition, it is important to point out that for higher γ_k values, the mean value of the equivalent variable step-size decreases. Therefore, in order to guarantee that the reliable nodes will present the fastest convergence, we have to guarantee that γ_k is an increasing function of $\sigma_{\eta,k}^2$. Restricting

the γ_k parameter to the form $\gamma_k = \sqrt{\tau_k \sigma_{\eta,k}^2}$, we propose to obtain the auxiliary parameter τ_k by using the following relation:

$$\tau_k = (\tau_{max} - \tau_{min}) \left(1 - e^{-b\sigma_{\eta,k}^2}\right) + \tau_{min} \quad (7)$$

where b is a constant that controls the impact of $\sigma_{\eta,k}^2$ in the auxiliary SM parameter value.

V. SIMULATION RESULTS

In this section, the three algorithms are simulated in different contexts. For all simulated scenarios, the network topology is the one depicted in Fig. 1, where there are $N = 12$ nodes. The vector to be estimated has equal M coefficients and unitary variance $\mathbf{w}_{opt} = [1, 1, \dots, 1] / \sqrt{M}$. In addition, the input signal $u_k(i)$ is a Gaussian random variable with zero mean and unitary variance. In the following, we analyze the algorithms under four different contexts: A) when $\sigma_{\eta,k}^2 = \sigma_{\eta}^2$, B) when the noise variances are not the same in the nodes, C) when the noise variances are different with μ_k and τ_k given by (6) and (7), and D) for different noise distributions. Under all scenarios, the algorithms are analyzed through the global instantaneous mean-square deviation (MSD) defined as:

$$\text{MSD}(i) = \frac{1}{N} \sum_{k=1}^N \|\psi_k(i) - \mathbf{w}_{opt}\|^2 \quad (8)$$

The MSD results were obtained through the mean of 2000 independent realizations.

A. Equal noise variance

In this subsection, we considered that $\sigma_{\eta,k}^2 = \sigma_{\eta}^2 = 10^{-2}$. For the NLMS algorithm, we assumed a step-size equal to $\mu_k = \mu = 0.22$, for the SM-NLMS the auxiliary parameter was set equal to $\tau_k = \tau = 2.8$ and for the RF-SM-NLMS we used $\tau_k = \tau = 3.2$ and $\tau_{k,fb} = \tau_{fb} = 2$. The τ_k values were chosen in order to keep the steady-value MSD approximately equal for both RF-SM-NLMS and SM-NLMS.

The results are depicted in Fig. 2. From it, we can notice that for MSD values higher than approximately -28 dB, the NLMS algorithm presents a slow convergence rate when compared to the others algorithms. It is also important to emphasize that it also presents a higher processing complexity than the SM-NLMS and the RF-SM-NLMS. In addition, the RF-SM-NLMS presents a reduction in channel occupancy and higher energy efficiency, since information is exchanged only in a fraction of time. For the RF-SM-NLMS, the average channel occupancy in the network in steady-state is 15%.

Since the RF-SM-NLMS only exchange information when the estimated vector obeys certain conditions, it can be though as the SM-NLMS with a time varying topology, *i.e.*, when information is not exchanged in the RF-SM-NLMS, it behaves as a SM-NLMS with reduced connections. Therefore, there is a small performance degradation of the RF-SM-NLMS, when it is compared to the SM-NLMS algorithm.

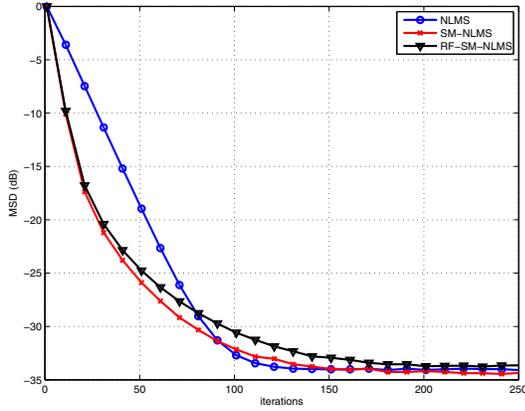


Fig. 2. Nodes with the same noise variance

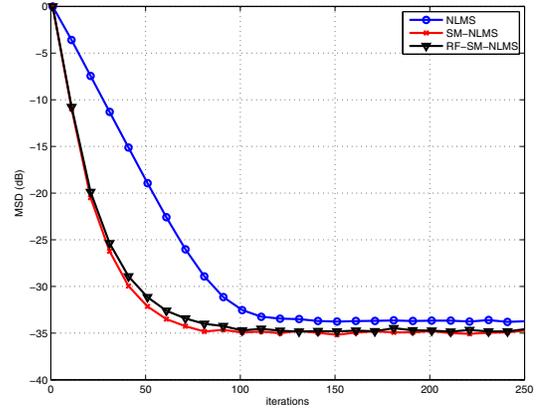


Fig. 3. Nodes with different noise variance

B. Different noise variance

As a second simulation scenario we considered that the noise variances were given by:

$$\sigma_{\eta,k}^2 = \begin{cases} 0.196, & k = 1, \dots, 6 \\ 10^{-4}, & k = 7, \dots, 10 \\ 10^{-3}, & k = 11, \dots, 12 \end{cases} \quad (9)$$

Note that the average noise variance is $\overline{\sigma_{\eta}^2} = 10^{-2}$ as in the previous case. We used the same parameters considered in Subsection V-A. From the results depicted in Fig. 3, we can note that the SM-NLMS algorithm provides faster convergence than the NLMS algorithm. Even the RF-SM-NLMS converges faster than the NLMS algorithm for all MSD range. In addition, these algorithms converges to a lower steady-state MSD when compared to the NLMS.

By the results shown in Fig. 3, we conclude that the SM version of the distributed algorithms seem to be a more suitable choice for this kind of scenario. In order to understand such behavior, we analyze the update probability of the SM-NLMS algorithm, presented in Fig. 4, where we can observe that the nodes with lower noise variance present a higher update probability. From [16], we can conclude that the update probability is a decreasing function of the auxiliary parameter τ . Therefore, in this heterogeneous scenario, the algorithm behaves as if the nodes with lower noise variance work with a lower τ . In this condition, the algorithm converges faster without an increase of the MSD level, since the faster convergence is provided by the low noise variance nodes.

It is worth noting that the performance degradation introduced by the RF-SM-NLMS in relation to the SM-NLMS is mitigated when compared to Subsection V-A scenario. This fact is explained by two reasons: (i) there is an increasing in the channel occupancy to 22% in this case versus 15% in the previous case (ii) the nodes with higher variance exchange information more often than the nodes with lower variance (for instance, 35% in node 8 versus 16% in node 1).

C. Different $\sigma_{\eta,k}^2$ with μ_k and τ_k given by (6) and (7)

In order to enhance the performance of the algorithms in the context in which the nodes experience different noise variance,

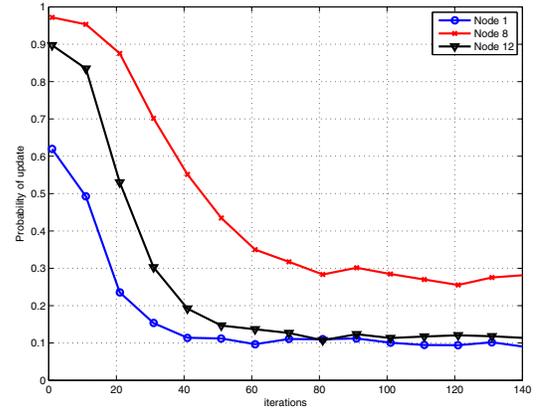


Fig. 4. Update probability for the nodes $k = 1, 8, 12$ in the scenario described in Subsection V-B

we applied the strategies described in Sec. IV. In (6), we set $\mu_{min} = 0.08$, $\mu_{max} = 0.45$, $a = 100$. For the SM-NLMS and the RF-SM-NLMS, we set $\tau_{min} = 2$, $\tau_{max} = 5.5$, $b = 100$ in (7). With these parameters, the μ and τ values varies with $\sigma_{\eta,k}^2$ as indicated in Fig. 6.

The MSD results are depicted in Fig. 5. We can see that the algorithms present a performance gain when these techniques are applied with faster convergence and lower MSD. However, both the SM-NLMS and the RF-SM-NLMS algorithms still present a superior performance when compared to the NLMS.

Then, we conclude that even if the performance of the NLMS is improved by the means of a variable step-size technique, we can also improve the performance of the SM-NLMS algorithm in a similar way. Therefore, the SM-NLMS technique still outperforms the NLMS technique in such heterogenous noise distribution environment.

D. Different noise distributions

In this subsection, we consider that the noise variance in each node is given by:

$$\sigma_{\eta,k}^2 = \begin{cases} \frac{10^{-2}}{\alpha^{(k-1)}}, & k = 1, \dots, 6 \\ 10^{-2} \left(2 - \frac{1}{\alpha^{(12-k)}} \right), & k = 7, \dots, 12 \end{cases} \quad (10)$$

where $\alpha \geq 1$ is a constant that is related to the noise distribution. If $\alpha = 1$, the network presents a uniform noise

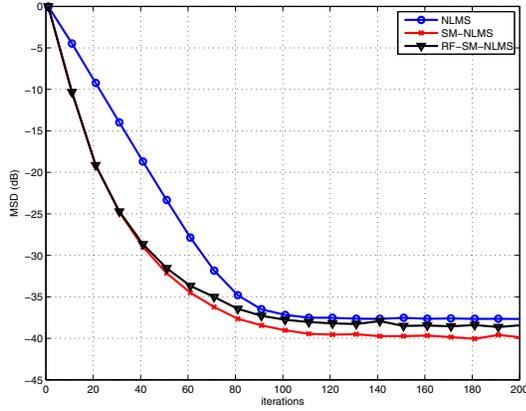


Fig. 5. Nodes with different noise variance and μ_k and τ_k given by (6) and (7)

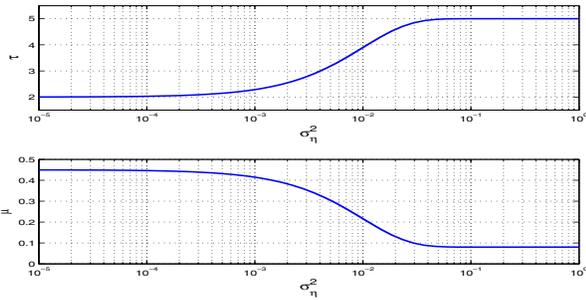


Fig. 6. τ and μ for various σ_η^2 values.

distribution, as α increases the noise distribution becomes less uniform. Note that for every value of α , the average noise variance is $\overline{\sigma_\eta^2} = 10^{-2}$.

In Fig. 7, we show how the algorithms behave for different noise distributions with the same parameters used in the Subsection V-A. We have observed that the NLMS is invariant to the noise distribution, as long as the mean noise variance is constant, then, in Fig. 7, we only represent the NLMS result for $\alpha = 1$. For the SM-NLMS, as well as the RF-SM-NLMS,

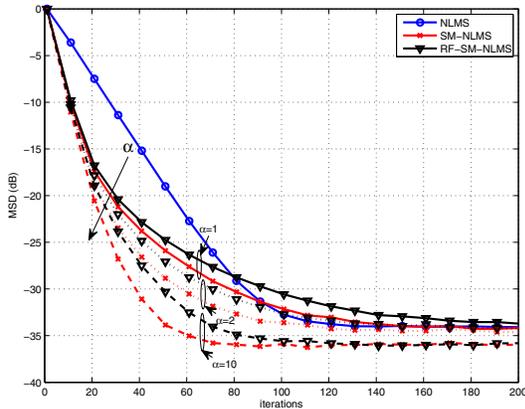


Fig. 7. Performance comparison for networks with different noise distributions given by (10)

there is a performance gain both in convergence speed and in MSD steady-state level as α increases.

VI. CONCLUSION

In this article we compared the distributed versions of the NLMS, the SM-NLMS and its reduced feedback version, the RF-SM-NLMS, under different contexts. We showed that the SM-NLMS and the RF-SM-NLMS are more advantageous than the NLMS specially in the case where some nodes in the network are subject to different noise variances. We also provided an alternative to improve the performance of the SM-NLMS and the RF-SM-NLMS in the presence of noisy nodes, inspired in the variable step-size solution proposed for the LMS algorithm in [14].

REFERENCES

- [1] J.-F. Chamberland and V.V. Veeravalli, "Wireless sensors in distributed detection applications," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 16–25, May 2007.
- [2] M. Guerriero, L. Svensson, and P. Willett, "Bayesian data fusion for distributed target detection in sensor networks," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3417–3421, June 2010.
- [3] J.N. Tsitsiklis, "Decentralized detection," *Advances in Signal Processing*, vol. 2, pp. 197–344, 1993.
- [4] R. Viswanathan and P.K. Varshney, "Distributed detection with multiple sensors i. fundamentals," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.
- [5] Qing Ling and Zhi Tian, "Decentralized sparse signal recovery for compressive sleeping wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3816–3827, July 2010.
- [6] S. Barbarossa and G. Scutari, "Decentralized maximum-likelihood estimation for sensor networks composed of nonlinearly coupled dynamical systems," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3456–3470, July 2007.
- [7] C.G. Lopes and A.H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3122–3136, July 2008.
- [8] C.G. Lopes and A.H. Sayed, "Incremental adaptive strategies over distributed networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4064–4077, Aug. 2007.
- [9] F.S. Cattivelli, C.G. Lopes, and A.H. Sayed, "Diffusion recursive least-squares for distributed estimation over adaptive networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1865–1877, May 2008.
- [10] S. Gollamudi, S. Nagaraj, S. Kapoor, and Yih-Fang Huang, "Set-membership filtering and a set-membership normalized lms algorithm with an adaptive step size," *IEEE Signal Processing Letters*, vol. 5, no. 5, pp. 111–114, May 1998.
- [11] S. Werner, Yih-Fang Huang, M.L.R. de Campos, and V. Koivunen, "Distributed parameter estimation with selective cooperation," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. ICASSP 2009.*, April 2009, pp. 2849–2852.
- [12] F.S. Cattivelli and A.H. Sayed, "Distributed detection over adaptive networks using diffusion adaptation," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 1917–1932, May 2011.
- [13] N. Takahashi, I. Yamada, and A.H. Sayed, "Diffusion least-mean squares with adaptive combiners: Formulation and performance analysis," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4795–4810, Sept. 2010.
- [14] A. Rastegarnia, M.A. Tinati, and A. Khalili, "A distributed incremental lms algorithm with reliability of observation consideration," in *IEEE International Conference on Communication Systems (ICCS), 2010*, Nov. 2010, pp. 67–70.
- [15] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, Springer, 2008.
- [16] M.V.S. Lima and P.S.R. Diniz, "On the steady-state mse performance of the set-membership nlms algorithm," in *7th International Symposium on Wireless Communication Systems (ISWCS), 2010*, Sept. 2010, pp. 389–393.

DIFFUSION LMS NETWORKS IN THE PRESENCE OF NOISY NODES: A CONVERGENCE RATE AND MSD ANALYSIS

Amanda de Paula and Cristiano Panazio

Departamento de Engenharia de Telecomunicações e Controle (PTC)
Escola Politécnica, University of São Paulo SP, Brazil,
e-mails: {amanda.cpanazio}@lcs.poli.usp.br

ABSTRACT

In this article, we provide a detailed analysis of the diffusion least mean square algorithm in the specific context where each node in the network presents different SNR values. We show, through an eigenvalue analysis, the condition that the algorithm step-size values should obey in order to provide a given convergence rate. Under this condition, it is shown how to set the algorithm step-size in each node that lead to the minimum mean square deviation. We also provide simulation results that corroborate our theoretical analysis.

Index Terms— adaptive algorithm, diffusion LMS, eigenvalue analysis, wireless sensor network

1. INTRODUCTION

Estimation of an unknown deterministic parameter in a wireless sensor network (WSN) is an extensively exploited topic in the literature [1–3]. It can be applied in a variety of contexts such as agricultural environments, where parameters such as temperature, humidity or wind speed are needed to be estimated. In addition, WSN can be also applied in the health area, where patients physiological data must be monitored.

Often, a fusion center is required in order to perform the estimation in a WSN. These implementations are known as centralized networks, this is the case in the systems described in [2] and [4]. One drawback of such implementations is that it usually have a high power demand. In order to overcome this problem, implementations that avoid the use of a fusion center have gained attention in the last years [5, 6].

In this context, recently distributed adaptive filters have been proposed in the literature [7], [8] and most of the classical adaptive algorithms have also been proposed in their distributed counterpart. For instance, distributed version of the least mean square (LMS) algorithm is described in [7], and the recursive least squares (RLS) is described in [9].

An important point that deserves attention in the distributed algorithms analysis is the case where the observation in each node in the network does not provide the same reliability regarding the estimation of the unknown parameter. In

this situation, the estimate provided by the noisy nodes, *i.e.*, the nodes that provide the least reliable parameter estimation, tends to deteriorate the global estimate in the network. This heterogenous scenario in a WSN can arrive in many situations. For instance, the spectrum sensing in a cognitive radio network in which the channel between each radio and the common primary user presents a different signal-to-noise ratio is an example of such situation [10]. In this case, the noisy cognitive radios would deteriorate the network global estimate.

In this paper, we consider that the distributed algorithm is applied employing a diffusion approach [7]. In this context, we show the condition under which the algorithm converges with a given rate and find the optimal step-size assignment for each node with the objective to minimize the mean-square deviation (MSD) of the estimated parameters. We consider both correlated and non-correlated input regressor model in a full mesh network, and in the simulation scenario section, we extend this results to more general network topologies.

The rest of this paper is organized as follows. In Section 2 we describe the diffusion algorithm used throughout this article. In Section 3, we analyze the algorithm convergence rate, in Section 4 the optimum step size configuration for a given convergence rate is derived, and in Section 5, we show some simulations results. Finally, the conclusions are presented in Section 6.

2. DISTRIBUTED ADAPTIVE ESTIMATION

In this article, we consider that the cooperation between the nodes occur accordingly to the diffusion rule [8]. In this mode of cooperation, each node can exchange its estimate with a set of nodes.

We assume that the nodes can reliably exchange information among the paths depicted in the network. Regarding the network topology, we define the neighborhood of a node k , \mathcal{N}_k , as the set of all nodes connected to it, including itself.

In a local node k , a linear model is assumed. Therefore, every node $k \in 1, \dots, N$ has access to the set of measure-

ments $\{\mathbf{u}_k(i), d_k(i)\}$, that are related as follows:

$$d_k(i) = \mathbf{u}_k(i)^T \mathbf{w}_{opt} + \eta_k(i) \quad (1)$$

where \mathbf{w}_{opt} is the optimum deterministic unknown vector to be estimated and $\eta_k(i)$ is the Gaussian measurement noise with zero mean and variance $\sigma_{\eta,k}^2$.

In the distributed adaptive estimation algorithm, in addition to the pair $\{\mathbf{u}_k(i), d_k(i)\}$, each node in the network has access to the estimates generated in its neighborhood, defined as $\psi_l, l \in \mathcal{N}_k$. Then, an aggregate estimate ϕ_k is obtained by the fusion of the local estimates. Restricting this function to a linear combination, ϕ_k is given by:

$$\phi_k(i) = \sum_{l \in \mathcal{N}_k} c_{k,l} \psi_l(i) \quad (2)$$

The coefficients $c_{k,l}$ can be chosen in several ways. In [7], it is stated that the coefficient matrix \mathbf{C} should be such that the sum of the elements in its rows must be equal to one.

A possible choice for these coefficients is given accordingly to the Metropolis rule [7]:

$$c_{k,l} = \begin{cases} \frac{1}{\max\{n_k, n_l\}}, & k \in \mathcal{N}_l \\ 0, & k \notin \mathcal{N}_l \\ 1 - \sum_{l \in \mathcal{N}_k, l \neq k} c_{k,l} & k = l \end{cases} \quad (3)$$

where n_k and n_l represent the number of nodes connected to the node k and to the node l , respectively.

Locally, each node in the network can apply several algorithms in order to update its estimate. It is common that versions of the LMS algorithm are applied in order to keep a low local node processing complexity. In the standard LMS diffusion network, the local estimate $\phi_k(i)$ is updated as described in Table 1, where μ_k is the LMS step-size for the k^{th} node.

Table 1

Algorithm 1 LMS

$$\begin{aligned} \phi_k(i) &= \sum_{l \in \mathcal{N}_k} c_{k,l} \psi_l(i) \\ e(i) &= d(i) - \phi_k(i)^T \mathbf{u}_k(i) \\ \psi_k(i+1) &= \phi_k(i) + \mu_k e(i) \mathbf{u}_k(i) \end{aligned}$$

3. LMS STEP-SIZE PARAMETER IN HETEROGENEOUS NETWORK

A question that arises in the heterogeneous network analysis is how to properly choose the step size parameter of each of the nodes in the network.

In order to answer this question, let us analyze a full mesh network. In this case, the combination matrix \mathbf{C} should be chosen as:

$$\mathbf{C} = \frac{1}{N} \mathbf{1}\mathbf{1}^T \quad (4)$$

where N is the number of nodes in the network and $\mathbf{1}$ is a $N \times N$ matrix in which all elements are equal to 1.

An approach to analyze the algorithm convergence rate is to introduce the vector coefficient error vector defined as:

$$\tilde{\boldsymbol{\psi}}(i) = \left[\tilde{\boldsymbol{\psi}}_1^T(i) \tilde{\boldsymbol{\psi}}_2^T(i) \cdots \tilde{\boldsymbol{\psi}}_N^T(i) \right]^T \quad (5)$$

where

$$\tilde{\boldsymbol{\psi}}_k = \boldsymbol{\psi}_k - \mathbf{w}_{opt} \quad (6)$$

It can be shown that the following recursion is valid [7]:

$$\mathbb{E} \left\{ \tilde{\boldsymbol{\psi}}(i) \right\} = (\mathbf{I}_{MN} - \mathbf{D}\mathbf{R}_u) \mathbf{G} \mathbb{E} \left\{ \tilde{\boldsymbol{\psi}}(i-1) \right\} \quad (7)$$

where $\mathbf{G} = \mathbf{C} \otimes \mathbf{I}_M$, $\mathbf{D} = \boldsymbol{\mu} \otimes \mathbf{I}_M$, $\mathbf{R}_U = \mathbf{I}_N \otimes \mathbf{R}$, \mathbf{I}_N is the identity matrix of dimension N and \mathbb{E} denotes the mathematical expectation. The symbol \otimes denotes the Kronecker product, the matrix $\boldsymbol{\mu}$ is a diagonal matrix whose elements are the step size parameter in each of the nodes and the matrix $\mathbf{R} = E \left\{ \mathbf{u}_k \mathbf{u}_k^T \right\}$ is the autocorrelation input matrix that is assumed to be constant to all k .

The convergence rate is determined by the spectral radius, *i.e.*, the magnitude of the largest eigenvalue of the matrix $\mathbf{A} = (\mathbf{I}_{MN} - \mathbf{D}\mathbf{R}_u) \mathbf{G}$.

Applying some Kronecker products properties, it can be shown that the matrix \mathbf{A} can be rewritten as:

$$\mathbf{A} = \mathbf{C} \otimes \mathbf{I}_M - \boldsymbol{\mu} \mathbf{C} \otimes \mathbf{R} \quad (8)$$

3.1. White regressor vector

Modeling the regressor vector as a unitary power white stochastic process, the matrix \mathbf{A} can be rewritten as:

$$\mathbf{A} = (\mathbf{I}_N - \boldsymbol{\mu}) \mathbf{C} \otimes \mathbf{I}_M \quad (9)$$

Defining $\mathbf{A}_1 = (\mathbf{I}_N - \boldsymbol{\mu}) \mathbf{C}$, one can also note that the eigenvalues of \mathbf{A} are equal to the eigenvalues of \mathbf{A}_1 . Therefore, the spectral radius of \mathbf{A} is equal to the spectral radius of \mathbf{A}_1 [11].

The matrix \mathbf{A}_1 applied to a vector \mathbf{v} results in:

$$\mathbf{A}_1 \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix} = \frac{1}{N} \sum_{i=1}^N v_i \begin{bmatrix} 1 - \mu_1 & 0 & \cdots & 0 \\ 0 & 1 - \mu_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 - \mu_N \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad (10)$$

Therefore:

$$\mathbf{A}_1 \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix} = \frac{1}{N} \sum_{i=1}^N v_i \underbrace{\begin{bmatrix} 1 - \mu_1 \\ 1 - \mu_2 \\ \vdots \\ 1 - \mu_N \end{bmatrix}}_{\mathbf{v}_0} \quad (11)$$

Hence, the vector \mathbf{v}_0 is an eigenvector of the matrix \mathbf{A}_1 associated to the eigenvalue $\lambda = 1 - \frac{1}{N} \sum_{i=1}^N \mu_i$. The other

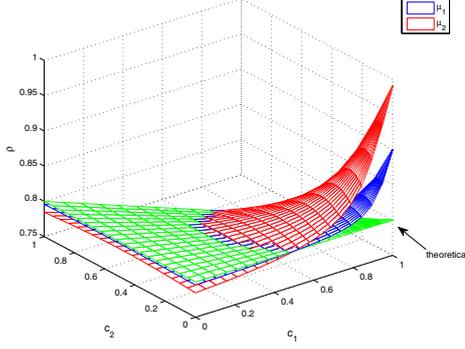


Fig. 1. Spectral radius for the step distributions μ_1 and μ_2

eigenvalues are null since \mathbf{C} is a rank one matrix. Therefore, the spectral radius is given by:

$$\rho = 1 - \frac{1}{N} \sum_{i=1}^N \mu_i \quad (12)$$

Furthermore, we can conclude that if the mean value of the nodes step sizes in the network is constant, the convergence rate of the algorithm is constant.

In spite of the above demonstration being valid for the full mesh network, where the combination matrix is given by (4), the result in (12) is still a good approximation for more general cases, where the combination matrix \mathbf{C} is any stochastic matrix, *i.e.*, the sum of the elements of each one of its row is equal to one. In order to illustrate this, we consider an example with a 3-nodes network, where the combination matrix is given by:

$$\mathbf{C} = \begin{bmatrix} c_1 & c_2 & 1 - c_1 - c_2 \\ 1 - c_1 - c_2 & c_1 & c_2 \\ c_2 & 1 - c_1 - c_2 & c_1 \end{bmatrix} \quad (13)$$

where $0 \leq c_1 \leq 1$ and $0 \leq c_2 \leq 1 - c_1$. Two different step-size configurations with the same mean value are considered: $\mu_1 = [0.3 \ 0.1 \ 0.2]$ and $\mu_2 = [0.39 \ 0.01 \ 0.2]$. Figure 1 shows the spectral radius related to combination matrices of the form (13). One can note that spectral radius is approximately given by the theoretical expression in (12) for a broad set of combination matrices. The only exception is the case where the parameter $c_1 \approx 1$, which can be interpreted as the non-cooperative mode. In such a case, the spectral radius will be equal to $\rho = 1 - \mu_{min}$, for the μ_1 step-size configuration we have $\rho = 0.9$ and for the μ_2 we have $\rho = 0.99$.

3.2. Colored input regressor

In this case, we consider a colored input regressor with correlation matrix given by: $\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, where \mathbf{U} is an unitary matrix and $\mathbf{\Lambda}$ is the diagonal eigenvalue matrix with elements $\mathbf{\Lambda}(k, k) = \lambda_k$.

In this context, we claim the the spectral radius of the matrix \mathbf{A} is given as stated by the following a proposition.

Proposition 3.1. *The spectral radius of the matrix \mathbf{A} is given by:*

$$\rho = 1 - \lambda_{min} \frac{1}{N} \sum_{i=1}^N \mu_i \quad (14)$$

where λ_{min} is the minimum eigenvalue of \mathbf{R} .

Proof. In order to prove this proposition, we note that the matrix \mathbf{R}_U can be written as $\mathbf{R}_U = (\mathbf{I}_N \otimes \mathbf{U}) (\mathbf{I}_N \otimes \mathbf{\Lambda}) (\mathbf{I}_N \otimes \mathbf{U}^H)$. Therefore we can rewrite the matrix \mathbf{A} as:

$$\mathbf{A} = (\mathbf{I}_{MN} - \mathbf{D} (\mathbf{I}_N \otimes \mathbf{U}) (\mathbf{I}_N \otimes \mathbf{\Lambda}) (\mathbf{I}_N \otimes \mathbf{U}^H)) \mathbf{G} \quad (15)$$

Recalling that $\mathbf{D} = \boldsymbol{\mu} \otimes \mathbf{I}_M$, we can apply the mixed product property of the Kronecker product in order and state that $\mathbf{D} (\mathbf{I}_N \otimes \mathbf{U}) = (\mathbf{I}_N \otimes \mathbf{U}) \mathbf{D}$. Using the same property, we can also state that $(\mathbf{I}_N \otimes \mathbf{U}^H) \mathbf{G} = \mathbf{G} (\mathbf{I}_N \otimes \mathbf{U}^H)$.

Therefore, the matrix \mathbf{A} can be written as:

$$\mathbf{A} = (\mathbf{I}_N \otimes \mathbf{U}) \mathbf{E} \mathbf{G} (\mathbf{I}_N \otimes \mathbf{U})^H \quad (16)$$

where, the matrix \mathbf{E} is a diagonal matrix, given by:

$$\mathbf{E} = \mathbf{I}_{NM} - \mathbf{D} (\mathbf{I}_N \otimes \mathbf{\Lambda}) \quad (17)$$

One can note the vectors of the form $\mathbf{v} = [\mathbf{0}_{k-1} \ 1 - \mu_1 \lambda_k \ \mathbf{0}_{M-k} \ \cdots \ \mathbf{0}_{k-1} \ 1 - \mu_N \lambda_k \ \mathbf{0}_{M-k}]^T$ are eigenvectors associated to the matrix $\mathbf{E}\mathbf{G}$ and that the respective eigenvalues are given by $1 - \lambda_k \frac{1}{N} \sum_{i=1}^N \mu_i$. The others $(N - 1)M$ eigenvalues are null since the matrix \mathbf{C} presents $N - 1$ null eigenvalues. Therefore, the maximum eigenvalue is given by $1 - \lambda_{min} \frac{1}{N} \sum_{i=1}^N \mu_i$ proving the proposition. \square

Similarly to the white input case, given that the mean step-size is constant, the convergence rate is also constant.

In order to illustrate this result, let us consider an example where the first line of the input regressor correlation matrix is given by $\mathbf{r} = [1 \ 0.2 \ 0.4 \ 0.1]$.

As in the white input regressor case, we will analyze a network with 3 nodes and combination matrix given by (13). Fig. 2 shows the spectral radius as a function of the combination matrix parameter. One can note that the theoretical expression in (14) is a good approximation concerning a wide selection of the combination matrix parameters.

It is also interesting to extend this analysis to a more realistic scenario, where larger networks can be employed. In order to do so, we will analyze the spectral radius of the algorithm, concerning the network depicted in Fig. 3. The spectral radius of the matrix \mathbf{A} is calculated for different step-size configurations drawn from a uniform distribution and normalized to mean equal do $\bar{\mu} = 0.01$. We considered a colored input matrix with correlation function give by $r(n) = 0.7^{|n|}$. The histogram of the spectral radius obtained with this scenario

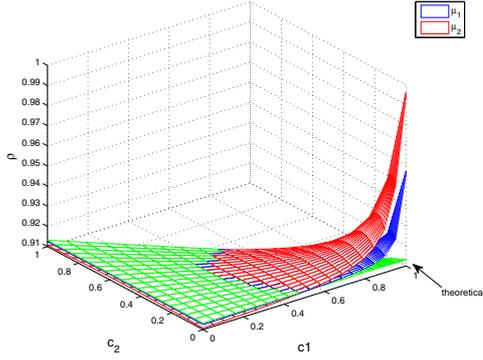


Fig. 2. Spectral radius for the step distributions μ_1 and μ_2 and colored input regressor vector.

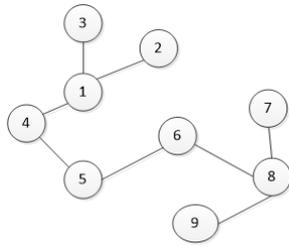


Fig. 3. Diffusion network

is depicted in Fig. 4. From this result, one could note that the spectral histogram is very close to an impulse located in the theoretical value of the spectral calculated assuming a full-mesh network, which corroborates our theoretical analysis.

4. OPTIMUM STEP-SIZE DISTRIBUTION

The adaptive filtering performance is often evaluated through the MSD, defined as follows:

$$\text{MSD} = \frac{1}{N} \sum_{k=1}^N \mathbb{E} \left\{ \left\| \tilde{\psi}_k \right\|^2 \right\} \quad (18)$$

In the last section, we have stated that if the mean step-size for the nodes in the network is constant, the converge rate is approximately the same. The question that arises is how to set the step-size of each node for a given convergence rate to achieve the minimum MSD.

In order to answer to this question, we considered again a 3-nodes full-mesh network. The input regressor power is constant for all the nodes in the network, although, the measurement noise power in each one of the nodes is different and given by: $\sigma_\eta^2 = [0.2 \ 0.04 \ 0.06]$. Fixing the mean step-size as $\bar{\mu} = 0.03$, we varied the step-size for the node 1 and node 2 and measured the MSD. The results are depicted in Fig. 5, where the MSDs are shown as a function of μ_1 and μ_2 . One

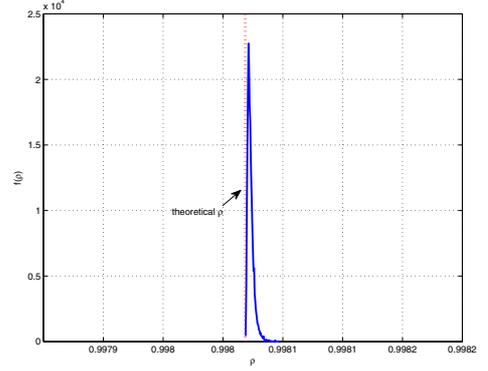


Fig. 4. Histogram of the \mathbf{A} matrix for step-size distributions drawn from an uniform distribution and normalized to $\bar{\mu} = 0.01$ mean.

can see that the minimum MSD is attained when the condition $\mu_1 \approx 0.01$, $\mu_2 \approx 0.05$ and, as a consequence, $\mu_3 \approx 0.03$. From this result one can conjecture that the optimum step-size should be inversely proportional to the measurement noise in each node:

$$\sigma_{\eta,1}^2 \mu_1 = \sigma_{\eta,2}^2 \mu_2 = \dots = \sigma_{\eta,N}^2 \mu_N \quad (19)$$

The condition (19) in conjunction with the condition $\bar{\mu} = \frac{1}{N} \sum_{k=1}^N \mu_k$ gives rise to the following proposition:

Proposition 4.1 (Optimum step-size configuration). *For a given mean step-size $\bar{\mu}$, the step-size in each node should be given by:*

$$\mu_i = \frac{\bar{\mu}}{\sigma_{\eta,i}^2} \mathcal{H}(\sigma_\eta^2) \quad (20)$$

where $\mathcal{H}(\sigma_\eta^2)$ is the harmonic mean of the power noise vector:

$$\mathcal{H}(\sigma_\eta^2) = \frac{1}{\frac{1}{N} \sum_{k=1}^N \frac{1}{\sigma_{\eta,k}^2}} \quad (21)$$

Proof. In order to prove this proposition, we state that the MSD can be written as:

$$\text{MSD} = \frac{1}{N^2} \left[\sum_{i=1}^M \frac{\lambda_i}{(1 - (1 - \lambda_i \bar{\mu})^2)} \right] \left[\sum_{k=1}^N \mu_k^2 \sigma_{\eta,k}^2 \right] \quad (22)$$

This expression is an approximation of the MSD value obtained by applying the independence theory and is more accurate for small step-size values. A similar expression was presented in [12] where an incremental network with white input regressor vector was studied.

Eliminating the constant terms in (22) regarding the step-size, the optimum step-size configuration will be given by the

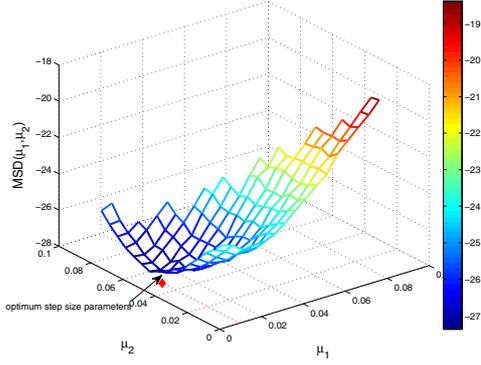


Fig. 5. MSD for different step-size configuration

solution of the following optimization problem:

$$\begin{aligned} & \underset{\boldsymbol{\mu}=[\mu_1 \ \mu_1 \ \dots \ \mu_N]}{\text{minimize}} && \sum_{k=1}^N \mu_k^2 \sigma_{\eta,k}^2 \\ & \text{subject to} && \frac{1}{N} \sum_{k=1}^N \mu_k = \bar{\mu} \end{aligned}$$

In [12], the authors arrive in the same optimization problem, which is solved by setting the Lagrangian gradient to zero. The Lagrangian related to this problem is given by:

$$L(\boldsymbol{\mu}, l) = \sum_{k=1}^N \mu_k^2 \sigma_{\eta,k}^2 + l \left(\frac{1}{N} \sum_{i=1}^N \mu_i - \bar{\mu} \right) \quad (23)$$

Setting the Lagrangian gradient to zero, we have that the following equations must be satisfied:

$$\frac{\partial}{\partial \mu_k} L(\boldsymbol{\mu}, l) = 0 \rightarrow \mu_k = -\frac{l}{2N\sigma_k^2} \quad (24)$$

Therefore, the constraint equation reduces to:

$$\frac{1}{N} \sum_{k=1}^N -\frac{l}{2N\sigma_{\eta,k}^2} = \bar{\mu} \quad (25)$$

Applying the lagrange multiplier l given by (25) in (24) results in (20). \square

Note that this result does not depend on the input regressor vector correlation matrix.

5. SIMULATION RESULTS

The results derived in the last sections assumed that a full mesh network was employed. However, we have pointed out that the results will still be valid for other several combination matrices. In this section, we will corroborate that approximation with some simulation results.

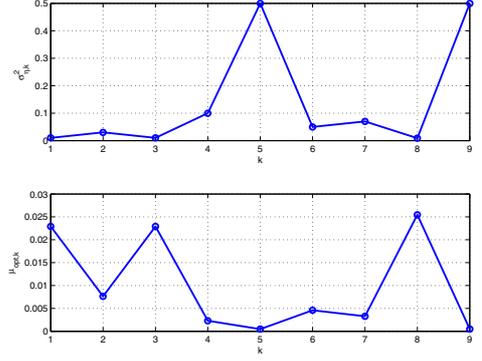


Fig. 6. Noise distribution and the respective optimum step-size

In other to do this, we will consider the 9-node network depicted in Fig. 3. In the first simulation scenario, we will consider that the input regressor vector is a white Gaussian noise with unitary power. In addition, the vector to be estimated has equal $M = 5$ coefficients and unitary variance, *i.e.*, $\mathbf{w}_{opt} = [1, 1, \dots, 1] / \sqrt{M}$.

The noise variance in each one of the nodes is depicted in Fig. 6. We have fixed the mean step-size value in $\bar{\mu} = 0.01$ and calculated the step-size in each node accordingly to (20).

Fig. 7 shows the MSD for this first simulation scenario. It was obtained through the mean of 100 independent realizations of the instantaneous MSD measurements:

$$\text{MSD}(i) = \frac{1}{N} \sum_{k=1}^N \|\psi_k(i) - \mathbf{w}_{opt}\|^2 \quad (26)$$

One can note that the convergence rate is indeed approximately the same, even if we do not consider a full mesh network. In addition to the curve with the optimum step-size vector, we show the MSD for step-sizes drawn from a uniform distribution and normalized to the required $\bar{\mu} = 0.01$. One can note that the step-size configuration that leads to the minimum MSD is the one from Fig. 6, which corroborates the Proposition (4.1). In addition, we can also observe that the minimum MSD given by (22) is also a tight bound to the MSD value. The difference between the MSD value attained by the optimum step-size configuration and the minimum MSD given by (22) arises from the fact that the network is not fully connected. In order to illustrate this, we also show the MSD curve considering the full mesh network, which shows that the minimum MSD matches (22) in this case.

We have also analyzed the behavior of the algorithm for a colored input regressor vector. In this case, we considered that the regressor vector \mathbf{u}_k was modeled as an autoregressive process with autocorrelation function given by $r(n) = 0.7^{|n|}$ and with a white Gaussian unitary variance signal as the input. The results are depicted in Fig. 8, from where we can observe that the convergence rate is constant, given that $\bar{\mu}$ is constant. In addition, we notice that the minimum MSD is achieved

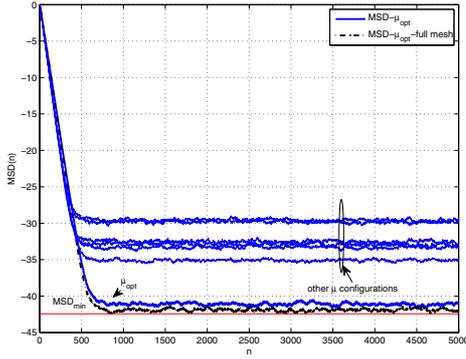


Fig. 7. MSD for different step-size configuration with $\bar{\mu} = 0.01$ and white regressor vector.

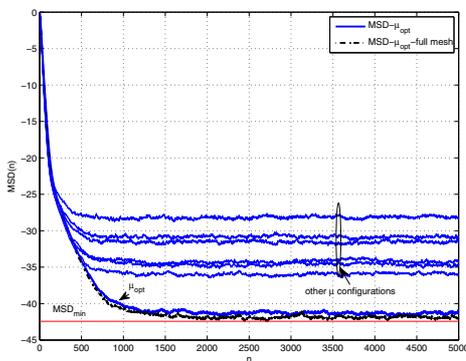


Fig. 8. MSD for different step-size configuration with $\bar{\mu} = 0.01$ and colored regressor vector.

when the step-size are chosen accordingly to the Proposition (4.1). As in the white input regressor vector, (22) is a tight bound for the MSD and matches the MSD value considering the full mesh network.

6. CONCLUSION

In this article we have shown that given that the mean value of the step-size values of the network is fixed, the diffusion LMS convergence rate is also fixed if the network is fully connected. We have also provided simulation results that corroborate that this conclusion could also be extended to more general networks. We showed that the minimum MSD is achieved if the step-size of each node in the network is inversely proportional to the noise variance in each one of the nodes. A disadvantage of this approach is that the nodes should have prior knowledge of the noise variance in the other nodes in the network. As a future work, one could propose strategies to parse this noise variance in the diffusion algorithm in order to overcome this prior knowledge requirement.

7. REFERENCES

- [1] J.-F. Chamberland and V.V. Veeravalli, “Wireless sensors in distributed detection applications,” *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 16–25, May 2007.
- [2] M. Guerriero, L. Svensson, and P. Willett, “Bayesian data fusion for distributed target detection in sensor networks,” *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3417–3421, June 2010.
- [3] J.N. Tsitsiklis, “Decentralized detection,” *Advances in Signal Processing*, vol. 2, pp. 197–344, 1993.
- [4] R. Viswanathan and P.K. Varshney, “Distributed detection with multiple sensors i. fundamentals,” *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.
- [5] Qing Ling and Zhi Tian, “Decentralized sparse signal recovery for compressive sleeping wireless sensor networks,” *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3816–3827, July 2010.
- [6] S. Barbarossa and G. Scutari, “Decentralized maximum-likelihood estimation for sensor networks composed of nonlinearly coupled dynamical systems,” *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3456–3470, July 2007.
- [7] C.G. Lopes and A.H. Sayed, “Diffusion least-mean squares over adaptive networks: Formulation and performance analysis,” *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3122–3136, July 2008.
- [8] C.G. Lopes and A.H. Sayed, “Incremental adaptive strategies over distributed networks,” *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4064–4077, Aug. 2007.
- [9] F.S. Cattivelli, C.G. Lopes, and A.H. Sayed, “Diffusion recursive least-squares for distributed estimation over adaptive networks,” *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1865–1877, May 2008.
- [10] F.S. Cattivelli and A.H. Sayed, “Distributed detection over adaptive networks using diffusion adaptation,” *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 1917–1932, May 2011.
- [11] C.D. Meyer, *Matrix analysis and applied linear algebra*, SIAM, 2000.
- [12] A. Khalili, A. Rastegarnia, J. A. Chambers, and W. M. Bazzi, “An optimum step-size assignment for incremental lms adaptive networks based on average convergence rate constraint,” *AEU - International Journal of Electronics and Communications*, vol. 67, no. 3, pp. 263–268, 2013.