

VÍTOR HELOIZ NASCIMENTO

**ANÁLISE DE ALGORITMOS PARA
FILTRAGEM ADAPTATIVA BASEADOS EM
MOMENTOS DE QUARTA ORDEM**

Tese apresentada à Escola Politécnica da Universidade de São Paulo como requisito para obtenção do Título de Professor Livre Docente, junto ao Departamento de Engenharia de Sistemas Eletrônicos.

São Paulo
2005

VÍTOR HELOIZ NASCIMENTO

**ANÁLISE DE ALGORITMOS PARA
FILTRAGEM ADAPTATIVA BASEADOS EM
MOMENTOS DE QUARTA ORDEM**

Tese apresentada à Escola Politécnica da Universidade de São Paulo como requisito para obtenção do Título de Professor Livre Docente, junto ao Departamento de Engenharia de Sistemas Eletrônicos.

São Paulo
2005

À minha família: Naomi, Benedicto, Magdalena e Carlos.

AGRADECIMENTOS

Aos meus colegas do LPS pelo apoio: Miguel, Denise, Flavio e Mario.

Ao Prof. José Bermudez, pela colaboração que deu origem a este trabalho.

Ao Prof. Luiz de Queiroz Orsini, pela orientação ao longo desses anos todos.

Aos amigos Normonds Alens e Max Gerken, que não chegaram a ver este trabalho.

RESUMO

Esta tese apresenta um novo modelo para dois algoritmos para filtragem adaptativa baseados em momentos de quarta ordem: o *least-mean fourth* (LMF) e o *least-mean mixed-norm* (LMMN). A originalidade do modelo apresentado aqui é que não se procura calcular a média quadrática do erro de estimação do algoritmo, mas sim a probabilidade do algoritmo ter um comportamento razoável (neste caso, convergir).

O trabalho mostra que o LMF e o LMMN não são estáveis na média quadrática se o regressor não for estritamente limitado (como ocorre, por exemplo, para a distribuição Gaussiana). Mesmo para a distribuição Gaussiana o LMF e o LMMN sempre têm uma probabilidade não nula de divergir, não importa quão pequeno seja o passo de adaptação. Esse resultado é demonstrado para um filtro escalar (com um único coeficiente) com regressor com uma distribuição normal modificada, e verificado através de várias simulações.

Além disso, é fornecido um limite superior para a probabilidade de divergência do LMF (e do LMMN), em função do comprimento do filtro, da potência dos sinais de entrada, do passo de adaptação, da variância do erro ótimo, para o caso de regressores Gaussianos. Os resultados apresentados aqui fornecem ferramentas para projetistas entenderem melhor o funcionamento do algoritmo LMF, e decidir quando é ou não conveniente o seu uso para uma dada aplicação.

ABSTRACT

This dissertation presents a new model for two adaptive filtering algorithms based on fourth-order moments: the least-mean fourth (LMF) and the least-mean mixed-norm (LMMN) algorithms. The novelty of the new model is its emphasis on computing the probability of a reasonable performance of a single realization of the algorithm (in this case, convergence), instead of looking for average performance indices such as mean-square error.

We show that the least-mean fourth (LMF) adaptive algorithm is not mean-square stable when the regressor input is not strictly bounded (as happens, for example, if the input has a Gaussian distribution). For input distributions with infinite support, even for the Gaussian distribution, the LMF has always a nonzero probability of divergence, no matter how small the step-size is chosen. We prove this result for a slight modification of the Gaussian distribution in an one-tap filter, and corroborate our findings with several simulations.

In addition, we give an upper bound for the probability of divergence of LMF as a function of the filter length, input power, step-size, and noise variance, for the case of Gaussian regressors. Our results provide tools for designers to better understand the behavior of the LMF algorithm, and decide on the convenience or not of its use for a given application.

SUMÁRIO

Lista de Figuras	iii
Lista de Tabelas	v
1 Introdução	6
1.1 Filtros Adaptativos	8
1.2 Modelamento Teórico de Filtros Adaptativos	13
1.3 Objetivos	19
2 Análise Tradicional do LMF e do LMMN	23
2.1 Análise do LMF	25
2.1.1 Primeiros resultados: [1]	25
2.1.2 Prova de Estabilidade: [2]	27
2.1.3 Análises sem Linearizações: [3, 4, 5, 6, 7, 8, 9, 10]	28
2.2 Análise do LMMN	33
3 O Mecanismo de Instabilidade no LMF	36
3.1 Um Exemplo Simples de Instabilidade	37
3.1.1 Prova de instabilidade para filtros escalares	38
3.2 Regressores Gaussianos	39
4 Probabilidade de Divergência no LMF	44

4.1	Recursão para $\ \mathbf{V}(n)\ ^2$	45
4.2	Estimando a Probabilidade de Divergência	47
4.3	Cálculo de $\hat{y}(n)$	51
4.4	Probabilidade de Divergência para o LMMN	52
5	Simulações	55
5.1	Simulações com o LMF	55
5.2	Simulações com o LMMN	58
5.3	Comparação com Modelos Tradicionais	59
6	Conclusões	64
	Apêndice A – Uma Aproximação para $(\mathbf{V}(n)^T \mathbf{X}(n))^{2k}$	67
	Apêndice B – Prova de que $D(n) \geq 0$	72
	Referências	75

LISTA DE FIGURAS

1.1	Curva de aprendizado de filtro LMS de comprimento 1.	15
1.2	Curva de aprendizado de filtro LMS de comprimento 1.	16
1.3	Curvas de aprendizado de filtro LMS de comprimento 1.	17
1.4	Curvas de aprendizado de filtro LMS de comprimento 1.	17
1.5	Saída de filtro LMS de comprimento 1.	18
1.6	Saída de filtro LMS de comprimento 128.	19
1.7	Saída de filtro LMS de comprimento 10.	20
2.1	Pontos de equilíbrio de (2.7), caso 1: ($y_c \ll y(0)_{max}$).	31
2.2	Pontos de equilíbrio de (2.7), caso 2: ($y_c = y(0)_{max}$).	32
2.3	Pontos de equilíbrio de (2.7), caso 3: (y_c and $y(0)_{max}$ complexos).	33
3.1	Três realizações do LMF com $M = 10$	36
3.2	Três realizações do LMF com regressor escalar.	43
5.1	Probabilidade de divergência, erro ótimo Gaussiano.	57
5.2	Probabilidade de divergência, regressores LA.	58
5.3	Probabilidade de divergência, regressores LA.	59
5.4	Probabilidade de divergência como função da potência do sina.	60
5.5	Probabilidade de divergência para o LMF, regressores IND.	61
5.6	Probabilidade de divergência para o LMMN, regressores LA.	61
5.7	Probabilidade de divergência para o LMMN, regressores LA.	62

5.8	Probabilidade de divergência para o LMMN, regressores LA. . . .	62
A.1	Cálculo de $\alpha_3(k)$	68

LISTA DE TABELAS

2.1	Valores de $M(1)/M(p)$ para várias distribuições de $e_0(n)$	26
3.1	Probabilidade de divergência observada para vários valores de passo.	42
5.1	Valores máximos μ_{\max} para o passo de adaptação do LMF.	63

1 INTRODUÇÃO

Um filtro adaptativo é essencialmente um estimador de parâmetros com duas características: recursividade (para permitir o acompanhamento de sistemas variantes no tempo), e simplicidade (para permitir a operação em tempo real a um custo aceitável). O grande interesse atual em pesquisa sobre filtragem adaptativa, verificado facilmente pela variedade de artigos sobre o assunto em revistas de importância, decorre principalmente de:

1. Grande utilização de filtragem adaptativa, principalmente em telecomunicações (equalização de canais, cancelamento de eco — acústico e de linha —, e mais atualmente em antenas inteligentes), mas também em cancelamento de interferências (como em medidas de sinais biológicos), identificação de sistemas para aplicação em controle adaptativo, cancelamento ativo de ruído, dentre outras (ver, por exemplo, [11, 12, 13, 14, 15, 16, 17, 18, 19, 20]);
2. Necessidade de filtros com muitos parâmetros em diversas aplicações (notadamente em cancelamento de eco acústico), o que aumenta a complexidade dos algoritmos;
3. Necessidade, na grande maioria das aplicações, de algoritmos rápidos e simples (isto é, que não demandem uma capacidade computacional muito elevada), de maneira a economizar tempo de processamento, energia e memória;

4. Necessidade de acompanhar fielmente a variação (que pode ser relativamente rápida) dos parâmetros que se deseja estimar.

A necessidade de filtros longos (com muitos parâmetros) em geral piora o desempenho dos algoritmos nos quesitos 3 e 4. Este fato, mais a sua natureza não-linear, estocástica e variante no tempo, fazem da análise e do desenvolvimento de algoritmos para filtragem adaptativa problemas de difícil solução. Dada a complexidade dos problemas envolvidos, o desenvolvimento de novos métodos de análise para filtros adaptativos é sempre muito bem-vindo, pois uma melhor compreensão do seu funcionamento facilita tanto o projeto de novos algoritmos, quanto a escolha do método mais adequado para cada situação específica (para ver a grande diversidade de algoritmos ou variantes de algoritmos existentes e propostos a cada ano, pode-se consultar, por exemplo, [12] e as últimas edições das revistas *IEEE Transactions on Signal Processing*, *IEEE Signal Processing Letters* e *Signal Processing*).

Esta tese descreve os trabalhos realizados pelo autor a respeito da análise teórica de uma família de algoritmos para filtragem adaptativa, que procura obter um desempenho melhor (comparado a algoritmos tradicionais) através do uso de funções-custo baseadas em momentos de ordem elevada. O maior foco do trabalho está no algoritmo *least-mean fourth* (LMF), proposto inicialmente por Walach e Widrow em 1984 [1]. A grande maioria dos resultados apresentados aqui foi apresentada em congressos, como o *IEEE International Conference on Acoustics, Speech, and Signal Processing* [10,21], o *Simpósio Brasileiro de Telecomunicações* [22], e também em um artigo aceito para publicação no *IEEE Transactions on Signal Processing* [23] (há também um outro artigo em preparação). O trabalho foi realizado em colaboração com o Prof. Dr. José Carlos M. Bermudez, da UFSC, e em parte também com seu ex-aluno Pedro I. Hübscher.

1.1 Filtros Adaptativos

De modo geral, um filtro adaptativo procura encontrar uma estimativa para uma variável $d(n)$ (*signal desejado*) em função de uma seqüência conhecida de *regressores*, $\{\mathbf{X}(k)\}_{k=0}^n$ (os $\mathbf{X}(n)$ são vetores M -dimensionais, podendo ser reais ou complexos),

$$\hat{d}(n) = h(\mathbf{X}(n), \mathbf{X}(n-1), \dots, \mathbf{X}(0)) \approx d(n).$$

Os elementos do vetor $\mathbf{X}(n)$ formam normalmente uma linha de atraso, isto é, $\mathbf{X}(n)$ é usualmente formado como abaixo (o símbolo T representa a operação de transposição de vetores e matrizes):

$$\mathbf{X}(n) = \begin{bmatrix} x(n) & x(n-1) & \dots & x(n-M+1) \end{bmatrix}^T.$$

Dependendo da aplicação, o sinal desejado $d(n)$ pode ser inteiramente conhecido (como em cancelamento ativo de ruído, cancelamento de eco e em controle adaptativo), pode ser conhecido apenas durante uma fase de aprendizado (equalização tradicional), ou pode ser desconhecido (equalização cega).

Atualmente na maioria das aplicações $h(\cdot)$ é uma função linear, mas há diversos trabalhos propondo o uso de funções não-lineares, por exemplo para controle ativo de ruído [24,25]. No caso linear, o filtro adaptativo procurará encontrar uma combinação linear $\mathbf{W}(n)^T \mathbf{X}(n)$ que seja a melhor aproximação de $d(n)$ segundo um critério previamente escolhido. Um dos critérios mais utilizados é o do mínimo erro médio quadrático, ou seja, o filtro busca minimizar o valor esperado $E e(n)^2$, onde $e(n) = d(n) - \hat{d}(n)$ é o erro de estimação. Outros critérios podem também ser usados: uma forma de se projetar novos algoritmos com características interessantes é alterar adequadamente a função-custo [11]. Por exemplo, em equalização cega procura-se utilizar critérios que não dependam de $d(n)$, mas apenas de al-

gumas de suas propriedades estatísticas (um exemplo importante é o critério do módulo constante, estudado para aplicação em telefonia celular [26, 27]).

Grande parte dos filtros adaptativos usados atualmente pode ser descrita pela recursão [28, 29]

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu f(e(n)) \mathbf{Z}(n),$$

em que μ é uma constante conhecida como passo de adaptação, $f(\cdot)$ é uma função conhecida, e $\mathbf{Z}(n)$ é um vetor que depende das seqüências $\{d(k)\}_{k=0}^n$ e $\{\mathbf{X}(k)\}_{k=0}^n$. Por exemplo, para $f(a) = a$ e $\mathbf{Z}(n) = \mathbf{X}(n)$ obtém-se o algoritmo *least-mean squares* (LMS):

$$\begin{aligned} \hat{d}(n) &= \mathbf{W}(n)^T \mathbf{X}(n), & e(n) &= d(n) - \hat{d}(n), \\ \mathbf{W}(n+1) &= \mathbf{W}(n) + \mu \mathbf{X}(n) e(n). \end{aligned} \tag{1.1}$$

Uma variante muito utilizada é o LMS normalizado (ou NLMS), obtido com o mesmo $f(a) = a$ e com $\mathbf{Z}(n) = \mathbf{X}(n)/(\epsilon + \|\mathbf{X}(n)\|^2)$ ($\|\mathbf{a}\|$ é a norma euclídeana do vetor \mathbf{a} , e ϵ é uma constante não-negativa).

$$\begin{aligned} \hat{d}(n) &= \mathbf{W}(n)^T \mathbf{X}(n), \\ e(n) &= d(n) - \hat{d}(n), \\ \mathbf{W}(n+1) &= \mathbf{W}(n) + \mu \frac{\mathbf{X}(n)}{\epsilon + \|\mathbf{X}(n)\|^2} e(n). \end{aligned} \tag{1.2}$$

Outro algoritmo de grande interesse, o *recursive least squares* (RLS) também pode ser obtido da expressão acima com escolhas convenientes de $f(\cdot)$ e de $\mathbf{Z}(n)$ [29].

$$\begin{aligned} \hat{d}(n) &= \mathbf{W}(n)^T \mathbf{X}(n), \\ e(n) &= d(n) - \hat{d}(n), \\ \mathbf{W}(n+1) &= \mathbf{W}(n) + \frac{\lambda^{-1} P(n) \mathbf{X}(n)}{1 + \lambda^{-1} \mathbf{X}(n)^T P(n) \mathbf{X}(n)} e(n), \\ P(n+1) &= \lambda^{-1} P(n) - \lambda^{-2} \frac{P(n) \mathbf{X}(n) \mathbf{X}(n)^T P(n)}{1 + \lambda^{-1} \mathbf{X}(n)^T P(n) \mathbf{X}(n)}. \end{aligned} \tag{1.3}$$

O parâmetro $0 \ll \lambda \leq 1$ é chamado fator de esquecimento do RLS.

No caso geral, o passo de adaptação μ é um parâmetro de projeto do filtro. É fácil notar que um passo de adaptação pequeno resulta num filtro que varia lentamente (o vetor $\mathbf{W}(n)$ pode variar apenas por uma quantidade pequena, proporcional a μ , a cada instante). Uma taxa de adaptação muito pequena (ou seja, um valor de μ muito próximo de zero) levará a uma taxa de convergência muito baixa, e a um filtro com dificuldade de seguir processos que variem rapidamente com o tempo [12, 30, 31, 32]; enquanto que um passo grande demais pode tornar o filtro muito sensível a ruído ou mesmo instável.

Em geral é necessário ter a convergência mais rápida possível [32]: no caso de equalizadores para telefonia, enquanto não há convergência não se pode transmitir sinal útil; no caso de aplicações em sinais biológicos, convergência rápida significa menos tempo gasto na coleta de dados (que pode envolver situações desconfortáveis para os pacientes). Para sinais não estacionários, a possibilidade de se utilizar convergência rápida possibilita o acompanhamento de processos que variam mais rapidamente no tempo.

O problema fundamental da filtragem adaptativa é desenvolver um algoritmo que possua, ao mesmo tempo, uma grande imunidade a ruído, seja numericamente robusto, tenha uma taxa de convergência elevada, e requeira poucas operações aritméticas a cada iteração. Os principais algoritmos derivados do LMS tendem a ser numericamente robustos e de baixa complexidade, mas são lentos, principalmente se a matriz de autocorrelação de $\mathbf{X}(n)$, $R_x = E(\mathbf{X}(n)\mathbf{X}(n)^T)$, tiver autovalores muito díspares (com amplitudes muito diferentes [12], o que acontece se os elementos da seqüência $\{x(n)\}$ de que é derivado $\mathbf{X}(n)$ tiverem alta correlação entre si).

O algoritmo RLS, por outro lado, possui uma taxa de convergência bastante rápida, mas apresenta problemas de instabilidade numérica e alto custo compu-

tacional, sendo necessário implementá-lo com muito cuidado [12, 33, 34]. Mesmo implementando com cuidado o algoritmo RLS, sua capacidade de acompanhar processos não estacionários nem sempre é adequada: como mostra [30], o RLS tem em algumas situações desempenho pior do que o do LMS nesse quesito.

Em algumas aplicações, em especial em cancelamento de eco acústico, ainda é difícil escolher um algoritmo adequado. Para cancelamento de eco acústico são necessários filtros normalmente muito longos (e portanto a complexidade computacional do algoritmo é um fator importante). Filtros longos tendem a convergir lentamente (pois para garantir a estabilidade dos filtros, o passo de adaptação deve ser aproximadamente inversamente proporcional ao comprimento do filtro), um problema que é agravado pelo fato da disparidade dos autovalores da matriz R_x ser bastante alta para os sinais de voz que precisam ser cancelados nessa aplicação [35].

A busca de algoritmos para filtragem adaptativa que sejam fáceis de implementar, robustos, rápidos, que rejeitem bem o ruído, e que acompanhem eficientemente processos não estacionários continua, portanto, sendo um problema atual. Vários algoritmos vêm sendo propostos para este fim:

1. Algoritmos de passo variável, que buscam variar o passo de adaptação de acordo com a situação (passo grande se alguma medida indicar distância do ponto ótimo — em geral são usadas estimativas da correlação entre o sinal de entrada e o erro de estimação) [36, 37];
2. Métodos que buscam descorrelacionar o sinal de entrada (como o algoritmo RLS), mas procurando minimizar problemas numéricos, como o algoritmo rápido de Newton [38], o algoritmo de projeções afins [39, 40] e algumas versões do RLS rápido [33];
3. O “LMS Wiener” [41], que procura melhorar o desempenho de seguimento

de parâmetros variáveis usando conhecimento prévio da estatística do sinal;

4. Algoritmos para filtros esparsos (em que vários coeficientes são nulos) [42, 43, 44];
5. Algoritmos com funções-custo modificadas, como o *least-mean fourth* (LMF) [6, 10];
6. Algoritmos adaptativos em subbandas [45, 46];
7. Algoritmos que usam funções ortonormais (como funções de Laguerre) no lugar de atrasos para gerar os filtros adaptativos, de maneira a diminuir o número de parâmetros do filtro [47, 48].
8. Algoritmos “aceleradores”, que utilizam diferenças de segunda ordem para a adaptação do vetor de parâmetros [49, 50, 51, 52, 53] (além de um artigo em fase de preparação).

Cada uma dessas propostas apresenta um compromisso diferente entre velocidade de convergência e complexidade, e funciona para condições diferentes.

Para se utilizar um algoritmo nas melhores condições, é importante se conhecer como a escolha dos parâmetros de projeto do algoritmo (em muitos casos, apenas o passo de adaptação μ) afeta o comportamento de um filtro adaptativo:

- Para qual intervalo de valores (μ_{\min}, μ_{\max}) do passo o filtro será estável?
- Qual é o passo que garante convergência mais rápida?

Mesmo quando o próprio passo é escolhido adaptativamente é necessário se conhecer a faixa (μ_{\min}, μ_{\max}), para garantir que o filtro resultante seja estável (ver por exemplo [54, pág. 72]).

1.2 Modelamento Teórico de Filtros Adaptativos

O estudo teórico de filtros adaptativos tradicionalmente se baseia na obtenção da curva de aprendizado do filtro em função de seus parâmetros de projeto. A curva de aprendizado (isto é, o gráfico de $E e(n)^2$ em função de n) fornece uma medida da velocidade e da precisão com que um filtro adaptativo pode reagir a variações em suas entradas. Grande parte dos trabalhos sobre análise teórica de filtros adaptativos procuram calcular a curva de aprendizado de um determinado filtro em função das propriedades estatísticas dos sinais de entrada $\mathbf{X}(n)$ e $d(n)$. No entanto, como será visto a seguir, mesmo o conhecimento exato da curva de aprendizado de um algoritmo pode não ser suficiente para descrever adequadamente o seu comportamento.

Para melhor compreender os vários resultados teóricos disponíveis sobre FAs, é conveniente dividir os métodos de análise existentes em 4 categorias:

- a) Análise na média quadrática,
- b) Estabilidade quase-certa (com probabilidade 1),
- c) Análise determinística,
- d) Análise experimental ou por simulações.

Cada método presta-se a situações e aplicações distintas, e, como será visto a seguir, complementa os resultados que podem ser obtidos com os outros três. Assim, o método (a) — análise na média quadrática — avalia o desempenho *médio* de um FA (medido usualmente através de $E e(n)$ e de $E e(n)^2$). Por sua vez, o método (b) permite dizer basicamente se um FA converge ou não na ausência de

ruído¹ — agora não se trata mais de um comportamento médio, mas do comportamento de um filtro para a quase totalidade dos sinais de entrada admissíveis (quase totalidade no sentido probabilístico). O método (c) procura identificar situações-limite, garantindo que um filtro terá comportamento satisfatório mesmo para o pior sinal de entrada e o pior ruído possíveis. Finalmente, o método (d) — simulação — permite que se verifiquem as previsões dos métodos anteriores, ou que se obtenham resultados em situações em que uma análise teórica não é possível com o atual estado da técnica.

Os quatro métodos possibilitam a obtenção de muitos resultados úteis, mas atualmente têm algumas limitações importantes que são listadas a seguir. Análises teóricas segundo (a) ou (b) só existem para casos muito particulares [55,32,56,57], ou com a aproximação de adaptação “suficientemente lenta” (ou seja, para filtros com passo de adaptação infinitesimal [28, 54, 29, 57]). Expressões para quão pequeno deve ser o passo para as análises serem válidas são raras e extremamente difíceis de serem calculadas [58, 59].

O método (c) - análise determinística — não sofre da limitação descrita acima, mas tende a apresentar resultados demasiadamente conservadores — os resultados são do tipo “pior caso”, tendo utilidade reduzida em grande parte das aplicações (exceto em controle automático, onde a análise de pior caso é muitas vezes a mais adequada [60, 61]). O método (d) tem as desvantagens óbvias de ser por demais trabalhoso, e de não permitir generalizações de resultados e previsões para situações novas, mas pode ser utilizado com praticamente qualquer modelo de sinal de entrada. Simulações são muito usadas para validar resultados teóricos, ou mesmo para extrair informações sobre a taxa de convergência ou o erro residual (em regime permanente) de um filtro adaptativo em situações para as quais não existem resultados teóricos.

¹ O método (b) pode ser estendido para fornecer o desempenho médio de filtros para o caso de passo infinitesimal [29].

Para aplicações que precisam de análises estatísticas (como são a maioria das situações em, por exemplo, Telecomunicações e Engenharia Biomédica), há poucos resultados que sejam aplicáveis para passos não infinitesimais (ou seja, sem a restrição de “convergência lenta”). Dentre os poucos resultados disponíveis, destacam-se [55, 56]. Infelizmente, estes resultados só podem ser utilizados com filtros pequenos (até 8 coeficientes), enquanto que em situações práticas é comum ser necessário utilizar filtros com milhares de coeficientes.

Para $\mu \approx 0$ os métodos (a) e (b) fornecem boas aproximações para as curvas de aprendizado, e estas predizem bem o comportamento real do filtro (ver fig. 1.1). Este resultado empírico foi justificado em [62] para o LMS, onde se prova que, no limite para adaptação lenta ($\mu \approx 0$), a taxa de convergência média (calculada usando o método (a)) do algoritmo LMS é igual à taxa de convergência de uma realização típica do mesmo algoritmo (calculada usando o método (b)).

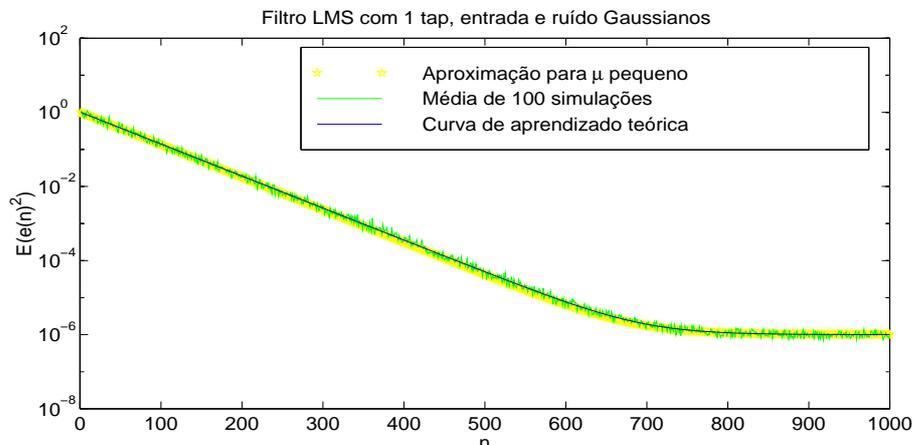


Figura 1.1: Curva de aprendizado de filtro LMS de comprimento 1 com passo $\mu = 0,01$. As entradas do filtro, $\{d(n)\}$ e $\{x(n)\}$, são processos aleatórios Gaussianos brancos, com $E x(n)^2 = 1$. Neste caso é possível calcular exatamente a curva de aprendizado teórica (representada em traço mais escuro). A aproximação para $\mu \approx 0$ está no traço mais claro, e a média de 100 simulações independentes pode ser vista no tom intermediário.

Nada disso vale para velocidades de adaptação maiores. Isto foi provado recentemente em [63], onde se demonstra que (para o algoritmo LMS com passo de adaptação “grande”), a evolução do erro $e(n)^2$ é bem explicada pela análise na

média quadrática apenas para os instantes iniciais da convergência, enquanto que a análise em probabilidade 1 explica melhor os efeitos observados após algumas iterações do algoritmo adaptativo (ver figs. 1.2–1.4).

A fig. 1.2 mostra a curva de aprendizado do mesmo filtro da fig. 1.1, mas com passo $\mu = 1/6$. Pode-se notar agora que a curva obtida com a aproximação para passos pequenos (curva suave tracejada) já apresenta um erro considerável para a curva teórica (curva suave em traço contínuo). No entanto, a curva teórica é ainda bem aproximada pela média de apenas 100 simulações (o que significa que as realizações do filtro tendem a ficar próximas à curva média).

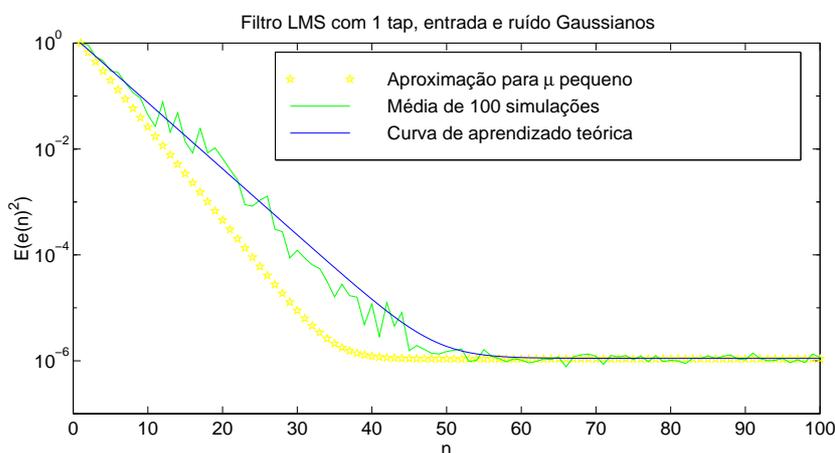


Figura 1.2: Curva de aprendizado de filtro LMS de comprimento 1 com passo $\mu = 1/6$. As entradas do filtro, $\{d(n)\}$ e $\{x(n)\}$, são como na fig. 1.1. A curva de aprendizado teórica está representada em traço contínuo escuro, a aproximação para $\mu \approx 0$ está em traço mais claro, tracejado, e a média de 100 simulações independentes está em traço contínuo claro.

Para passos ainda maiores, o comportamento típico de um filtro pode não ser bem representado pela curva de aprendizado média (veja figs. 1.3 e 1.4). Na fig. 1.3, as curvas inferiores suave e “acidentada” representam respectivamente a curva de aprendizado teórica e a média de 100 simulações, agora para $\mu = 1/3$ (que é a situação de máxima taxa de convergência de $Ee(n)^2$). As curvas da figura anterior foram repetidas para comparação. Note como agora as simulações tendem a ficar *abaixo* da curva teórica (este efeito, que ainda é pequeno, fica mais pronunciado à medida que se aumenta o passo de adaptação).

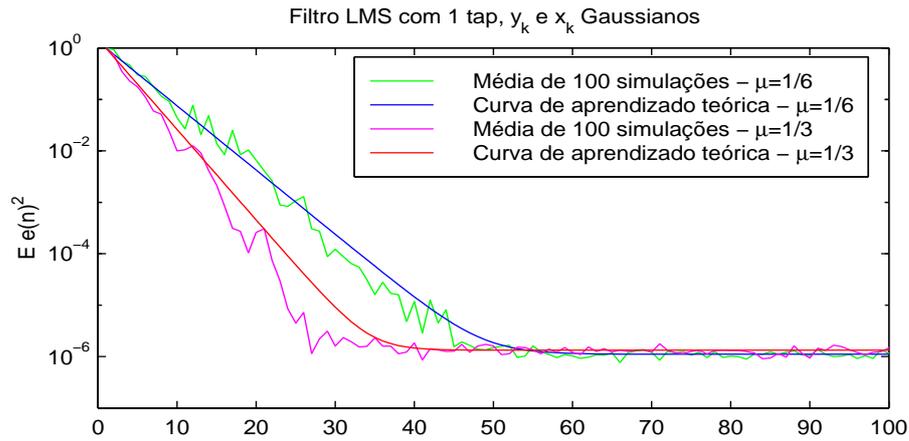


Figura 1.3: Curvas de aprendizado de filtro LMS de comprimento 1 com passos $\mu = 1/6$ e $1/3$. As entradas do filtro, $\{d(n)\}$ e $\{x(n)\}$, são como na fig. 1.1.

Na fig. 1.4, pode-se ver a curva de aprendizado teórico (curva suave superior) para passo $\mu = 0,5$, e a média de 100 simulações com este passo na curva mais abaixo de todas, em traço claro. Note como a taxa de convergência “teórica” é igual à do caso $\mu = 1/6$. No entanto, a convergência observada nas simulações foi muito mais rápida, como mostra a curva em traço claro, que está bastante próxima à curva obtida com $\mu = 1/3$.

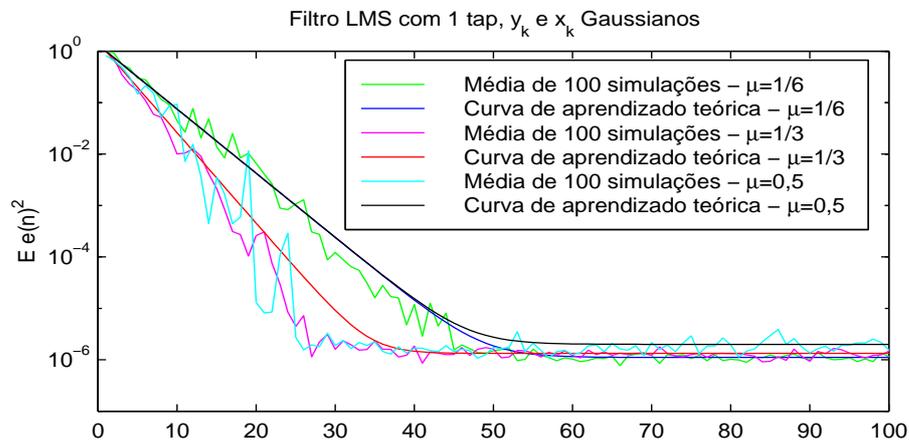


Figura 1.4: Curvas de aprendizado de filtro LMS de comprimento 1 com passos $\mu = 1/6$, $1/3$ e $1/2$. As entradas do filtro, $\{d(n)\}$ e $\{x(n)\}$, são como na fig. 1.1.

A diferença entre a curva de aprendizado teórico e as simulações para $\mu = 1/2$ é explicada pelo comportamento diferente das taxas de convergência *na média quadrática* e *com probabilidade 1* quando o passo de adaptação é aumentado. De fato, [63] prova que a primeira (taxa de convergência na média quadrática)

descreve bem os instantes iniciais da convergência do filtro, enquanto que após um certo número de iterações o filtro tenderá a convergir com a taxa predita pela análise em probabilidade 1. Este efeito é bem ilustrado na fig. 1.5: note que o filtro da fig. 1.5 *diverge* na média quadrática, e nas iterações iniciais ($0 \leq n \leq 10$), o erro $e(n)^2$ realmente aumenta (a curva correspondente à simulação permanece próxima à curva de $E e(n)^2$ teórica, acima). Para $10 < n \leq 30$, no entanto, o erro $e(n)^2$ passa a *diminuir*, com uma taxa próxima à taxa de convergência com probabilidade 1 (curva inferior). Do instante $n = 40$ em diante, o filtro está em regime.

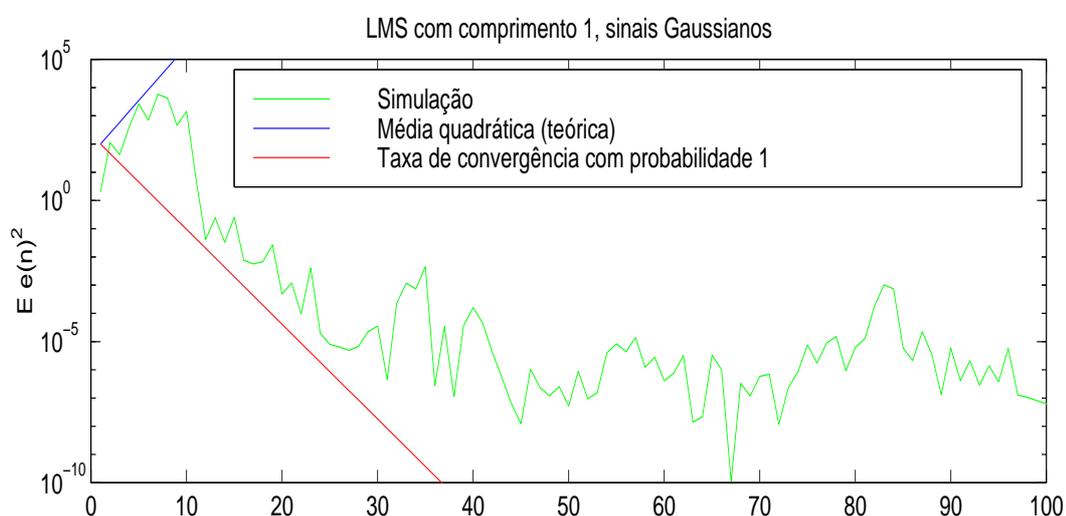


Figura 1.5: Saída de filtro LMS de comprimento 1 com passo $\mu = 0,5$. As entradas do filtro, $\{d(n)\}$ e $\{x(n)\}$, são como na fig. 1.1. A curva central é uma única realização do filtro, não uma média. A curva superior é $E e(n)^2$, e a curva superior representa a taxa de convergência obtida com a análise em probabilidade 1.

Como mostram os exemplos acima, o comportamento de um filtro adaptativo com passo *não* infinitesimal não é necessariamente bem descrito nem por uma análise na média quadrática nem por uma análise de convergência em probabilidade 1. Na verdade, esses dois métodos de análise são *complementares*, pelo menos para o algoritmo LMS. Em algumas situações, um filtro *estável* com probabilidade 1 (ou quase-certamente estável) pode não apresentar desempenho adequado (na simulação da fig. 1.4, $e(n)^2$ cresce de 1 até valores da ordem de 10^4

antes de convergir e ficar em torno de 10^{-4}); enquanto que em outros casos um filtro *instável* na média quadrática pode convergir na grande maioria das vezes, e apresentar desempenho razoável (ver fig. 1.6). Estes fatos podem ser mascarados em simulações. De fato, médias de simulações podem em algumas situações apresentar resultados bastante diferentes dependendo do número de simulações usado para se aproximar a média $Ee(n)^2$, como se vê nas figs 1.6 e 1.7.

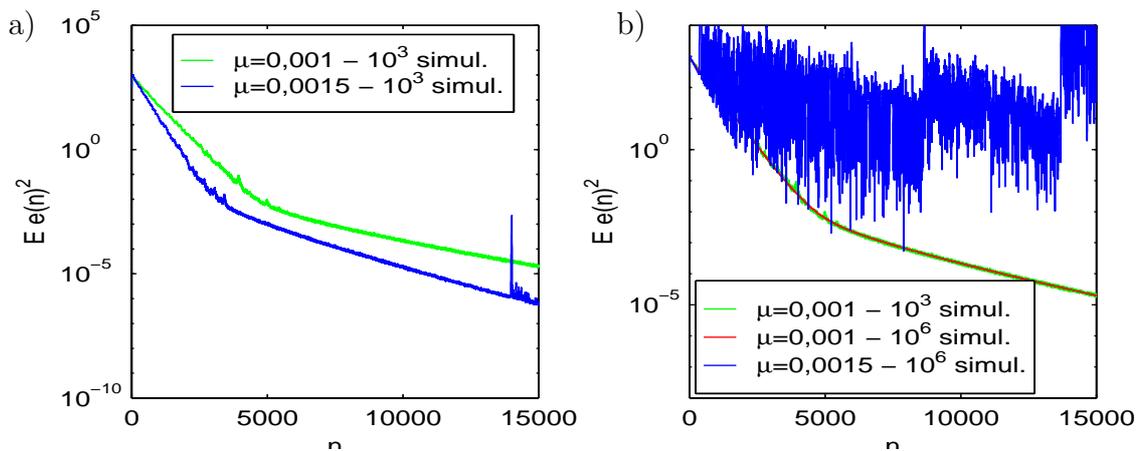


Figura 1.6: Saída de filtro LMS de comprimento 128. A entrada $\{\mathbf{x}_n\}$ é um processo estocástico correlacionado (filtro AR de 2ª ordem com ruído branco Gaussiano na entrada). Na figura (a) à esquerda, a curva inferior é a média de 1.000 simulações do filtro com passo $\mu = 0,0015$, enquanto que a curva superior é a média de também 1.000 simulações com $\mu = 0,001$. Na figura à direita, temos acima a média de 10^6 simulações do filtro com passo $\mu = 0,0015$, mostrando divergência. Para comparação, está em abaixo a média de 10^6 simulações com passo $\mu = 0,001$, sobreposta à média de 1.000 simulações com o mesmo passo.

1.3 Objetivos

O objetivo principal deste trabalho é apresentar uma forma de estender os resultados mencionados na seção anterior, obtidos para o algoritmo LMS em [63], para algoritmos baseados em momentos de ordem superior, em particular o LMF.

Filtros adaptativos supervisionados baseados em momentos de quarta ordem foram propostos no início da década de 1980, com o objetivo de obter um desempenho melhor (maior velocidade de convergência e/ou menor desajuste em regime)

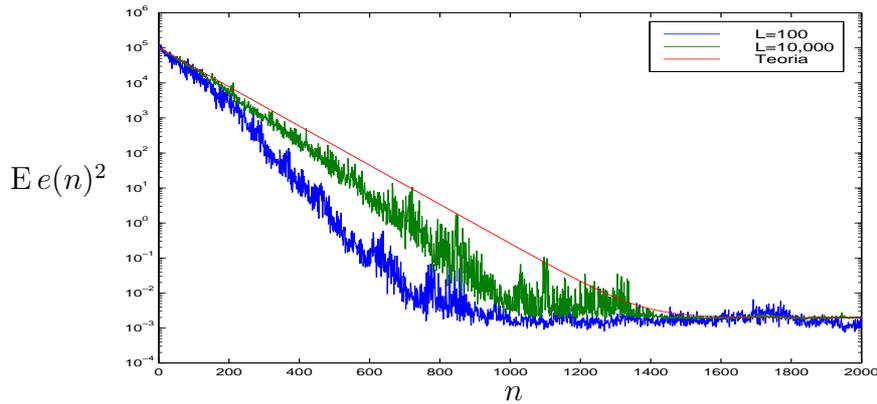


Figura 1.7: Saída de filtro LMS de comprimento 10. A entrada $\{\mathbf{X}(n)\}$ é um processo estocástico branco Gaussiano, e o passo de adaptação é $\mu = 0,15$. A curva inferior é a média de 100 simulações, e a curva central, de 10^4 simulações. Em superior está a curva $E e(n)^2$ teórica.

do que o tradicional algoritmo LMS, sem chegar à complexidade computacional do RLS. O primeiro exemplo de filtro supervisionado baseado em momentos de quarta ordem foi o *least-mean fourth* (LMF), que é um algoritmo de gradiente, bastante semelhante ao LMS, mas baseado na minimização do valor esperado da quarta potência do erro [1,12]. O LMF calcula novas aproximações para um vetor de parâmetros segundo a recursão

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu e(n)^3 \mathbf{X}(n), \quad (1.4)$$

em que $e(n) = d(n) - \mathbf{W}(n)^T \mathbf{X}(n)$, $\mathbf{W}(n) \in \mathbb{R}^M$ e $\mathbf{X}(n) \in \mathbb{R}^M$.

Este algoritmo é motivado pela minimização da função-custo $E e(n)^4$, usando o método do gradiente estocástico para a minimização. O trabalho original [1] na verdade propõe toda uma família de algoritmos, usando uma potência p -ésima do erro,

$$\text{sinal}(e(n)) |e(n)|^p, \quad (1.5)$$

para $p \geq 1$, e mostra que filtros baseados em momentos de ordem elevada podem ser vantajosos se o erro de estimação ótimo for sub-gaussiano². No entanto, o

²O erro de estimação ótimo é definido no capítulo 2, pág. 23.

Uma variável aleatória v é *sub-gaussiana* se a sua *curtose*,

$$\kappa(v) = E\{v^4\} - 3E^2\{v^2\},$$

mesmo trabalho já adverte para possíveis problemas de estabilidade de filtros de ordem elevada, em especial para a dependência da estabilidade com a condição inicial do filtro. Mais recentemente foram propostos modelos para esta dependência, evidenciando a importância de se escolher uma boa condição inicial para garantir a estabilidade do LMF [4, 6, 10, 9].

Na década de 1990 foi sugerida uma solução intermediária entre o LMS e o LMF, denominada (com uma certa falta de imaginação) de algoritmo *least-mean mixed-norm* (LMMN) [64], com a recursão

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu(\lambda e(n) + (1-\lambda)e(n)^3)\mathbf{X}(n), \quad (1.6)$$

em que $0 \leq \lambda \leq 1$ é um parâmetro que permite escolher entre um comportamento mais próximo do LMS ou do LMF. A motivação para a introdução desta alteração era a de aumentar a robustez do algoritmo.

Vários autores analisaram o comportamento médio quadrático do LMF e do LMMN, como por exemplo [65, 66, 3, 4, 6, 5, 7, 8, 9, 10, 64, 67], procurando aproximações para a curva de aprendizado dos algoritmos e para o valor máximo do passo de adaptação. Muitas destas referências assumem (ou admitem o caso) que o regressor $\mathbf{X}(n)$ é gaussiano, incluindo alguns trabalhos deste autor.

No entanto será demonstrado nesta tese que o comportamento do LMF e do LMMN na verdade não é estável na média quadrática para *nenhum* valor do passo de adaptação μ , se o regressor $\mathbf{X}(n)$ for gaussiano. Os dois algoritmos em questão, por conta da não-linearidade cúbica em sua recursão, têm sempre uma probabilidade não-nula de que a norma euclideana $\|\mathbf{W}(n)\|$ divirja se o regressor não for estritamente limitado, ou seja, se não houver um valor $B < \infty$ para o qual $\|\mathbf{X}(n)\| < B$ para todo n . Também será apresentada uma aproximação

for negativa. Note que a curtose de uma variável aleatória Gaussiana é nula, distribuições com caudas mais longas (como a exponencial, por exemplo) têm curtose positiva, e distribuições com caudas mais curtas (por exemplo, uniforme) têm curtose negativa.

(limite superior) para a probabilidade de divergência do LMF e do LMMN em função dos parâmetros dos filtros (μ e λ) e das distribuições de $\mathbf{X}(n)$ (suposto Gaussiano) e do ruído de medida.

Esses resultados mostram que o LMF e o LMMN têm realmente problemas importantes de robustez, e indicam maneiras de contorná-los através de uma escolha adequada dos parâmetros dos filtros (a probabilidade de divergência tende a zero quando o passo de adaptação tende a zero). Note-se que [66, 2] provam a estabilidade do LMF para o caso de regressor estritamente limitado e passo suficientemente pequeno. Este resultado, apesar de interessante, não indica se os filtros são estáveis ou não para regressores Gaussianos, e não fornece indicações sobre o tamanho do passo que pode ser usado com segurança.

É importante salientar que a análise para regressores Gaussianos não é apenas acadêmica: variáveis aleatórias Gaussianas são bastante concentradas em torno de sua média, por exemplo, a probabilidade de ocorrer um valor distante de mais de 3 desvios-padrão da média de uma variável Gaussiana é da ordem de $2,2 \times 10^{-5}$. Os resultados apresentados neste trabalho mostram que mesmo uma probabilidade extremamente pequena de ocorrer um valor elevado no regressor podem desestabilizar completamente o algoritmo, o que significa na prática que o algoritmo é bastante sensível a *outliers*. A proposta de um modelo para a relação entre o passo de adaptação e a probabilidade de divergência do algoritmo permite uma decisão mais bem-informada sobre a conveniência ou não do uso do LMF ou do LMMN em uma dada situação.

Como mencionado anteriormente, a maior parte dos resultados apresentados aqui foi apresentada em congressos [10, 21, 22], e será publicada em artigo (já aceito) no *IEEE Transactions on Signal Processing* [23].

2 ANÁLISE TRADICIONAL DO LMF E DO LMMN

Neste capítulo é apresentado um resumo dos principais resultados disponíveis na literatura sobre os algoritmos LMF e LMMN. O objetivo é mostrar o estado atual do modelamento teórico desses algoritmos, e apontar os problemas que motivaram este trabalho. Para não alongar demais o texto, não serão repetidas aqui as demonstrações dos resultados descritos, apenas fornecidas referências para os artigos originais.

Para referência futura é conveniente apresentar agora as equações de erro de ambos os algoritmos. Inicialmente, recorde que, dadas duas seqüências estocásticas $\{\mathbf{X}(n) \in \mathbb{R}^M\}$ e $\{d(n) \in \mathbb{R}\}$ conjuntamente estacionárias, o estimador ótimo linear de $d(n)$ dado $\mathbf{X}(n)$ (conhecido como *solução de Wiener*) é dado por [12, 11]

$$\mathbf{w}_* = \mathbf{R}_X^{-1} \mathbf{p},$$

em que $\mathbf{R}_X \triangleq E(\mathbf{X}(n)\mathbf{X}(n)^T)$ é a matriz de autocorrelação de $\mathbf{X}(n)$, e $\mathbf{p} \triangleq E(d(n)\mathbf{X}(n))$ é o vetor de correlação cruzada entre $\mathbf{X}(n)$ e $d(n)$ (supõe-se, como acontece na maioria dos casos, que \mathbf{R}_X tenha inversa). O estimador ótimo linear satisfaz a condição de ortogonalidade

$$E(e_0(n)\mathbf{X}(n)) = \mathbf{0},$$

em que $e_0(n) = d(n) - \mathbf{w}_*^T \mathbf{X}(n)$ é o erro de estimação ótimo, com variância σ_0^2 . Em outras palavras, dadas as seqüências $\{\mathbf{X}(n)\}$ e $\{d(n)\}$, sempre é possível

achar um modelo

$$d(n) = \mathbf{w}_*^T \mathbf{X}(n) + e_0(n), \quad (2.1)$$

em que o erro $e_0(n)$ não é correlacionado com $\mathbf{X}(n)$.

A análise dos algoritmos geralmente é simplificada se as recursões forem escritas em termos do vetor erro

$$\mathbf{V}(n) \triangleq \mathbf{w}_* - \mathbf{W}(n),$$

pois

$$\begin{aligned} e(n) &= d(n) - \mathbf{W}(n)^T \mathbf{X}(n) = \mathbf{w}_*^T \mathbf{X}(n) + e_0(n) - \mathbf{W}(n)^T \mathbf{X}(n) = \\ &= \mathbf{V}(n)^T \mathbf{X}(n) + e_0(n). \end{aligned}$$

Usando esta relação, as recursões do LMF (1.4) e do LMMN (1.6) podem ser reescritas como

1. LMF:

$$\mathbf{V}(n+1) = \mathbf{V}(n) - (\mathbf{V}(n)^T \mathbf{X}(n) + e_0(n))^3 \mathbf{X}(n), \quad (2.2)$$

2. LMMN:

$$\begin{aligned} \mathbf{V}(n+1) &= \mathbf{V}(n) - [\lambda(\mathbf{V}(n)^T \mathbf{X}(n) + e_0(n)) + \\ &\quad + (1-\lambda)(\mathbf{V}(n)^T \mathbf{X}(n) + e_0(n))^3] \mathbf{X}(n). \end{aligned} \quad (2.3)$$

Um parâmetro importante para medir o desempenho de um algoritmo adaptativo é o seu desajuste, ou seja, a razão entre a potência de $e(n) - e_0(n) = \mathbf{V}(n)^T \mathbf{X}(n)$ e a potência de $e_0(n)$,

$$D \triangleq \frac{\text{E} [(\mathbf{V}(n)^T \mathbf{X}(n))^2]}{\text{E}(e_0(n)^2)}, \quad (2.4)$$

considerando entradas estacionárias e que o filtro já atingiu o regime permanente.

2.1 Análise do LMF

Modelos estocásticos para o LMF começaram a ser propostos já em [1], que apresenta um modelo bastante simplificado para o comportamento médio e médio quadrático da família de algoritmos em que a potência do erro em (1.4) é $p \geq 1$ (veja (1.5)). Esse primeiro modelo dá a motivação para o uso do LMF, mostrando que o desajuste do algoritmo em regime permanente poderia ser menor do que o do LMS para o caso do erro ótimo $e_0(n)$ ter uma distribuição sub-gaussiana. A análise de estabilidade do algoritmo, no entanto, tem alguns erros. Modelos mais precisos foram sendo propostos ao longo dos anos, destacando-se os apresentados em [2, 66, 4, 6, 8, 9, 10] (os modelos descritos em [65, 7] também podem ser aplicados ao LMF). A seguir serão apresentadas as principais conclusões dos artigos mencionados acima, destacando-se as hipóteses feitas.

2.1.1 Primeiros resultados: [1]

Nessa referência, de 1984, Walach e Widrow usam as seguintes hipóteses:

1. Os vetores $\mathbf{W}(n)$ e $\mathbf{X}(n)$ são independentes entre si para todo n ,
2. O erro ótimo $e_0(n)$ é independente de $\mathbf{X}(m)$, para todo m, n ,
3. $\mathbf{X}(n)$ e $e_0(n)$ são simetricamente distribuídos em torno de $\mathbf{0}$ e 0 , de modo que $E e_0(n)^{2k+1} = 0$ para todo $k \in \mathbb{N}$, e analogamente para os elementos do vetor $\mathbf{X}(n)$,
4. O algoritmo já está em regime permanente, e o vetor $\mathbf{V}(n)$ mantém-se em uma bola $B = \{\mathbf{V} : \|\mathbf{V}\| < B_0 < \infty\}$ em torno da origem,
5. O passo de adaptação μ é pequeno o suficiente para que termos com potências de μ maiores que a primeira possam ser desprezados.

Nestas condições, os autores calculam uma aproximação para o valor do desajuste dos algoritmos, em função de p , e comparam esse desajuste com o do LMS (que corresponde ao caso $p = 1$) para o caso em que as constantes de tempo dos algoritmos são iguais, obtendo ($M(p)$ é o desajuste do algoritmo que usa potência p do erro de estimação)

$$\frac{M(1)}{M(p)} = \frac{(2p - 1)^2 \mathbb{E}(e_0(n)^2) [\mathbb{E}(e_0(n)^{2p-2})]^2}{\mathbb{E}(e_0(n)^{4p-2})}. \quad (2.5)$$

A tabela 2.1.1 fornece o valor da razão acima para vários valores de p e para diferentes distribuições de $e_0(n)$.

p	Gaussiana	Uniforme	Seno	Onda quadrada
2	0,6	2,3	3,6	9
3	0,24	3,67	7,14	25
4	0,08	5	11,4	49

Tabela 2.1: Valores de $M(1)/M(p)$ para várias distribuições de $e_0(n)$. [Reproduzida de [1].]

As constantes de tempo calculadas em [1] valem apenas para o regime permanente, isto é, equivalem à taxa de convergência do algoritmo para variações pequenas em torno do valor ótimo \mathbf{w}_* . O artigo também fornece uma prova de que os algoritmos são estáveis para passos suficientemente pequenos, e fornece uma estimativa para o valor do passo máximo. A demonstração não está correta, no entanto, porque a argumentação parte da hipótese 4 acima, que já é parte do que se pretende demonstrar (se $\mathbf{V}(n)$ se mantém em uma bola B finita, logicamente sua matriz de autocorrelação é finita, que é o que se entende por “estabilidade na média quadrática”).

Os resultados de [1] mostram que o LMF pode ser útil dependendo da distribuição do erro de estimação ótimo. O próprio artigo adverte porém que o algoritmo apresenta problemas de robustez, principalmente sensibilidade à condição inicial. Como toda a análise foi feita considerando-se $\mathbf{V}(n) \approx \mathbf{0}$, essa sensibilidade

às condições iniciais não foi quantificada.

Linearizando a recursão da matriz de autocorrelação de $\mathbf{V}(n)$, mostra-se em [65, 66] que, em regime,

$$\mathbb{E}[\mathbf{V}(n)\mathbf{V}(n)^T] \approx \frac{\mu \mathbb{E}(e_0(n)^6)}{6\sigma_0^2} \mathbf{I},$$

em que \mathbf{I} é a matriz identidade. A relação acima é válida para passo suficientemente pequeno e $\mathbf{X}(n)$ e $d(n)$ estritamente limitados.

2.1.2 Prova de Estabilidade: [2]

A demonstração precisa da estabilidade do LMF para passo suficientemente pequeno veio em 1992, com o artigo de Sethares [2]. Esse artigo demonstra que o algoritmo LMF é estável se o passo de adaptação for suficientemente pequeno, se $d(n)$ for estritamente limitado (isto é, se existir $B < \infty$ tal que $|d(n)| < B$ para todo n), e se os regressores produzirem uma *excitação persistente*, isto é, se existirem $0 < \alpha \leq \beta < \infty$ e $N < \infty$ tais que, para todo n ,

$$\alpha \mathbf{I} \leq \sum_{k=n}^{n+N} \mathbf{X}(k)\mathbf{X}(k)^T \leq \beta \mathbf{I},$$

em que \mathbf{I} é a matriz identidade de dimensão M , e a notação $\mathbf{A} \leq \mathbf{B}$ significa que a matriz $\mathbf{A} - \mathbf{B}$ é positiva semi-definida. Essa condição implica que o regressor é também estritamente limitado. O trabalho usa as mesmas ferramentas usadas para a análise determinística do LMS [68]. Os resultados são apenas locais, ou seja, prova-se que caso o erro ótimo $e_0(n)$ seja identicamente nulo, a origem $\mathbf{0}$ é um ponto de equilíbrio estável de (2.2). O resultado é estendido para o caso com ruído, provando-se que, nas condições acima e se a condição inicial $\mathbf{V}(0)$ for suficientemente próxima da origem, então $\mathbf{V}(n)$ permanecerá em uma vizinhança de $\mathbf{0}$ para todo n . Não são fornecidas aproximações para o valor-limite do passo de adaptação, nem para o tamanho da vizinhança de $\mathbf{0}$ em que o vetor erro

permanece para cada caso.

O resultado acima é interessante, pois demonstra que, sob condições adequadas, o algoritmo LMF deve funcionar corretamente. No entanto, a sua utilidade prática é pequena, pois as “condições adequadas” não são quantificadas.

2.1.3 Análises sem Linearizações: [3, 4, 5, 6, 7, 8, 9, 10]

As análises iniciais do LMF foram todas baseadas em linearizações do filtro em torno de algum ponto de operação: em [1], em torno do regime permanente (supondo-se a-priori que o filtro converge), em [2], em torno do ponto ótimo $\mathbf{V} = \mathbf{0}$. Para evitar linearizações os trabalhos tratados nesta seção em geral assumem que o regressor é Gaussiano, exceto [7, 8], que chegam a resultados válidos para regressores não Gaussianos, desde que sua matriz de autocorrelação seja múltipla da identidade. Serão descritos aqui os resultados mais recentes, de [9, 10].

A primeira destas referências propõe um modelo para a fase de aprendizado do LMF, usando as hipóteses 1–3 da seção 2.1.1, acrescidas das hipóteses

6. O regressor $\mathbf{X}(n)$ tem distribuição Gaussiana, com média nula e matriz de autocorrelação \mathbf{R}_X ,
7. O comprimento do filtro é longo o suficiente para que sejam válidas aproximações como

$$\mathbb{E}(\mathbf{V}(n)^T \mathbf{R}_X \mathbf{V}(n) \mathbf{V}(n) \mathbf{V}(n)^T) \approx \mathbb{E}(\mathbf{V}(n)^T \mathbf{R}_X \mathbf{V}(n)) \mathbb{E}(\mathbf{V}(n) \mathbf{V}(n)^T),$$

8. $\mathbf{V}(n)$ é pequeno o suficiente para que

$$\mathbb{E} \left[(\mathbf{V}(n)^T \mathbf{X}(n))^{2k} \mathbf{X}(n) \mathbf{X}(n)^T \right] \approx \mathbf{0}$$

se $k > 1$.

Usando essas hipóteses, [9] acha uma recursão aproximada para a matriz de autocorrelação de $\mathbf{V}(n)$, $\mathbf{K}(n) \triangleq \mathbb{E}(\mathbf{V}(n)\mathbf{V}(n)^T)$ como segue:

$$\begin{aligned} \mathbf{K}(n+1) \approx & \mathbf{K}(n) - 3\mu [\sigma_0^2 + \text{Tr}\{\mathbf{R}_X \mathbf{K}(n)\}] [\mathbf{R}_X \mathbf{K}(n) + \mathbf{K}(n) \mathbf{R}_X] + \\ & + \mu^2 \left\{ \mathbb{E}(e_0(n)^6) \mathbf{R}_X + 15 \mathbb{E}(e_0(n)^4) [\text{Tr}\{\mathbf{R}_X \mathbf{K}(n)\} \mathbf{R}_X + 2\mathbf{R}_X \mathbf{K}(n) \mathbf{R}_X] \right\}. \end{aligned} \quad (2.6)$$

Essa recursão fornece uma boa aproximação para a curva de aprendizado do LMF (dado que a hipótese 1 da seção 2.1.1 implica que $\mathbb{E} \left[(\mathbf{V}(n)^T \mathbf{X}(n))^2 \right] = \text{Tr}(\mathbf{R}_X \mathbf{K}(n))$), o que foi confirmado por diversas simulações em [9]. A recursão também é não-linear, e diverge se a condição inicial $\mathbf{K}(0)$ não for suficientemente pequena, como observado já em [1]. No entanto, predições de região de estabilidade (passo e condição inicial) feitas com (2.6) não são muito próximas do observado em simulações.

Procurou-se corrigir esse problema em [10], evitando-se usar a aproximação feita na hipótese 8 acima, substituída por

8a. A distribuição de $(\mathbf{V}(n)^T \mathbf{X}(n))^2$ é razoavelmente concentrada em torno de sua média de forma que

$$\mathbb{E} \left[(\mathbf{V}(n)^T \mathbf{X}(n))^{2k} \mathbf{X}(n) \mathbf{X}(n)^T \right] \approx \left[\mathbb{E} (\mathbf{V}(n)^T \mathbf{X}(n))^2 \right]^k \mathbf{R}_X.$$

Também foi acrescentada uma nova hipótese,

$$9. \mathbf{R}_X = \sigma_x^2 \mathbf{I}.$$

Com essas hipóteses, [10] obtém uma recursão aproximada para $y(n) \triangleq \mathbb{E}(\|\mathbf{V}(n)\|^2) = \text{Tr}(\mathbf{K}(n))$:

$$y(n+1) \approx (1-a)y(n) - by(n)^2 + cy(n)^3 + d, \quad (2.7)$$

em que M é o comprimento do filtro, e

$$\begin{aligned}
 a &= A_1\mu - A_2\mu^2, & b &= B_1\mu - B_2\mu^2, \\
 c &= C\mu^2, & d &= D\mu^2, \\
 A_1 &= 6\sigma_0^2\sigma_x^2, & A_2 &= 15E[e_0(n)^4]\sigma_x^4(M+2), \\
 B_1 &= 6\sigma_x^4, & B_2 &= 15\sigma_0^2\sigma_x^6(3M+12), \\
 C &= \sigma_x^8(15M+90), & D &= E[e_0(n)^6]\sigma_x^2M.
 \end{aligned}$$

Definindo a função

$$f(y) = (1 - a)y - by^2 + cy^3 + d,$$

nota-se que os pontos de equilíbrio de (2.7) são soluções de $f(y) = y$, e portanto (2.7) tem no máximo 3 zeros, dependendo do valor dos parâmetros. A análise de (2.7) mostra que há três situações possíveis (y_c é o ponto de equilíbrio estável e positivo, se existir, e $y(0)_{\max}$ é o ponto de equilíbrio instável e positivo, se existir):

1. Três pontos de equilíbrio, apenas um dos quais é estável e positivo (ver figura 2.1),
2. Apenas um ponto de equilíbrio positivo, instável (ver figura 2.2),
3. Nenhum ponto de equilíbrio positivo (figura 2.3).

Mostra-se em [10] que $y_c \leq y(0)_{\max}$, e que, para um dado valor de passo, a região de condição inicial $E(\|\mathbf{V}(0)\|^2) = y(0)$ para a qual $y(n)$ converge para y_c é $0 \leq y(0) < y(0)_{\max}$. Note que tanto y_c quanto $y(0)_{\max}$ são funções do passo de adaptação μ . Para cada μ , pode-se resolver a equação $f(y) = y$: se existirem raízes positivas, a maior será $y(0)_{\max}$ (valor máximo da condição inicial), e a menor será y_c (valor de y em regime, ou seja,

$$\lim_{n \rightarrow \infty} E(\|\mathbf{V}(n)\|^2) = y_c.$$

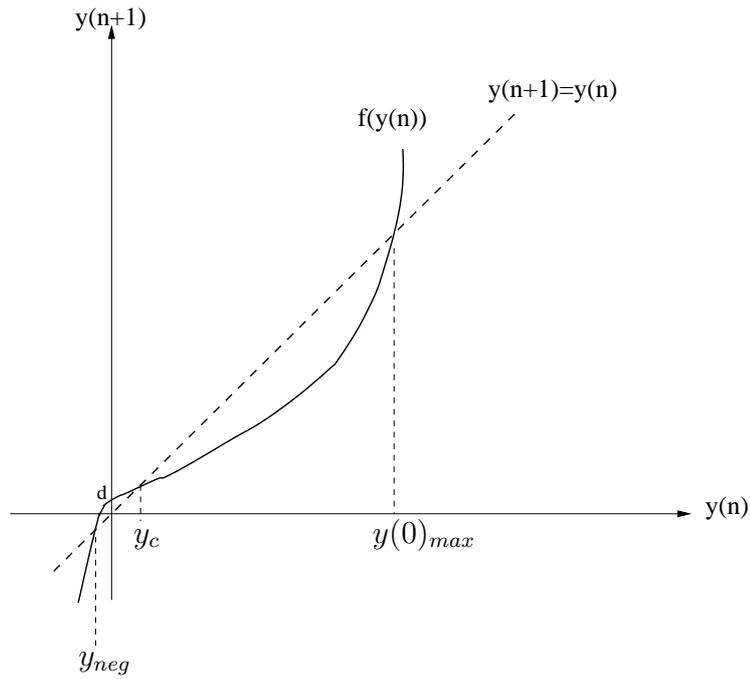


Figura 2.1: Pontos de equilíbrio de (2.7), caso 1: $(y_c \ll y(0)_{max})$.

Se para um certo valor de passo a equação não tiver raízes positivas, ou se tiver apenas uma, o filtro não é estável para esse passo.

O valor máximo para a condição inicial aumenta à medida que μ tende a 0, aumentando a região de atração do ponto de equilíbrio y_c . Por outro lado, existe um valor de passo máximo (μ_{max}), para o qual a equação $f(y) = y$ tem apenas uma raiz positiva, correspondente a um ponto de equilíbrio instável (figura 2.3). Portanto, a faixa de passos para a qual o LMF converge deve ser

$$0 < \mu < \mu_{max},$$

e esse passo máximo pode ser calculado achando-se as raízes de

$$P_4\mu^4 + P_3\mu^3 + P_2\mu^2 + P_1\mu + P_0 = 0, \quad (2.8)$$

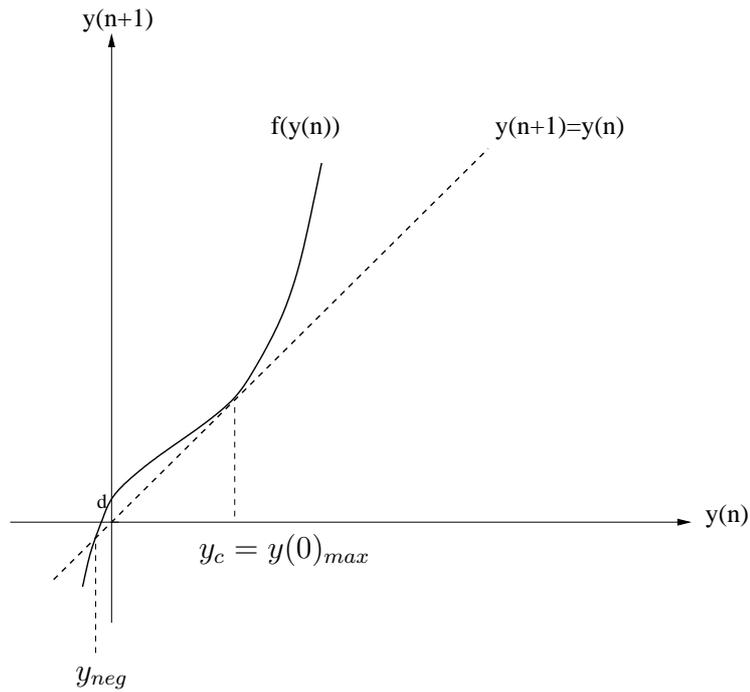


Figura 2.2: Pontos de equilíbrio de (2.7), caso 2: ($y_c = y(0)_{max}$).

em que:

$$P_4 = -4A_2^3C + A_2^2B_2^2 + 18A_2B_2CD - 4B_2^3D - 27C^2D^2,$$

$$P_3 = 12A_1A_2^2C - 2(A_1A_2B_2^2 + A_2^2B_1B_2) - 18(A_1B_2 + A_2B_1)CD + 12B_1B_2^2D,$$

$$P_2 = -12A_1^2A_2C + A_1^2B_2^2 + A_2^2B_1^2 + 4A_1A_2B_1B_2 + 18A_1B_1CD - 12B_1^2B_2D,$$

$$P_1 = 4A_1^3C - 2(A_1A_2B_1^2 + A_1^2B_1B_2) + 4B_1^3D,$$

$$P_0 = A_1^2B_1^2.$$

Este resultado concorda razoavelmente bem com simulações. No entanto, a análise das simulações levantou algumas dúvidas, cujas respostas deram origem a este trabalho, e que serão explicadas no próximo capítulo. Antes de passar a esse assunto serão revistos rapidamente alguns resultados da literatura sobre o LMMN.

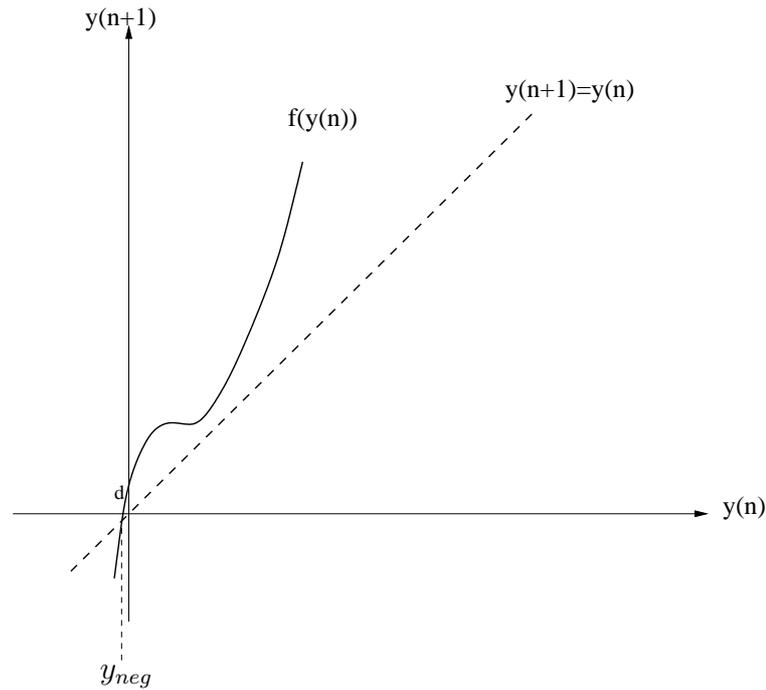


Figura 2.3: Pontos de equilíbrio de (2.7), caso 3: (y_c and $y(0)_{max}$ complexos).

2.2 Análise do LMMN

O LMMN foi proposto em 1994 por Chambers, Tanrikulu e Constantinides [64], procurando uma forma de aumentar a robustez do LMF, e em geral conseguir um desempenho melhor do que os do LMS e do LMF.

As análises do LMMN habitualmente aproveitam e combinam resultados para o LMS e para o LMF. Assim, [64] propõe uma faixa para o passo de adaptação combinando resultados para o LMS e a faixa calculada para o LMF em [1]. Resultados mais precisos foram propostos em [67], que provou a estabilidade do algoritmo para sinais determinísticos, seguindo os resultados de [2], apresentados na seção 2.1.2, e propôs uma aproximação para o desajuste e para a autocorrelação de $\mathbf{V}(n)$ em regime usando o método da equação diferencial (ODE) [54], chegando às seguintes expressões para $\mathbf{K}(n)$ e para o desajuste em regime:

$$\lim_{n \rightarrow \infty} \mathbf{K}(n) \approx \frac{\mu \lambda^2 \sigma_0^2 + 2\lambda(1-\lambda) E(e_0(n)^4) + (1-\lambda)^2 E(e_0(n)^6)}{\lambda + 3(1-\lambda)\sigma_0^2} \mathbf{I}, \quad (2.9)$$

$$D = \frac{\mu}{2\sigma_0^2} \frac{\lambda^2 \sigma_0^2 + 2\lambda(1-\lambda) \mathbb{E}(e_0(n)^4) + (1-\lambda)^2 \mathbb{E}(e_0(n)^6)}{\lambda + 3(1-\lambda)\sigma_0^2} \text{Tr}(\mathbf{R}_X). \quad (2.10)$$

Esses resultados foram obtidos usando a hipótese de que $\mathbf{V}(n)$ se mantém em uma “pequena região em torno de seus valores ótimos”, o que significa que a análise não demonstra a convergência, apenas calcula o valor-limite dos erros para o caso do algoritmo convergir. A análise é válida para passo suficientemente pequeno, e também ignora termos de ordem dois ou superior em $\mathbf{V}(n)$ para a conversão para equação diferencial.

Yousef e Sayed estenderam esses resultados para uma faixa maior de passos de adaptação em [69], em um primeiro passo para estender para análise estocástica o método de análise de filtros adaptativos através de uma estrutura em realimentação proposto inicialmente em [70, 71] para análise determinística. A referência [69] também não demonstra a estabilidade do LMMN, apenas calcula os valores em regime do desajuste e de $\mathbf{K}(n)$ para o caso do algoritmo convergir.

Fazendo a hipótese de que em regime $|e_0(n)| \gg |\mathbf{V}(n)^T \mathbf{X}(n)|$, [69] propõe a seguinte aproximação para o desajuste:

$$D = \frac{\mu a \text{Tr}(\mathbf{R}_X)}{2b - \mu c \text{Tr}(\mathbf{R}_X)}, \quad (2.11)$$

em que

$$a = \lambda^2 \sigma_0^2 + 2\lambda(1-\lambda) \mathbb{E}(e_0(n)^4) + (1-\lambda)^2 \mathbb{E}(e_0(n)^6),$$

$$b = \lambda + 3(1-\lambda)\sigma_0^2,$$

$$c = \lambda^2 + 12\lambda(1-\lambda) + 15(1-\lambda) \mathbb{E}(e_0(n)^4).$$

As aproximações listadas neste capítulo descreveram em linhas gerais os principais modelos propostos nos últimos 20 anos sobre o LMF e o LMMN. Para não

alongar demais o capítulo, não foram descritos os modelos propostos em todos os artigos citados (e logicamente não foram citados todos os trabalhos a respeito desses algoritmos), mas procurou-se dar uma idéia do estado da arte atual da pesquisa a respeito dos algoritmos. Os próximos capítulos tratam das limitações dos modelos atuais, propõem uma nova forma de analisar o desempenho de algoritmos adaptativos, e finalmente mostram como interpretar corretamente os modelos aqui descritos.

3 O MECANISMO DE INSTABILIDADE NO LMF

Como mencionado no capítulo anterior, o modelo obtido em [10] para o comportamento do LMF e para a faixa de valores do passo de adaptação para a qual o algoritmo é estável na média quadrática, concorda razoavelmente com simulações. No entanto, fica uma dúvida, que parte da observação do comportamento típico do LMF quando diverge: diferentemente do que ocorre com o LMS, que apresenta picos de erro quando começa a divergir [63, 72], a divergência do LMF é catastrófica, como mostra a figura 3.1.

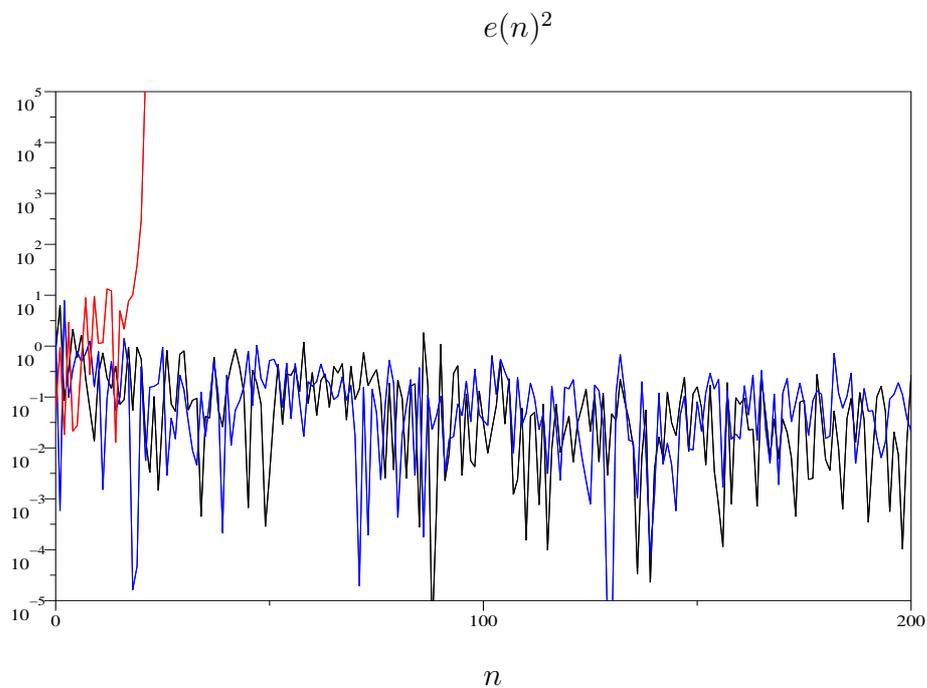


Figura 3.1: Três realizações do LMF com $M = 10$, $\mathbf{w}_* = [0 \ 0 \ \dots \ 0]^T$, $\mathbf{V}(0) = \mathbf{W}(0) = [1 \ 0 \ \dots \ 0]^T$, $\mu = 0,02$, $\mathbf{X}(n)$ Gaussiano com $\mathbf{R}_X = \mathbf{I}$.

A figura 3.1 mostra três realizações do LMF, obtidas sob as mesmas condições e começando da mesma condição inicial. Duas convergem sem problemas, mas uma diverge. Voltando para a análise teórica, nota-se que a recursão (2.7) tem uma região de atração finita, isto é, a condição inicial deve ser menor que $y(0)_{\max}$ para o filtro convergir na média quadrática. No entanto, considere uma realização do filtro. Mesmo que a condição inicial $\mathbf{V}(0)$ seja determinística (e dentro da faixa de convergência), após algumas poucas iterações haverá toda uma faixa de valores possíveis para $y(n)$. Se em uma realização do filtro, para um dado instante, $\|\mathbf{V}(n)\|^2$ ultrapassar $y(0)_{\max}$, o que acontece? As simulações mostram que, à medida que o passo é aumentado, é mais e mais comum ter-se uma realização divergindo. Experimentalmente observou-se que, nas condições do experimento da figura 3.1, em aproximadamente 3% das realizações o LMF diverge.

Nos próximos capítulos será apresentado um modelo para descrever melhor o comportamento do LMF à medida que o passo de adaptação é aumentado, e será fornecida uma interpretação melhor para os valores de $y(0)_{\max}$ e μ_{\max} obtidos no capítulo 2. O presente capítulo usa um caso simples (um filtro com apenas um coeficiente, e uma distribuição normal modificada para o regressor) para mostrar como é a divergência do LMF. Os resultados para a distribuição modificada são confirmados com simulações para o caso de regressores Gaussianos.

3.1 Um Exemplo Simples de Instabilidade

O objetivo aqui é apresentar um exemplo simples mostrando que o LMF tem uma probabilidade não-nula de divergência para uma distribuição bem-comportada do regressor, *qualquer que seja* o valor do passo de adaptação (desde que não-nulo, é claro). Esse exemplo, apesar de idealizado, esclarece o mecanismo de divergência do LMF (parte desta seção foi apresentada em [22]).

3.1.1 Prova de instabilidade para filtros escalares

Considere o algoritmo LMF (1.4) com $M = 1$, utilizado para identificar uma constante w_* , dada uma seqüência $\{x(n)\}$ com função densidade de probabilidade dada por

$$p_x(x) = \begin{cases} 0, & \text{se } |x| < \epsilon, \\ \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(|x| - \epsilon)^2}{2\sigma_x^2}}, & \text{se } |x| \geq \epsilon, \end{cases} \quad (3.1)$$

para $\epsilon > 0$. Assumindo também que não há ruído, de modo que $d(n) = w_*x(n)$, e lembrando que $v(n) = w_* - w(n)$, a equação do LMF fica

$$v(n+1) = (1 - \mu x(n)^4 v(n)^2) v(n). \quad (3.2)$$

Primeiramente será mostrado que há um valor $0 < K < \infty$ tal que, se $|v(n)| > K$ para qualquer n , então $\lim_{n \rightarrow \infty} |v(n)| = \infty$. Em seguida será provado que a probabilidade de $|v(n+1)| > K$ dado $|v(n)| = \alpha$ é não-nula para qualquer $\alpha > 0$.

Seja K tal que ($\delta > 0$ é um número positivo qualquer)

$$\mu\epsilon^4 K^2 - 1 > 1 + \delta \Leftrightarrow K > \sqrt{\frac{2 + \delta}{\mu\epsilon^4}}. \quad (3.3)$$

Dada a desigualdade (3.3) e como $|x(n)| \geq \epsilon$ por (3.1), necessariamente

$$|1 - \mu x(n)^4 K^2| > |1 - \mu\epsilon^4 K^2| > 1 + \delta. \quad (3.4)$$

Se acontecer de $|v(n)| > K$, de (3.2) resulta que

$$\begin{aligned} |v(n+1)| &= |1 - \mu x(n)^4 v(n)^2| |v(n)| \\ &> |1 - \mu\epsilon^4 K^2| |v(n)| > (1 + \delta) |v(n)|, \end{aligned} \quad (3.5)$$

e conclui-se que $|V(n)| \rightarrow \infty$ se para um instante n qualquer acontecer de $|V(n)| > K$.

O argumento é completado mostrando que a probabilidade de $|V(n+1)| > K$ dado $|V(n)| = \alpha$ é não-nula para todo $\alpha > 0$. Defina, para um dado $\alpha > 0$,

$$\beta(\alpha) \triangleq \frac{K}{\mu\alpha^3} + \frac{1}{\mu\alpha^2}. \quad (3.6)$$

Então, para $|v(n)| = \alpha$, uma entrada $x(n)$ tal que $x(n)^4 > \beta(\alpha)$ leva a

$$\mu x(n)^4 v(n)^2 - 1 > \frac{K}{|v(n)|}, \quad (3.7)$$

e portanto,

$$|v(n+1)| = |1 - \mu x(n)^4 v(n)^2| |v(n)| > K. \quad (3.8)$$

As expressões (3.5) e (3.8) mostram que $|v(n)| \rightarrow \infty$ se $x(n)^4 > \beta(\alpha)$ para $|v(n)| = \alpha$. Portanto, para provar que a probabilidade de divergência é não-nula, falta provar que a probabilidade de $x(n)^4 > \beta(\alpha)$, dado que $|v(n)| = \alpha$, é não-nula. Usando (3.1), segue que

$$\begin{aligned} & \Pr \{ |v(n+1)| > K \mid |v(n)| = \alpha \} > \\ & > \Pr \{ x(n)^4 > \beta(\alpha) \mid |v(n)| = \alpha \} \\ & = 2 \int_{\beta(\alpha)^{1/4}}^{\infty} p_x(x) dx > 0, \end{aligned} \quad (3.9)$$

em que $\Pr\{A|B\}$ é a probabilidade de A dado B . Isso conclui a prova.

3.2 Regressores Gaussianos

Quando $x(n)$ é normal ($\epsilon = 0$ em (3.1)), a prova simples apresentada acima não é válida. Contudo, agora são apresentadas simulações que mostram que o resultado ainda é válido para regressores Gaussianos.

Considere que o LMF é aplicado nas mesmas condições da seção 3.1, mas com $\epsilon = 0$. Nas simulações realizadas, foram calculados:

- A probabilidade de divergência do LMF, estimada como segue: Foram simuladas $L = 10^6$ realizações do algoritmo, começando com a mesma condição inicial $V(0) = 1$ e sem ruído ($e_0(n) \equiv 0$). Uma realização foi contada como “divergência” sempre que o erro absoluto $|v(n)|$ ultrapassou 10^{100} (variar este valor em uma ampla faixa não altera os resultados),
- A probabilidade $P_{>}$ de $|V(1)| > |V(0)|$,
- O valor $V_{1/2}$ para o qual a probabilidade $\Pr\{|V(n+1)| > |V(n)| \mid |V(n)| = V_{1/2}\} = 0,5$,
- A probabilidade $P_{>V_{1/2}}$ que $|V(1)| > V_{1/2}$, dada a condição inicial.

A probabilidade de divergência foi obtida experimentalmente. Todos os outros valores podem ser obtidos como segue: inicialmente, obtém-se a função densidade de probabilidade de $v(n+1)$ dado $v(n)$. De (3.2), segue que

$$\begin{aligned}
\Pr\{v(n+1) < z \mid v(n) = Z > 0\} &= \\
&= \Pr\{(1 - \mu x(n)^4 Z^2)Z < z\} = \\
&= \Pr\left\{x(n)^4 > \frac{1 - z/Z}{\mu Z^2}\right\}.
\end{aligned} \tag{3.10}$$

A função densidade de probabilidade de $x(n)^4$ é dada por

$$\begin{aligned}
p_{x^4}(y) &= \frac{d \Pr\{x^4 \leq y\}}{dy} = \\
&= \frac{d \left(2 \int_0^{y^{1/4}} \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{x^2}{2\sigma_x^2}} dx \right)}{dy} = \\
&= \frac{1}{\sqrt{8\pi}\sigma_x y^{3/4}} e^{-\frac{\sqrt{y}}{2\sigma_x^2}}, \quad y \geq 0.
\end{aligned} \tag{3.11}$$

Portanto,

$$\begin{aligned}
\Pr\{v(n+1) \leq z \mid v(n) = Z > 0\} &= \\
&= \Pr\left\{x(n)^4 \geq \frac{1-z/Z}{\mu Z^2}\right\} = \\
&= \int_{\frac{1-z/Z}{\mu Z^2}}^{\infty} \frac{1}{\sqrt{8\pi}\sigma_x y^{3/4}} e^{-\frac{\sqrt{y}}{2\sigma_x^2}} dy.
\end{aligned} \tag{3.12}$$

Finalmente, a função densidade de probabilidade desejada é obtida diferenciando-se (3.12) com relação a z :

$$\begin{aligned}
p_{v(n+1) \mid v(n)}(z \mid v(n) = Z) &= \\
&= \frac{d\Pr\{v(n+1) \leq z \mid v(n) = Z > 0\}}{dz} \\
&= \frac{1}{\sqrt{8\pi}\sigma_x \mu^{1/4} Z^{3/4} (Z-z)^{3/4}} e^{-\frac{\sqrt{Z-z}}{2\sigma_x^2 \sqrt{\mu Z^3}}}.
\end{aligned} \tag{3.13}$$

Escolhendo $\sigma_x^2 = 1$, pode-se usar (3.13) com $v(0) = 1$ (fixo) para determinar as probabilidades $P_{>} = \Pr\{|v(1)| > |v(0)|\}$ e $P_{>V_{1/2}} = \Pr\{|v(1)| > |V_{1/2}| \mid v(0) = 1\}$, e o ponto $V_{1/2} > 0$ para o qual $\Pr\{|v(1)| > v(0) \mid v(0) = V_{1/2}\} = 0,5$. Esses valores são dados, para vários valores do passo μ , na tabela 3.1. A tabela também mostra N_{div} , o número observado de realizações do LMF para o qual $|v(n)| > 10^{100}$ para algum n .

A última coluna na tabela 3.1 mostra que a probabilidade de divergência aumenta com o passo de adaptação. No entanto, mesmo para o maior passo na tabela, o filtro tem um comportamento adequado para a maioria das realizações. A figura 3.2, semelhante à figura 3.1, mostra três realizações para o LMF com $M = 1$ coeficiente, com entrada Gaussiana e iid¹ $x(n)$ com variância unitária, passo $\mu = 0,03$ (probabilidade de divergência de 0,16% de acordo com a tabela 3.1), e condição inicial $v(0) = 1$. A figura mostra duas realizações em que o algoritmo convergiu, e uma em que o algoritmo divergiu. Note que não são necessárias muitas iterações para verificar a divergência. Em todas as simulações re-

¹Independente e identicamente distribuída

Tabela 3.1: Probabilidade de divergência observada para vários valores de passo, sempre para $M = 1$ e condição inicial $|v(0)| = 1$. As definições das colunas são dadas no texto.

μ	$P_{>}$	$V_{1/2}$	$P_{>V_{1/2}}$	N_{div}/L
0,01	$1,7 \times 10^{-4}$	31,1	$5,2 \times 10^{-14}$	7×10^{-6}
0,02	$1,6 \times 10^{-3}$	22,0	$5,8 \times 10^{-9}$	$3,0 \times 10^{-4}$
0,03	$4,3 \times 10^{-3}$	17,9	$5,4 \times 10^{-7}$	$1,6 \times 10^{-3}$
0,04	$7,8 \times 10^{-3}$	15,5	$6,5 \times 10^{-6}$	$4,4 \times 10^{-3}$
0,05	$1,2 \times 10^{-2}$	13,9	$3,3 \times 10^{-5}$	$8,5 \times 10^{-3}$
0,06	$1,6 \times 10^{-2}$	12,7	$1,0 \times 10^{-4}$	$1,4 \times 10^{-2}$
0,07	$2,1 \times 10^{-2}$	11,7	$2,4 \times 10^{-4}$	$2,0 \times 10^{-2}$
0,08	$2,5 \times 10^{-2}$	11,0	$4,7 \times 10^{-4}$	$2,8 \times 10^{-2}$
0,09	$3,0 \times 10^{-2}$	10,4	$8,0 \times 10^{-4}$	$3,5 \times 10^{-2}$
0,10	$3,4 \times 10^{-2}$	9,8	$1,3 \times 10^{-3}$	$4,3 \times 10^{-2}$
0,20	$7,5 \times 10^{-2}$	7,0	$1,2 \times 10^{-2}$	$1,3 \times 10^{-1}$

alizadas, isso se mostrou o comportamento típico do algoritmo. Adicionalmente, note que a probabilidade de divergência depende da condição inicial: quanto maior $|v(0)|$, maior será a probabilidade de divergência. Esse comportamento está de acordo com as observações de [10].

No entanto, se a probabilidade de divergência (isto é, do erro $|v(n)|$ tender a infinito) do algoritmo é não-nula, o algoritmo não pode ser estável na média quadrática, ou seja, para qualquer passo de adaptação positivo, a variância de $v(n)$ (no caso escalar, no caso vetorial, a matriz de autocorrelação $\mathbf{K}(n)$) necessariamente deve divergir. Essa conclusão é contrária à maioria dos resultados sobre análise do LMF publicados na literatura (exceto aqueles que assumem que o regressor e o erro ótimo são estritamente limitados). Nos próximos capítulos esse ponto será melhor analisado, e será apresentada uma interpretação para os resultados publicados na literatura, calculando a faixa de passos que garantem a estabilidade do LMF, mesmo para regressor Gaussiano.

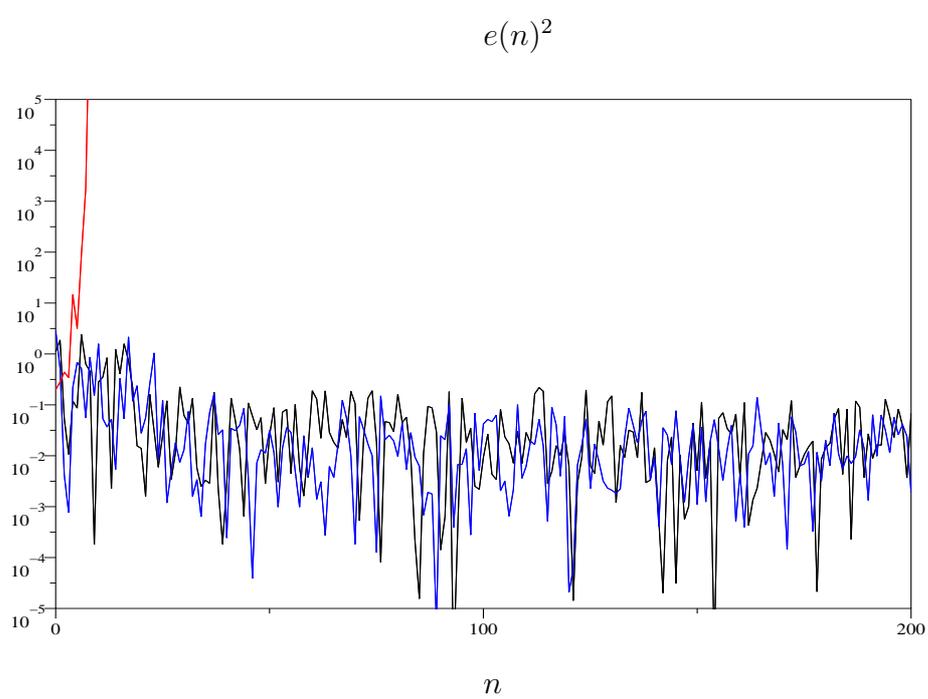


Figura 3.2: Três realizações do LMF com regressor escalar, $w_* = 0$, $v(0) = w(0) = 1$, $\mu = 0,03$, $x(n) \sim N(0, 1)$.

4 PROBABILIDADE DE DIVERGÊNCIA NO LMF

Neste capítulo será abordado o problema de estimar a probabilidade de divergência do LMF para filtros de qualquer comprimento. Antes de mais nada, é necessário definir claramente o que se quer dizer por divergência. Um comportamento indesejado de um filtro poderia ocorrer de várias maneiras: o erro de estimação poderia crescer a valores elevados antes de começar a diminuir, ou ele pode permanecer em valores aceitáveis a maior parte do tempo, mas com picos ocasionais de valor elevado, ou finalmente poderia crescer indefinidamente. A definição usual de estabilidade em filtragem adaptativa envolve a variância das estimativas; isto é, deseja-se que a variância do erro de estimação permaneça limitada e não muito maior que a variância do erro ótimo. Dado o comportamento que foi observado no capítulo anterior, neste capítulo será adotada a seguinte definição de divergência:

Definição 4.1 (Divergência). Neste trabalho diz-se que uma realização do algoritmo LMF *divergiu* se $\lim_{n \rightarrow \infty} \|\mathbf{W}(n)\| = \infty$. Por outro lado, será dito que uma realização do algoritmo *convergiu* se ela não tiver divergido (note que esta não é a definição usual de convergência).

A questão de interesse é a seguinte: dada a condição inicial $\mathbf{W}(0)$, o passo de adaptação μ , o comprimento M do filtro, e as estatísticas do regressor e do erro ótimo $e_0(n)$, qual é a probabilidade que uma realização do filtro divergir?

Para o filtro escalar com a distribuição Gaussiana modificada do capítulo anterior, se o erro absoluto $|v(n)|$ atingir um certo valor K , o algoritmo LMF necessariamente irá divergir, como visto no capítulo anterior. A análise é mais complicada para a distribuição normal sem modificações, pois neste caso, não importa o quanto grande fique $|v(n)|$, sempre haverá uma pequena, mas não-nula probabilidade que o erro retorne a valores razoáveis (no entanto, essa probabilidade rapidamente decresce à medida que $|v(n)|$ aumenta). Isso torna a estimação da probabilidade de divergência um problema consideravelmente complexo. Neste capítulo é proposto um modelo aproximado, que captura a essência da dependência da probabilidade de divergência em μ , M , na variância da seqüência de entrada σ_x^2 , e na distribuição do erro ótimo $e_0(n)$.

4.1 Recursão para $\|\mathbf{V}(n)\|^2$

Inicialmente será procurada uma recursão aproximada para $\|\mathbf{V}(n)\|^2$, assumindo que $\{\mathbf{X}(n)\}$ é iid e Gaussiana, e que os elementos de $\mathbf{X}(n)$ são não-correlacionados. Portanto,

$$\begin{cases} \mathbf{E}\{\mathbf{X}(n)\mathbf{X}(m)^T\} = \mathbf{0} \text{ para } m \neq n, \text{ e} \\ \mathbf{R}_x = \mathbf{E}\{\mathbf{X}(n)\mathbf{X}(n)^T\} = \sigma_x^2 \mathbf{I}, \end{cases}$$

em que \mathbf{I} é a matriz identidade. Também será assumido que a função densidade de probabilidade de $e_0(n)$ em (2.1) é par, de modo que $\mathbf{E}\{e_0(n)^k\} = 0$ sempre que k for um inteiro ímpar.

Recordando que (como visto no capítulo 2),

$$e(n) = \mathbf{V}(n)^T \mathbf{X}(n) + e_0(n), \quad (4.1)$$

em que $\mathbf{V}(n) = \mathbf{w}_* - \mathbf{W}(n)$, a recursão do LMF (1.4) fica, em termos de $\mathbf{V}(n)$,

$$\mathbf{V}(n+1) = \mathbf{V}(n) - \mu \left(\mathbf{V}(n)^T \mathbf{X}(n) + e_0(n) \right)^3 \mathbf{X}(n). \quad (4.2)$$

Definindo $p(n) = \mathbf{V}(n)^T \mathbf{X}(n)$ para simplificar a notação,

$$\begin{aligned} \mathbf{V}(n+1) = \mathbf{V}(n) - \mu \left[p(n)^3 + 3p(n)^2 e_0(n) + 3p(n) e_0(n)^2 + \right. \\ \left. + e_0^3(n) \right] \mathbf{X}(n). \end{aligned} \quad (4.3)$$

Define-se também $y(n) \triangleq \mathbf{V}(n)^T \mathbf{V}(n) = \|\mathbf{V}(n)\|^2$. De (4.3) obtém-se

$$\begin{aligned} y(n+1) = y(n) - 2\mu \left[p(n)^4 + 3p(n)^3 e_0(n) + 3p(n)^2 e_0(n)^2 + \right. \\ \left. + p(n) e_0(n)^3 \right] + \mu^2 \left[p(n)^6 + 6p(n)^5 e_0(n) + \right. \\ \left. + 15p(n)^4 e_0(n)^2 + 20p(n)^3 e_0(n)^3 + 15p(n)^2 e_0(n)^4 + \right. \\ \left. + 6p(n) e_0(n)^5 + e_0(n)^6 \right] \|\mathbf{X}(n)\|^2. \end{aligned} \quad (4.4)$$

O objetivo é estimar a probabilidade de $\lim_{n \rightarrow \infty} y(n) = \infty$.

Para prosseguir, é necessário simplificar (4.4). Para isso, faz-se a aproximação $y(n) \approx E\{y(n) | y(n-1), \mathbf{X}(n)\}$. Esta aproximação troca o erro ótimo $e_0(n)$ e suas potências por suas respectivas médias, resultando

$$\begin{aligned} y(n+1) = y(n) - 2\mu \left[p(n)^4 + 3p(n)^2 \sigma_0^2 \right] + \\ + \mu^2 \left[p(n)^6 + 15p(n)^4 \sigma_0^2 + 15p(n)^2 \psi_0^4 + \eta_0^6 \right] \|\mathbf{X}(n)\|^2, \end{aligned}$$

em que foram usadas as hipóteses de que $E\{e_0(n)\} = E\{e_0(n)^3\} = E\{e_0(n)^5\} = 0$, e definiu-se $\psi_0^4 = E\{e_0(n)^4\}$, $\eta_0^6 = E\{e_0(n)^6\}$.

É necessário agora aproximar $(\mathbf{V}(n)^T \mathbf{X}(n))^{2k}$, escrevendo esses valores em termos de $y(n)^k$ e de $\|\mathbf{X}(n)\|^{2k}$. Usando a hipótese de que os elementos de $\mathbf{X}(n)$ são Gaussianos e não-correlacionados (e portanto independentes), conclui-se que

o vetor $\mathbf{X}(n)$ aponta para qualquer direção no espaço \mathbb{R}^M com probabilidade igual. A hipótese de $\mathbf{X}(n)$ ser iid implica que as direções de $\mathbf{V}(n)$ e de $\mathbf{X}(n)$ são independentes. Esta observação é usada no Apêndice A para chegar às seguintes aproximações:

$$\frac{(\mathbf{V}(n)^T \mathbf{X}(n))^{2k}}{(\|\mathbf{V}(n)\| \|\mathbf{X}(n)\|)^{2k}} \approx \mathbb{E} \left\{ \frac{(\mathbf{V}(n)^T \mathbf{X}(n))^{2k}}{(\|\mathbf{V}(n)\| \|\mathbf{X}(n)\|)^{2k}} \right\} \triangleq \alpha_M(k), \quad (4.5)$$

com

$$\alpha_M(1) = \frac{1}{M},$$

$$\alpha_M(2) = \frac{3}{M(M+2)},$$

e

$$\alpha_M(3) = \frac{15}{M(M+2)(M+4)}.$$

Substituindo $(\mathbf{V}(n)^T \mathbf{X}(n))^{2k}$ por $\alpha_M(k)y(n)^k \|\mathbf{X}(n)\|^{2k}$, obtém-se

$$\begin{aligned} y(n+1) \approx & \left[1 - \mu (6\sigma_0^2 - 15\mu\psi_0^4 \|\mathbf{X}(n)\|^2) \frac{\|\mathbf{X}(n)\|^2}{M} - \right. \\ & - 3\mu(2 - 15\mu\sigma_0^2 \|\mathbf{X}(n)\|^2) \frac{\|\mathbf{X}(n)\|^4}{M(M+2)} y(n) + \\ & \left. + 15\mu^2 \frac{\|\mathbf{X}(n)\|^8}{M(M+2)(M+4)} y(n)^2 \right] y(n) + \mu^2 \eta_0^6 \|\mathbf{X}(n)\|^2. \end{aligned} \quad (4.6)$$

4.2 Estimando a Probabilidade de Divergência

O próximo passo é estimar a probabilidade de que a seqüência $y(n)$ dada pela recursão (4.6) cresça indefinidamente, dada a condição inicial $y(0) = \mathbf{V}(0)^T \mathbf{V}(0)$, e μ , σ_x^2 , σ_0^2 , ψ_0^4 , η_0^6 , e M .

Denote por $D(n)$ o fator entre colchetes multiplicando $y(n)$ em (4.6), isto é,

$$\begin{aligned} D(n) &= 1 - \mu (6\sigma_0^2 - 15\mu\psi_0^4\|\mathbf{X}(n)\|^2) \frac{\|\mathbf{X}(n)\|^2}{M} - \\ &\quad - 3\mu(2 - 15\mu\sigma_0^2\|\mathbf{X}(n)\|^2) \frac{\|\mathbf{X}(n)\|^4}{M(M+2)}y(n) + \\ &\quad + 15\mu^2 \frac{\|\mathbf{X}(n)\|^8}{M(M+2)(M+4)}y(n)^2. \end{aligned} \quad (4.7)$$

Prova-se no Apêndice B que $D(n)$ é sempre não-negativo.

Uma aproximação P_d para a probabilidade de divergência do LMF pode ser obtida como segue. A recursão (4.6) converge se houver um D_0 dado para o qual $0 \leq D(n) < D_0 < 1$ para todo n , já que neste caso $y(n+1) \leq D_0y(n) + \mu\eta_0^6\|\mathbf{X}(n)\|^2$, e portanto

$$y(n) \leq D_0^n y(0) + \mu\eta_0^6 \sum_{k=0}^{n-1} D_0^{n-1-k} \|\mathbf{X}(k)\|^2. \quad (4.8)$$

Se $D(n) \leq D_0 < 1$ para n , $\|\mathbf{X}(n)\|^2$ deve necessariamente ser limitado (pois $D(n) \rightarrow \infty$ quando $\|\mathbf{X}(n)\| \rightarrow \infty$). Esta propriedade de $\|\mathbf{X}(n)\|$, juntamente com a equação (4.8), implicam que $y(n)$ permanece limitada (e portanto, de acordo com a Definição 4.1, o algoritmo converge).

No entanto, pode ocorrer do algoritmo (quer dizer, uma realização do algoritmo) convergir mesmo se alguns $D(n)$ forem maiores do que um. Assim,

$$P_c = \Pr\{0 < D(n) < 1 \text{ for all } n \geq 0\}$$

é um limite inferior para a probabilidade de convergência, e

$$P_d = 1 - P_c$$

é um limite superior para a probabilidade de divergência.

Para avaliar P_c , pode-se:

1. Calcular as probabilidades $\Pr\{D(n) < 1 \mid y(n) = \hat{y}(n)\}$, para $0 \leq n \leq N$,

começando com um valor dado de $y(0)$,

2. Um inconveniente na expressão anterior é que $y(n)$, para $n > 0$, varia para cada realização do filtro. Para poder prosseguir, é necessário achar uma aproximação $\hat{y}(n)$ para $y(n)$.¹

3. Calcular

$$P_c \approx \prod_{n=0}^N \Pr\{D(n) < 1 \mid y(n) = \hat{y}(n)\} \quad (4.9)$$

Admitindo por hora que uma aproximação $\hat{y}(n)$ esteja disponível, será calculada a probabilidade referida no Passo 1. De (4.7), nota-se que $D(n) < 1$ é equivalente a

$$\begin{aligned} & 1 - \mu (6\sigma_0^2 - 15\mu\psi_0^4 \|\mathbf{X}(n)\|^2) \frac{\|\mathbf{X}(n)\|^2}{M} - \\ & - 3\mu(2 - 15\mu\sigma_0^2 \|\mathbf{X}(n)\|^2) \frac{\|\mathbf{X}(n)\|^4}{M(M+2)} y(n) + \\ & + 15\mu^2 \frac{\|\mathbf{X}(n)\|^8}{M(M+2)(M+4)} y(n)^2 < 1, \end{aligned}$$

que, definindo $z \triangleq \|\mathbf{X}(n)\|^2$, se reduz a

$$\begin{aligned} Q(z) \triangleq & 6\sigma_0^2 + \left(\frac{6}{M+2} y(n) - 15\mu\psi_0^4 \right) z - \frac{45\mu\sigma_0^2}{M+2} y(n) z^2 - \\ & - 15 \frac{\mu}{(M+2)(M+4)} y(n)^2 z^3 > 0. \end{aligned}$$

$Q(z)$ não tem zero positivo se $y(n) = \sigma_0^2 = \psi_0^4 = 0$; caso contrário $Q(z)$ tem um e apenas um zero positivo $z_0(n)$ para $y(n) \geq 0$, como mostrado agora:

1. Se o erro ótimo for identicamente nulo (isto é, se $\sigma_0^2 = \psi_0^4 = 0$), então $Q(z)$ ser reduz a

$$\begin{aligned} Q(z)|_{\text{erro nulo}} = & z \left(\frac{6}{M+2} y(n) - \right. \\ & \left. - \frac{15\mu}{(M+2)(M+4)} y(n)^2 z^2 \right), \end{aligned}$$

¹O uso de $\hat{y}(n)$, de (4.5), e dos valores esperados de potências de $e_0(n)$, faz a análise apresentada aqui aproximada. Os resultados serão validados no capítulo 5.

e o único zero positivo é dado por

$$z_0(n)|_{\text{erro nulo}} = \sqrt{\frac{2(M+4)}{5\mu y(n)}}.$$

2. Se $y(n) = 0$ e $\sigma_0^2 > 0$, $Q(z)$ tem grau um, e o único zero positivo é

$$z_0(n)|_{y(n)=0} = \frac{2\sigma_0^2}{5\mu\psi_0^4}.$$

3. Para o caso de erro ótimo não nulo e $y(n) > 0$, note que $Q(z)$ tem a forma

$$a + bz - cz^2 - dz^3,$$

em que $a, c, d > 0$, mas b é indefinido. Como $a > 0$, segue que $Q(0) > 0$, e como $d > 0$, segue que

$$\lim_{z \rightarrow \infty} Q(z) = -\infty,$$

e portanto há ao menos um zero positivo para $Q(z)$. O argumento seguinte prova que esse zero é único. A primeira e a segunda derivadas de $Q(z)$ são

$$Q'(z) = b - 2cz - 3dz^2,$$

$$Q''(z) = -2c - 6dz.$$

Como $Q''(z) < 0$ para $z > 0$, $Q'(z)$ é estritamente decrescente para $z > 0$. $Q'(z)$ terá portanto um único zero positivo se $b > 0$, e nenhum se $b \leq 0$. Como $Q'(z)$ é estritamente decrescente, $Q'(z)$ pode ter no máximo um zero positivo $z_1 > 0$. Isso implica que $Q(z)$ pode ter no máximo um ponto extremo $z_1 > 0$. Há duas possibilidades para $Q(z)$, portanto:

- (a) se $b > 0$, $Q(z)$ começa em $Q(0) > 0$, cresce para $0 < z < z_1$, alcança um máximo em z_1 , e depois decresce para $-\infty$; ou
- (b) se $b \leq 0$, $Q(z)$ começa em $Q(0) > 0$ e decresce para $-\infty$ para $z > 0$.

Nos dois casos, $Q(z)$ cruza o eixo z uma (e uma única) vez para $z > 0$, e o

zero $z_0(n) > 0$ deve ser único. Os conjuntos $\{z \in \mathbb{R} \mid z > 0, Q(z) > 0\}$ e $\{z \in \mathbb{R} \mid 0 < z < z_0(n)\}$ são portanto idênticos, e a probabilidade que se deseja calcular é dada por (lembre que $z_0(n)$ depende de $y(n)$)

$$\begin{aligned} \Pr\{D(n) < 1 \mid y(n) = \hat{y}(n)\} &= \\ &= \Pr\{\|\mathbf{X}(n)\|^2 \leq z_0(n) \mid y(n) = \hat{y}(n)\}. \end{aligned} \quad (4.10)$$

Como os elementos de $\mathbf{X}(n)$ são Gaussianos e iid, $\|\mathbf{X}(n)\|^2/\sigma_x^2$ segue uma distribuição χ^2 com M graus de liberdade, e a última probabilidade em (4.10) pode ser facilmente calculada.

Ainda é necessário encontrar uma aproximação $\hat{y}(n)$ para $y(n) = \|\mathbf{V}(n)\|^2$, para todo instante n . A escolha mais razoável seria a mediana da distribuição de $\|\mathbf{V}(n)\|^2$, mas esse parâmetro não é facilmente calculável. Outra aproximação seria usar $\hat{y}(n) = E\{\|\mathbf{V}(n)\|^2\}$. Contudo, esta escolha também não é conveniente, pois $E\{\|\mathbf{V}(n)\|^2\}$ diverge, como observado na seção 3.2. Na próxima seção é apresentada uma alternativa para $\hat{y}(n)$ que é facilmente calculável e fornece resultados razoáveis.

4.3 Cálculo de $\hat{y}(n)$

O valor esperado de (4.2) para $\mathbf{X}(n)$ Gaussiano com correlação $\sigma_x^2 \mathbf{I}$ foi calculado em [9], com o resultado

$$\begin{aligned} E\{\mathbf{V}(n+1)\} &\approx E\{\mathbf{V}(n)\} - \\ &- 3\mu [\sigma_0^2 + \sigma_x^2 E\{\mathbf{V}(n)^T \mathbf{V}(n)\}] \sigma_x^2 E\{\mathbf{V}(n)\} \end{aligned} \quad (4.11)$$

A estimativa $\hat{y}(n)$ é conseguida através das aproximações

$$\mathbf{V}(n) \approx E\{\mathbf{V}(n)\},$$

e

$$\hat{y}(n) \approx \mathbf{E}\{\mathbf{V}(n)\}^T \mathbf{E}\{\mathbf{V}(n)\},$$

que são razoáveis no começo da adaptação, quando $\mathbf{V}(n)$ é dominado por seu valor médio (lembrando que a divergência é mais provável nas primeiras iterações, como observado nas simulações do capítulo 3). A multiplicação de (4.11) por sua transposta fornece uma recursão para $\mathbf{E}\{\mathbf{V}(n)\}^T \mathbf{E}\{\mathbf{V}(n)\}$, que pode ser usada para definir $\hat{y}(n)$, como segue

$$\begin{aligned} \hat{y}(n+1) &= [1 - 3\mu\sigma_x^2(\sigma_0^2 + \sigma_x^2\hat{y}(n))]^2 \hat{y}(n), \\ \hat{y}(0) &= y(0) = \mathbf{E}\{\mathbf{V}(0)^T \mathbf{V}(0)\} = \mathbf{V}(0)^T \mathbf{V}(0). \end{aligned} \tag{4.12}$$

Observe que essa escolha para $\hat{y}(n)$ só é razoável se μ for pequeno. Se o passo for grande o suficiente para que $1 - 3\mu\sigma_x^2(\sigma_0^2 + \sigma_x^2\hat{y}(0)) = 0$, então $\hat{y}(n) = 0$ for $n \geq 1$, o que está longe do comportamento real do algoritmo.

Todos os elementos para a estimativa da probabilidade de divergência do LMF estão agora disponíveis:

4.4 Probabilidade de Divergência para o LMMN

O mesmo algoritmo pode ser usado para estimar a probabilidade de divergência do LMMN, bastando para isso usar

$$\mu_{\text{eq}} = (1 - \lambda)\mu \tag{4.15}$$

no algoritmo 1.

A estimativa da probabilidade de divergência obtida pelo algoritmo 1 é um limite superior para a probabilidade real, portanto pode ser usada por projetistas para escolher os parâmetros dos filtros para que a probabilidade real seja menor do que um dado valor pré-escolhido. Como várias aproximações foram usadas para se chegar ao algoritmo 1, o próximo capítulo apresenta diversas simulações

Probabilidade de divergência

$$P_d \approx 1 - P_c, \quad (4.13)$$

em que

$$P_c = \prod_{n=0}^N \Pr \{ \|\mathbf{X}(n)\|^2 \leq z_0(n) \mid y(n) = \hat{y}(n) \},$$

em que $\|\mathbf{X}(n)\|^2/\sigma_x^2$ segue uma distribuição χ^2 com M graus de liberdade, e $z_0(n)$ é o único zero positivo de (se $\hat{y}(n) = 0$ e $\sigma_0^2 = 0$, $P_c = 1$)

$$Q(z) = 6\sigma_0^2 + \left(\frac{6}{M+2}\hat{y}(n) - 15\mu\psi_0^4 \right) z - \frac{45\mu\sigma_0^2}{M+2}\hat{y}(n)z^2 - \frac{15\mu}{(M+2)(M+4)}\hat{y}(n)^2z^3, \quad (4.14)$$

com $\sigma_0^2 = E(e_0(n)^2)$, $\psi_0^4 = E(e_0(n)^4)$, e $\hat{y}(n)$ é calculado a partir da recursão

$$\begin{aligned} \hat{y}(n+1) &= [1 - 3\mu\sigma_x^2(\sigma_0^2 + \sigma_x^2\hat{y}(n))]^2 \hat{y}(n), \\ \hat{y}(0) &= y(0) = E\|\mathbf{V}(0)\|^2, \end{aligned}$$

para $\mu < [3\sigma_x^2(\sigma_0^2 + \sigma_x^2\hat{y}(0))]^{-1}$.

Algoritmo 1: Cálculo da probabilidade de divergência.

comparando os valores obtidos pelo algoritmo 1 com resultados de simulações.

5 SIMULAÇÕES

Neste capítulo são comparadas as estimativas para a probabilidade de divergência do LMF e do LMMN obtidas com o algoritmo 1, e resultados de diversas simulações. Foram feitas comparações do modelo teórico para erro ótimo $e_0(n)$ Gaussiano e uniforme, e para regressores Gaussianos de dois tipos: *vetores* $\mathbf{X}(n)$ independentes (referidos como “regressores IND” no restante deste capítulo) e vetores $\mathbf{X}(n)$ formados a partir de uma linha de atraso, como é usual em filtragem adaptativa (referidos por “regressores LA”). A matriz de autocorrelação de $\mathbf{X}(n)$ é sempre igual a $\sigma_x^2 \mathbf{I}$ (ou seja, no caso IND, $E\{\mathbf{X}(n)\mathbf{X}(n-k)\} = 0$ para todo $k \neq 0$, e no caso LA, $E\{\mathbf{X}(n)\mathbf{X}(n-k)\}$ é não-nulo e constante apenas na k -ésima diagonal.)

5.1 Simulações com o LMF

O primeiro exemplo valida o procedimento usado para as simulações. São comparadas as probabilidades de divergência em função do passo de adaptação, obtidos sob diferentes condições. Cada curva na figura 5.1 foi obtida através de L realizações independentes do LMF para N_{it} iterações. Para este exemplo, foi adotado $M = 100$, $\sigma_x^2 = 1$, $\sigma_z^2 = 0.01$, e condição inicial $y(0) = \|\mathbf{V}(0)\|^2 = 1$. Uma curva foi marcada como “divergente” se $\|\mathbf{V}(N_{it})\| \geq 10^{100}$, e a probabilidade de divergência observada foi calculada por

$$P_{d,o} = \frac{\text{Número de curvas divergentes}}{L}.$$

Há sete curvas na figura 5.1. As duas à esquerda foram obtidas usando o modelo teórico usando como número de iterações no algoritmo 1 os valores $N_{it} = 100$ (o) e 1000 (traço contínuo). As cinco curvas à direita foram obtidas por simulações (probabilidades observadas $P_{d,o}$). A curva mais à direita (+) foi obtida usando regressores LA. As quatro simulações restantes foram obtidas usando regressores IND.

As simulações mostram que basicamente não há diferenças entre as curvas obtidas com $L = 10^3$ e $L = 10^4$ realizações (casos (a) e (b) na figura). Com relação ao número de iterações, nota-se um deslocamento da curva estimada para a esquerda (ou seja, maior probabilidade de divergência para valores menores de μ) quando N_{it} é aumentado de 10^2 para 10^3 , mas praticamente não há variação para $L \geq 10^3$. Com relação ao tipo de regressor, comparando as curvas para regressores LA e IND obtidas para $N_{it} = L = 10^4$ (a curva marcada (+) e a curva marcada \diamond para o caso (b)), fica claro que as diferenças entre simulações com regressores IND e LA têm pouco impacto. O modelo teórico realmente forneceu um limite superior para a probabilidade de divergência observada em ambos os casos.

Na figura 5.2 foram usados apenas regressores LA. As curvas em traço contínuo foram obtidas por simulações, com $y(0) = 0,01$ e $\sigma_x^2 = 1$, $\sigma_0^2 = 0,01$, $N_{it} = 10^4$, e $L = 10^3$; e as curvas tracejadas mostram o modelo teórico P_d , usando 10^3 iterações (os mesmos resultados teóricos foram obtidos usando 10^4 iterações). Note que para erro $y(0)$ menor (menor que o valor $y(0) = 1$ usado previamente), as aproximações teóricas ficam mais próximas das simulações.

A figura 5.3 mostra novamente resultados apenas para regressores LA, $M = 10^2$, $\sigma_x^2 = 1$, $N_{it} = 10^4$, e $L = 10^3$, mas agora com erro ótimo Gaussiano com $\sigma_0^2 = 0,1$. As curvas acima são com $y(0) = 0,1$, e as abaixo, para $y(0) = 1$. Novamente, o modelo teórico é melhor para valores menores de $y(0)$.

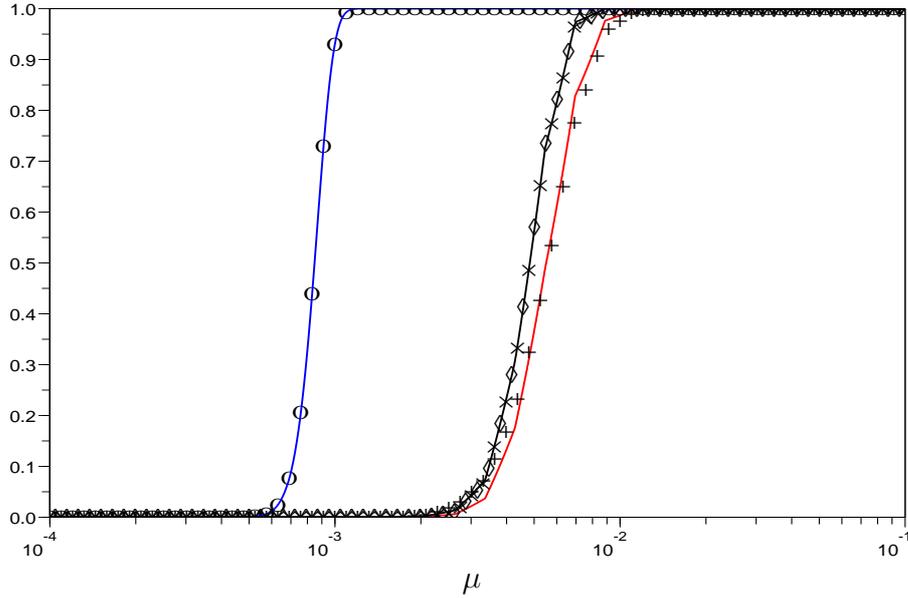
Probabilidade de divergência observada para $M = 100$ 

Figura 5.1: Probabilidade de divergência, erro ótimo Gaussiano, $M = 100$, $\sigma_x^2 = 1$, $y(0) = 1$, $\sigma_0^2 = 0.01$. À esquerda vêem-se duas curvas superpostas obtidas com o modelo teórico proposto aqui, calculadas com 10^2 (o) e 10^3 (traço contínuo) iterações. A curva mais à direita, marcada por +, foi obtida por simulações usando regressores LA e $N_{it} = L = 10^4$. A curva em traço contínuo próxima a ela foi obtida com regressores IND, $N_{it} = 10^2$ e $L = 10^4$. No centro há três curvas superpostas, obtidas com regressores IND e (a) $N_{it} = 10^3$ e $L = 10^4$ (traço contínuo), (b) $N_{it} = L = 10^4$ (\diamond), e (c) $N_{it} = 2.5 \times 10^4$ e $L = 10^3$ (\times).

Na figura 5.4 são apresentadas simulações para regressores LA e $M = 100$, agora com diferentes valores para σ_x^2 . Em todos os casos as curvas em traço contínuo são simulações feitas com $N_{it} = 10^4$, $L = 10^3$, e $\sigma_z^2 = 0.01$; e as curvas tracejadas foram obtidas com o modelo teórico. Em (a), $y(0) = 0, 1$ e $\sigma_x^2 = 2$, em (b), $y(0) = 0, 1$ e $\sigma_x^2 = 4$, em (c), $y(0) = 1$ e $\sigma_x^2 = 0.5$, e em (d), $y(0) = 1$ e $\sigma_x^2 = 4$.

Resultados semelhantes são obtidos para outras distribuições do erro ótimo $e_0(n)$. A figura 5.5 apresenta resultados para erro ótimo com distribuição uniforme.

Probabilidade de divergência observada.

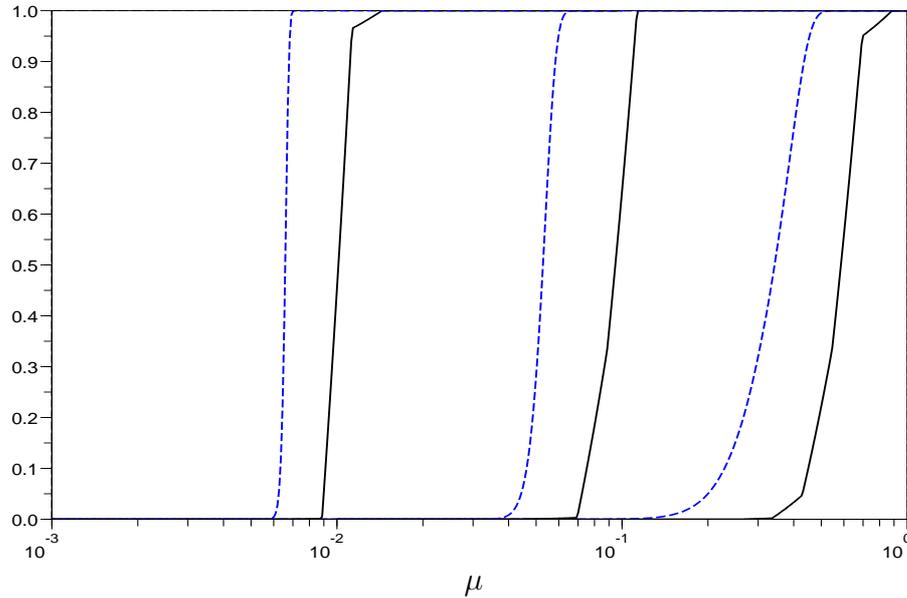


Figura 5.2: Probabilidade de divergência, regressores LA, erro ótimo Gaussiano, $\sigma_x^2 = 1$, $y(0) = 0.01$, $\sigma_0^2 = 0.01$, e $M = 10, 100$, e 1000 . Há um par de curvas para cada valor de M . As curvas em traço contínuo foram obtidas por simulações, e as tracejadas, pelo modelo teórico. O par mais à esquerda é para $M = 10^3$, o par central, para $M = 10^2$, e o par mais à direita, para $M = 10$. Em todos os casos, $N_{it} = 10^4$ e $L = 10^3$.

5.2 Simulações com o LMMN

Simulações realizadas com o LMMN apresentam resultados semelhantes. Nas figuras 5.6–5.8 podem ser vistas a probabilidade de divergência observada para o LMMN com regressores LA e $\delta = 0,8$. As curvas teóricas foram obtidas usando o passo $\mu_{eq} = 2\mu(1 - \delta)$ no algoritmo 1.

As figuras mostram que o modelo teórico realmente fornece um limite superior para a probabilidade de divergência do LMF e do LMMN. As simulações realizadas mostraram que o modelo teórico se aproxima da probabilidade observada para valores menores da condição inicial $y(0)$, e se a variância do erro ótimo não é maior do que a condição inicial. Foram realizadas várias outras simulações, variando todos os parâmetros possíveis, sempre com resultados similares. Si-

Probabilidade de divergência observada.

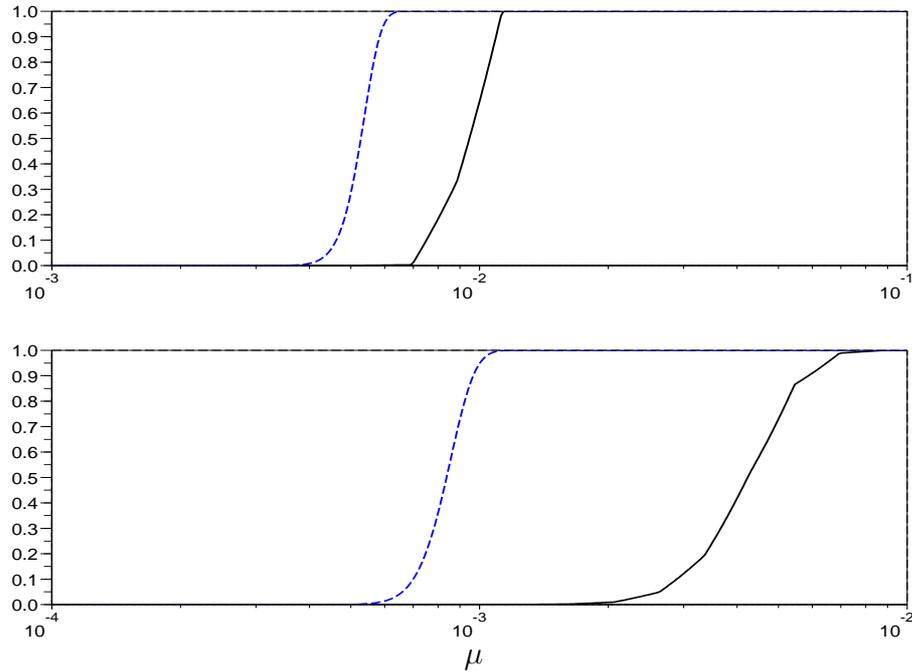


Figura 5.3: Probabilidade de divergência, regressores LA, erro ótimo Gaussiano, $\sigma_0^2 = 0.1$, $\sigma_x^2 = 1$, e $M = 100$. Acima: $y(0) = 0, 1$, abaixo: $y(0) = 1$. Em todos os casos, $N_{it} = 10^4$ e $L = 10^3$. As curvas em traço contínuo são simulações, as tracejadas foram obtidas com o modelo teórico usando 10^3 iterações.

mulações em que $\sigma_0^2 \geq y(0)$ são mais difíceis de realizar, pois nesse caso o número de iterações para que o valor de $P_{d,o}$ se estabilize tende a ser muito grande, e as simulações muito demoradas. O mesmo ocorre se σ_x^2 for reduzido demais.

5.3 Comparação com Modelos Tradicionais

O modelo teórico aqui apresentado pode ser comparado com os limites para o passo de adaptação propostos em [10]. Esses limites estão listados na tabela 5.1, para diferentes condições.

Comparando os valores máximos para o passo de adaptação preditos pela tabela 5.1 com as simulações apresentadas nas figuras 5.1–5.3, pode-se notar que os limites propostos em [10] para o passo de adaptação ficam no início da região em que a probabilidade de divergência está crescendo rapidamente. Portanto, os

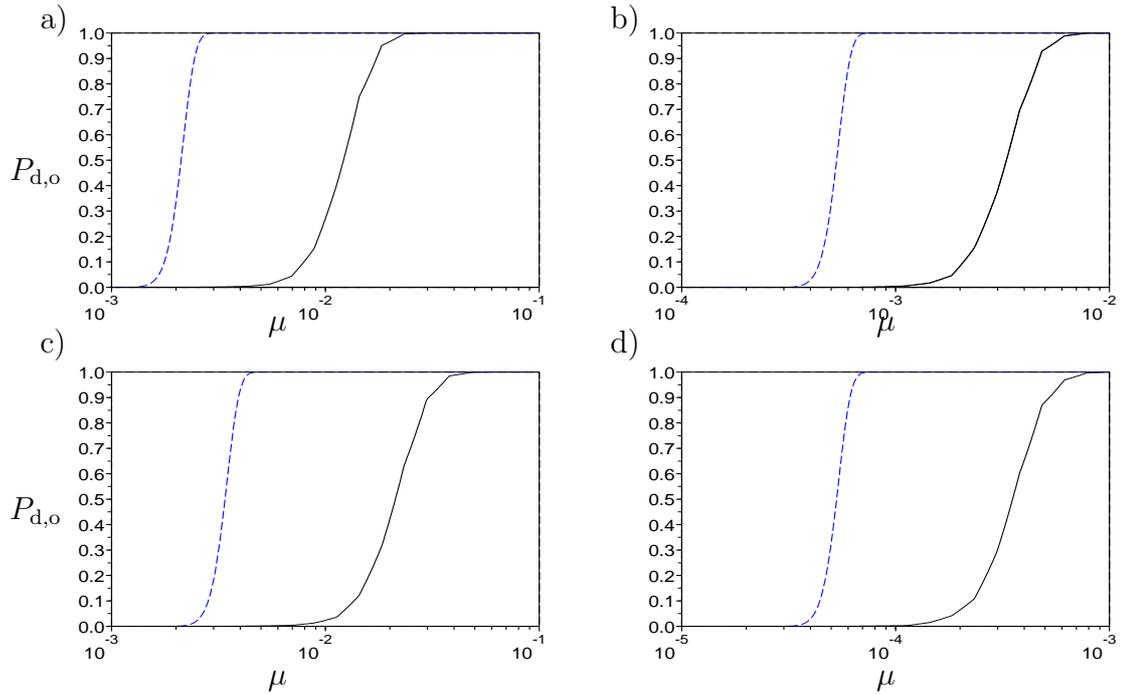


Figura 5.4: Probabilidade de divergência como função da potência do sinal, regressores LA, erro ótimo Gaussiano, $M = 100$. Traço contínuo: simulações, $N_{\text{it}} = 10^4$, $L = 10^3$, e $\sigma_0^2 = 0,01$. Curvas tracejadas: modelo teórico, calculado com 10^3 iterações. (a), $y(0) = 0,1$ e $\sigma_x^2 = 2$; (b), $y(0) = 0,1$ e $\sigma_x^2 = 4$; (c), $y(0) = 1$ e $\sigma_x^2 = 0,5$; (d), $y(0) = 1$ e $\sigma_x^2 = 4$.

resultados de [10] são realmente úteis para determinar uma faixa de passos de adaptação utilizáveis, dada a relativa simplicidade da expressão para o cálculo de μ_{max} lá proposta, como mostrado na equação (2.8) do capítulo 2.

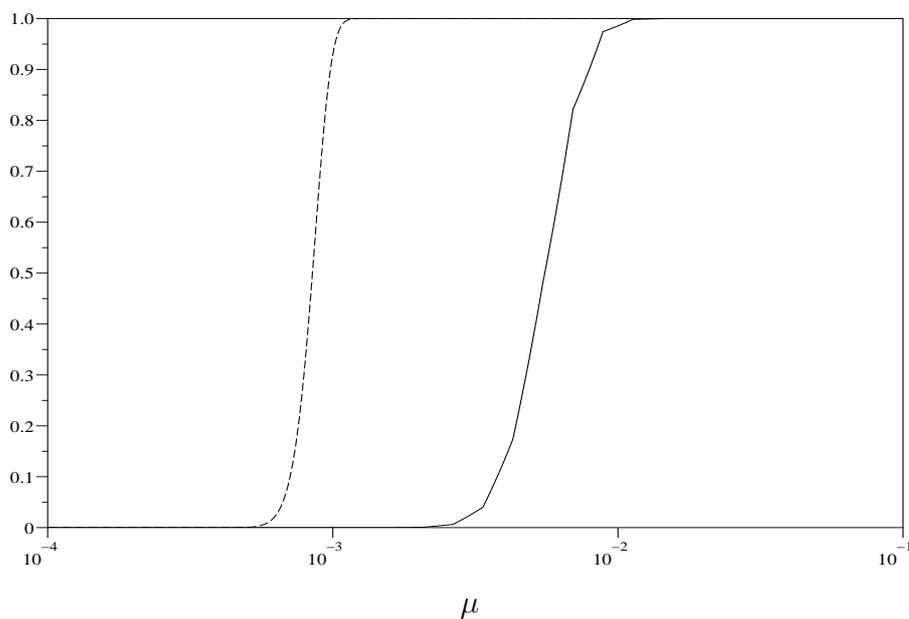
Observed probability of divergence and P_d , $M = 100$ 

Figura 5.5: Probabilidade de divergência para o LMF, regressores IND, erro ótimo uniformemente distribuído, $M = 100$, $\sigma_0^2 = 0.01$. Traço contínuo: probabilidade observada em 10^5 realizações com 100 iterações. Curva tracejada: modelo teórico, calculado também com 100 iterações.

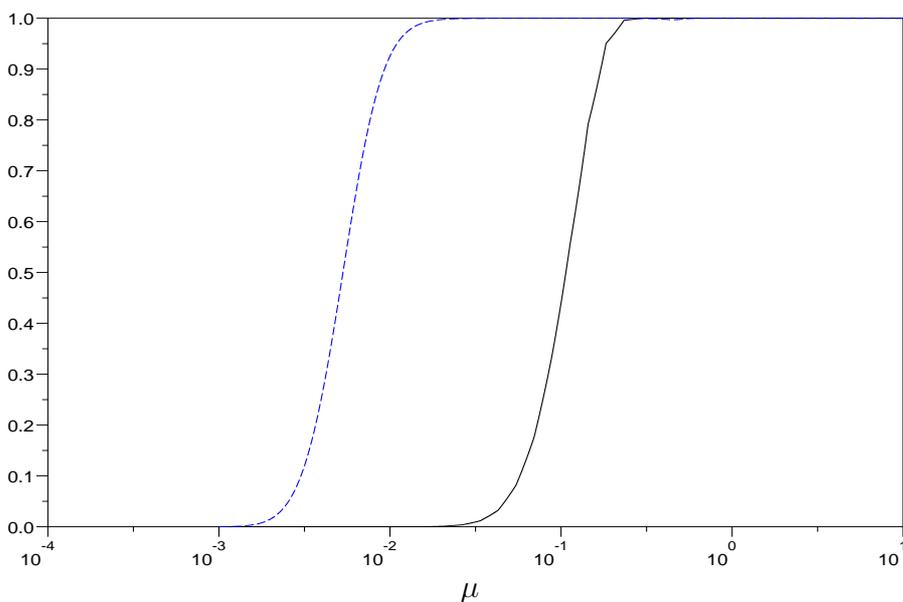
Probabilidade de divergência para o LMMN, $M = 10$ 

Figura 5.6: Probabilidade de divergência para o LMMN, regressores LA, erro ótimo Gaussiano, $M = 10$, $\sigma_0^2 = 0,01$, $\delta = 0,8$. Traço contínuo: probabilidade observada em 10^5 realizações com 100 iterações. Curva tracejada: modelo teórico.

Probabilidade de divergência para o LMMN, $M = 100$

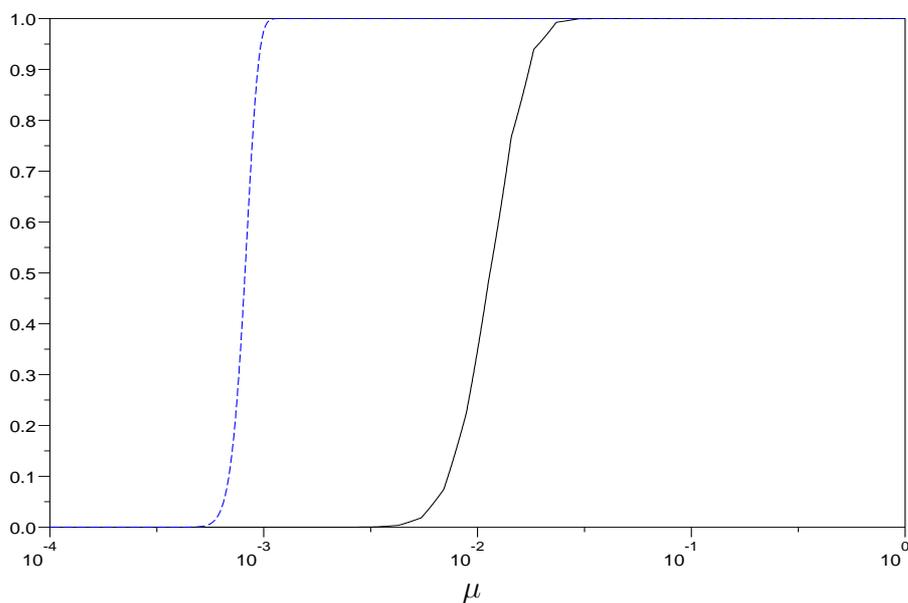


Figura 5.7: Probabilidade de divergência para o LMMN, regressores LA, erro ótimo Gaussiano, $M = 100$, $\sigma_0^2 = 0,01$, $\delta = 0,8$. Traço contínuo: probabilidade observada em 10^5 realizações com 100 iterações. Curva tracejada: modelo teórico.

Probabilidade de divergência para o LMMN, $M = 1.000$

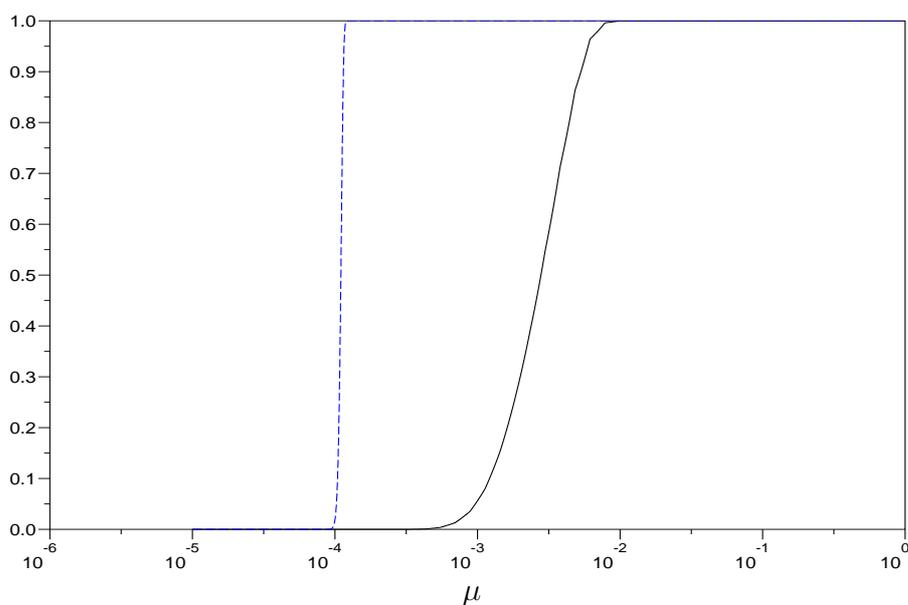


Figura 5.8: Probabilidade de divergência para o LMMN, regressores LA, erro ótimo Gaussiano, $M = 1.000$, $\sigma_0^2 = 0,01$, $\delta = 0,8$. Traço contínuo: probabilidade observada em 10^5 realizações com 100 iterações. Curva tracejada: modelo teórico.

Tabela 5.1: Valores máximos μ_{\max} para o passo de adaptação do LMF para diferentes variâncias do erro ótimo e comprimentos de filtro e regressores Gaussianos, conforme [10].

σ_0^2	M	μ_{\max}	$y(0)$
0,01	100	0,00370	1
0,01	10	0,7692	0.01
0,01	100	0,097087	0.01
0,01	1000	0,00997008	0.01
0,1	100	0,00313	1
0,1	100	0,009709	0.1

6 CONCLUSÕES

Neste trabalho mostra-se que o LMF e o LMMN não podem ser estáveis na média quadrática se o regressor não for estritamente limitado, e propõe-se um modelo teórico (limite superior) para a probabilidade de divergência desses algoritmos para regressores Gaussianos, ao contrário do que é assumido em vários estudos sobre esses algoritmos. Na prática (já que todos os regressores são limitados), os resultados apresentados aqui significam que tanto o LMF quanto o LMMN são bastante sensíveis a valores elevados nos regressores, mesmo que esses valores elevados ocorram muito raramente, como é o caso da distribuição Gaussiana. Outra conclusão é que os passos de adaptação para o LMF e para o LMMN devem ser escolhidos com muito cuidado, principalmente quando uma boa aproximação para a solução ótima de Wiener (isto é, uma boa condição inicial para o algoritmo) não estiver disponível.

Os comportamentos do LMF e do LMS são bem diferentes nesse respeito. Se o vetor de erro $\mathbf{V}(n)$ crescer a um valor elevado em uma dada realização do LMS, o algoritmo tende a fazer com que $\mathbf{V}(n)$ retorne rapidamente a um valor razoável [63]. O LMF e o LMMN, devido às suas não-linearidades cúbicas, perdem completamente a estabilidade se o vetor de erro for levado a um valor elevado.

Os resultados aqui apresentados não significam que as aproximações propostas na literatura para o erro médio quadrático do LMF e do LMMN são inúteis.

Como mostrado na seção 5.3, as aproximações sugeridas em [10] fornecem uma boa aproximação para o ponto (em termos do passo de adaptação) em que a probabilidade de divergência começa a crescer rapidamente. Para passos pequenos, a probabilidade de divergência é muito pequena, e pode-se considerar que as aproximações propostas na literatura estão de fato calculando

$$E\{e(n)^2 \mid \text{o filtro não divergiu}\}.$$

O limite superior proposto aqui fornece ferramentas para projetistas decidirem quando o uso do LMF ou do LMMN seria uma boa escolha. Como passos muito próximos do limite de estabilidade geralmente não levam a desempenho adequado, o fato das aproximações não serem muito próximas dos valores reais para erros iniciais elevados não prejudicar o seu uso. Por outro lado, os resultados apresentados aqui são uma nova forma de estudar o desempenho de filtros adaptativos, que podem levar tanto a uma melhor compreensão do funcionamento de filtros não-lineares quanto a novas idéias para melhorar o desempenho e a robustez de algoritmos para filtragem adaptativa.

Há vários caminhos abertos para se prosseguir na linha de pesquisa aberta com este trabalho, em particular:

- Melhorar as aproximações para a probabilidade de divergência obtidas pelo algoritmo 1. Uma idéia para realizar essa tarefa é melhorar a aproximação $\hat{y}(n)$ (4.12), trocando-se a aproximação

$$E(\mathbf{V}(n)^T \mathbf{V}(n)) \approx E(\mathbf{V}(n))^T E(\mathbf{V}(n)) = \hat{y}(n),$$

por outra que leve em conta que uma variável aleatória com distribuição não-trivial sempre satisfaz $E(\mathbf{V}(n)^T \mathbf{V}(n)) = \gamma E(\mathbf{V}(n))^T E(\mathbf{V}(n))$, com $\gamma \geq 1$. Isso poderia ser feito assumindo-se uma distribuição “média” para $\mathbf{V}(n)$ para o cálculo de γ .

- Aplicar as idéias apresentadas aqui para outros filtros, em particular para estudar a convergência de algoritmos não-supervisionados, como o CMA [26]. De particular interesse seria o estudo da probabilidade de uma realização de um filtro não-supervisionado convergir para um mínimo local.
- A idéia de probabilidade de divergência, ou probabilidade de uma dada realização de um filtro ter um desempenho adequado, não é aplicável apenas a filtros com regressores potencialmente ilimitados, como a distribuição Gaussiana. Mesmo para regressores com distribuição uniforme, por exemplo, pode ser vantajoso aceitar uma pequena probabilidade de divergência (ou seja, usar um passo de adaptação que não garanta convergência) para se ter um desempenho melhor de velocidade de convergência e erro ao acompanhar sinais não estacionários. Note que mesmo no caso de entradas estritamente limitadas, em que [2] demonstra a estabilidade do LMF para passos suficientemente pequenos, um comportamento como o descrito aqui irá acontecer a partir de um certo valor de passo de adaptação, ou seja, os algoritmos não começarão a divergir em todas as realizações para um certo valor-limite do passo, apenas a probabilidade de divergência irá aumentando de maneira contínua.
- Estender a idéia de probabilidade de divergência para probabilidade de se ter um desempenho “adequado”, por exemplo, probabilidade do erro de uma realização de um filtro decrescer abaixo de um certo valor desejado ϵ em um dado número N de iterações.

APÊNDICE A – UMA APROXIMAÇÃO PARA $(\mathbf{V}(N)^T \mathbf{X}(N))^{2K}$

Na seção 4.1 foi usada a aproximação

$$(\mathbf{V}(n)^T \mathbf{X}(n))^{2k} \approx \alpha_M(k) \|\mathbf{V}(n)\|^{2k} \|\mathbf{X}(n)\|^{2k}, \quad k = 1, 2, 3, \quad (\text{A.1})$$

com $\alpha_M(k)$ dado por (4.5), $\alpha_M(k) = \text{E} \left\{ \left(\frac{\mathbf{V}(n)^T \mathbf{X}(n)}{\|\mathbf{V}(n)\| \|\mathbf{X}(n)\|} \right)^{2k} \right\}$.

Neste apêndice são derivadas as expressões para $\alpha_M(k)$ para $k = 1, 2, 3$, sob as seguintes hipóteses:

1. A seqüência *vetorial* $\{\mathbf{X}(n)\}$ é iid,
2. Os elementos de $\mathbf{X}(n)$ são independentes (e portanto $\text{E}\{\mathbf{X}(n)\mathbf{X}(n)^T\} = \sigma_x^2 \mathbf{I}$).

A partir dessas hipóteses e de (4.5), segue que

1. O vetor erro $\mathbf{V}(n)$ é independente de $\mathbf{X}(n)$,
2. Se $\mathbf{V}(n) = \mathbf{0}$ ou $\mathbf{X}(n) = \mathbf{0}$, (A.1) vale qualquer que seja o valor (finito) escolhido para $\alpha_M(k)$,

3. Pode-se reescrever $\alpha_M(k)$ como

$$\alpha_M(k) = \mathbb{E} \left\{ \mathbb{E} \left[\left(\frac{\mathbf{V}(n)^T \mathbf{X}(n)}{\|\mathbf{V}(n)\| \|\mathbf{X}(n)\|} \right)^{2k} \middle| \mathbf{V}(n) \right] \right\},$$

4. Note que no produto interno acima, $\mathbf{V}(n)$ é fixo. Já que $\mathbf{X}(n)$ tem elementos independentes, $\mathbf{X}(n)$ pode apontar para qualquer direção em \mathbb{R}^M com igual probabilidade, e portanto o produto interno é independente da direção particular tomada por $\mathbf{V}(n)$. Portanto, a esperança mais externa é sobre uma constante. Lembrando a hipótese 2. e definindo $\mathbf{e}_1 = [1 \ 0 \ \dots \ 0]^T \in \mathbb{R}^M$, $\bar{\mathbf{X}} \triangleq \mathbf{X}(n)/\|\mathbf{X}(n)\|$, obtém-se

$$\alpha_M(k) = \mathbb{E} \left[(\mathbf{e}_1^T \bar{\mathbf{X}})^{2k} \right], \quad (\text{A.2})$$

em que \mathbf{e}_1 e $\bar{\mathbf{X}}$ têm norma Euclideana unitária.

Antes de atacar a solução geral para (A.2), é conveniente considerar o caso tridimensional. $\bar{\mathbf{X}}$ é um vetor cuja ponta permanece na esfera unitária, o o produto escalar $\mathbf{e}_1^T \bar{\mathbf{X}}$ é igual ao co-seno do ângulo θ entre $\bar{\mathbf{X}}$ e \mathbf{e}_1 (veja a figura A.1).

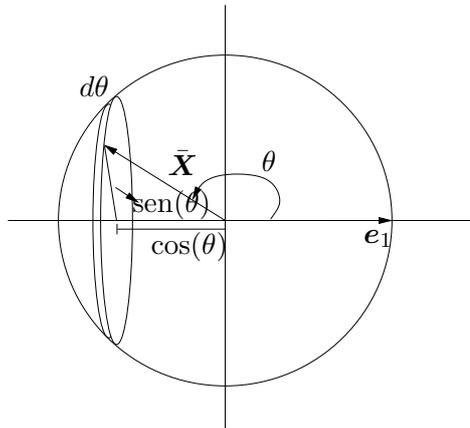


Figura A.1: Cálculo de $\alpha_3(k)$.

Todos os vetores $\bar{\mathbf{X}}$ com pontas no circunferência indicada na figura têm o mesmo ângulo θ com relação a \mathbf{e}_1 . A distância de qualquer ponto no circunferência ao eixo \mathbf{e}_1 é $\sin(\theta)$ (já que o comprimento de $\bar{\mathbf{X}}$ é igual a um, e

$\text{sen}(\theta) = \sqrt{1 - \cos(\theta)^2}$ para $0 \leq \theta \leq \pi$). O elemento de área para todo $\bar{\mathbf{X}}$ com ângulo θ é portanto $2\pi \text{sen}(\theta)d\theta$. Assim pode-se calcular $\alpha_3(k)$ pela expressão

$$\alpha_3(k) = \frac{\int_0^\pi \cos(\theta)^{2k} 2\pi \text{sen}(\theta) d\theta}{\text{Área da esfera unitária}}.$$

A área da esfera unitária é $A_3 = 4\pi$. Esse valor é obtido de

$$A_3 = \int_0^\pi 2\pi \text{sen}(\theta) d\theta.$$

O caso geral M -dimensional é similar. Fixando o ângulo θ entre $\bar{\mathbf{X}}$ e \mathbf{e}_1 , o vetor unitário $\bar{\mathbf{X}}$ fica restrito a uma hiper-esfera $M - 1$ -dimensional com raio $\text{sen}(\theta)$ (pois a projeção de $\bar{\mathbf{X}}$ com relação a \mathbf{e}_1 foi fixada em $\cos(\theta)$, as outras $M - 1$ coordenadas de $\bar{\mathbf{X}}$ devem ter comprimento total $\text{sen}(\theta) = \sqrt{1 - \cos(\theta)^2}$, para $0 \leq \theta \leq \pi$). A expressão para $\alpha_M(k)$ é então

$$\alpha_M(k) = \frac{\int_0^\pi \cos(\theta)^{2k} A_{M-1} \text{sen}(\theta)^{M-2} d\theta}{\int_0^\pi A_{M-1} \text{sen}(\theta)^{M-2} d\theta},$$

em que A_{M-1} é a “área” da superfície de uma hiper-esfera $M - 1$ -dimensional de raio 1, $A_{M-1} \text{sen}(\theta)^{M-2}$ é a área da superfície de uma hiper-esfera $M - 1$ -dimensional de raio $\text{sen}(\theta)$, e o denominador é a área da superfície da hiper-esfera M -dimensional de raio 1 (isto é, o denominador é A_M). Note que A_{M-1} pode ser cancelado, então

$$\alpha_M(k) = \frac{\int_0^\pi \cos(\theta)^{2k} \text{sen}(\theta)^{M-2} d\theta}{\int_0^\pi \text{sen}(\theta)^{M-2} d\theta}. \quad (\text{A.3})$$

Pode-se simplificar mais essa expressão integrando o numerador por partes. Começando com o caso $k = 1$, seja $u = \cos(\theta)$ e $dv = \cos(\theta) \text{sen}(\theta)^{M-2} d\theta$. Obtém-se $du = -\text{sen}(\theta) d\theta$, $v = 1/(M - 1) \text{sen}(\theta)^{M-1}$, e, como $\cos(\theta) \text{sen}(\theta)^{M-1}|_0^\pi = 0$,

$$\alpha_M(1) = \frac{1}{M - 1} \frac{\int_0^\pi \text{sen}(\theta)^M d\theta}{\int_0^\pi \text{sen}(\theta)^{M-2} d\theta}. \quad (\text{A.4})$$

Para $k = 2$ e 3 a integração por partes precisa ser repetida mais uma e duas

vezes, respectivamente, com o resultado final

$$\alpha_M(2) = \frac{3}{(M+1)(M-1)} \frac{\int_0^\pi \text{sen}(\theta)^{M+2} d\theta}{\int_0^\pi \text{sen}(\theta)^{M-2} d\theta} \quad (\text{A.5})$$

$$\alpha_M(3) = \frac{15}{(M+3)(M+1)(M-1)} \frac{\int_0^\pi \text{sen}(\theta)^{M+4} d\theta}{\int_0^\pi \text{sen}(\theta)^{M-2} d\theta}. \quad (\text{A.6})$$

Agora é necessário calcular $\int_0^\pi \text{sen}(\theta)^n d\theta$ para qualquer inteiro $n > 1$. Esta integral pode ser calculada diretamente ou com o auxílio de uma tabela, como [73]. Para o cálculo, procede-se assim:

Para n par, escreve-se

$$\begin{aligned} \text{sen}(\theta)^n &= \left(\frac{e^{j\theta} - e^{-j\theta}}{2j} \right)^n = \\ &= \frac{e^{nj\theta} - ne^{(n-2)j\theta} + \dots - ne^{(2-n)j\theta} + e^{-nj\theta}}{2^n (-1)^{n/2}} = \\ &= \frac{(-1)^{n/2}}{2^{n-1}} \left[\cos(n\theta) - n \cos((n-2)\theta) + \dots + \right. \\ &\quad \left. + (-1)^{n/2} \binom{n}{n/2} \right], \end{aligned}$$

em que $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. Integrando a expressão anterior de 0 a π , o único termo que não desaparece é o último, resultando

$$\int_0^\pi \text{sen}(\theta)^n d\theta = \frac{1}{2^n} \binom{n}{n/2} \pi, \quad n \text{ par}, \quad (\text{A.7})$$

em que $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Para n ímpar, escreve-se

$$\begin{aligned} \int_0^\pi \text{sen}(\theta)^n d\theta &= \int_0^\pi (\text{sen}(\theta)^2)^{\frac{n-1}{2}} \text{sen}(\theta) d\theta = \\ &= \int_0^\pi [1 - \cos(\theta)^2]^{\frac{n-1}{2}} \text{sen}(\theta) d\theta. \end{aligned}$$

Agora aplica-se a mudança de variáveis $x = \cos(\theta)$, $dx = -\operatorname{sen}(\theta)d\theta$, assim

$$\begin{aligned}
 \int_0^\pi \operatorname{sen}(\theta)^n d\theta &= \int_{-1}^1 (1-x^2)^{\frac{n-1}{2}} dx = \\
 &= \int_{-1}^1 \left[1 - \binom{\frac{n-1}{2}}{1} x^2 + \dots + (-1)^{\frac{n-1}{2}} x^{n-1} \right] dx = \\
 &= 2 \sum_{k=0}^{\frac{n-1}{2}} (-1)^k \binom{\frac{n-1}{2}}{k} \frac{1}{2k+1} = \sqrt{\pi} \cdot \frac{\Gamma((n+1)/2)}{\Gamma((n+2)/2)}.
 \end{aligned} \tag{A.8}$$

A última expressão foi obtida usando o Maple VI. $\Gamma(n)$ é a função gama.

Usando (A.7) e (A.8) obtém-se (4.5) (que vale para todo n .) Como $\alpha_1(k) = 1$ para todo k , (4.5) vale para $M \geq 1$.

APÊNDICE B - PROVA DE QUE $D(N) \geq 0$

De (4.7), tem-se (para simplificar a notação, a dependência das variáveis em n será omitida neste apêndice)

$$\begin{aligned} D = D(y) &= 1 - \mu (6\sigma_0^2 - 15\mu\psi_0^4\|\mathbf{X}\|^2) \frac{\|\mathbf{X}\|^2}{M} - \\ &- \mu(6 - 45\mu\sigma_0^2\|\mathbf{X}\|^2) \frac{\|\mathbf{X}\|^4}{M(M+2)}y + \\ &+ 15\mu^2 \frac{\|\mathbf{X}\|^8}{M(M+2)(M+4)}y^2 \end{aligned}$$

Prova-se agora que $D(y) \geq 0$ para todo $y \geq 0$. A prova segue os passos abaixo:

1. Primeiro, prova-se que $D(0) \geq 0$ vale sempre.
2. Em seguida, encontra-se o mínimo D_{\min} de $D(y)$, e condições para que $D_{\min} < 0$.
3. A seguir mostra-se que quando $D_{\min} < 0$, o ponto y_{\min} em que o mínimo é atingido é necessariamente negativo.
4. Conclui-se que, como o coeficiente de y^2 em $D(y)$ é positivo, $D(y)$ é estritamente crescente para $y > y_{\min}$, e portanto 1) e 3) implicam que $D(y) \geq 0$ para $y \geq 0$ se $D_{\min} < 0$ (se $D_{\min} \geq 0$, não há o que provar).

Demonstra-se agora que 1), 2), 3). Lembrando que para qualquer variável aleatória x vale que $0 \leq E(x^2 - E(x^2))^2 = E(x^4) - (E(x^2))^2$, tem-se $\psi_0^4 \geq \sigma_0^4$, e

portanto,

$$\begin{aligned} D(0) &= 1 + 15\mu^2\psi_0^4 \frac{\|\mathbf{X}\|^4}{M} - 6\mu\sigma_0^2 \frac{\|\mathbf{X}\|^2}{M} > \\ &> 1 + 15\mu^2\sigma_0^4 \frac{\|\mathbf{X}\|^4}{M} - 6\mu\sigma_0^2 \frac{\|\mathbf{X}\|^2}{M}. \end{aligned}$$

$D(0)$ pode ser reescrito como

$$\begin{aligned} D(0) &\geq \left(1 - 3\mu\sigma_0^2 \frac{\|\mathbf{X}\|^2}{M}\right)^2 + 15\mu^2\sigma_0^4 \frac{\|\mathbf{X}\|^4}{M} - \\ &- 9\mu^2\sigma_0^4 \frac{\|\mathbf{X}\|^4}{M^2} \geq 0. \end{aligned}$$

Isto prova 1). Para 2), considere primeiramente um polinômio genérico de segundo grau $c + by + ay^2$, with $a > 0$. Seu mínimo é atingido em

$$y_{\min} = \frac{-b}{2a},$$

e o valor mínimo é

$$c + b\frac{-b}{2a} + a\left(\frac{-b}{2a}\right)^2 = c - \frac{b^2}{4a}.$$

Aplicando esse resultado a $D(y)$, lembrando que $\psi_0^4 \geq \sigma_0^4$, e definindo $\beta = \mu\sigma_0^2\|\mathbf{X}\|^2$, obtém-se

$$\begin{aligned} y_{\min} &= \frac{2 - 15\mu\sigma_0^2\|\mathbf{X}\|^2}{10\mu\|\mathbf{X}\|^4}(M + 4), \\ D_{\min} &= 1 + 15\mu^2\psi_0^4 \frac{\|\mathbf{X}\|^4}{M} - 6\mu\sigma_0^2 \frac{\|\mathbf{X}\|^2}{M} - \\ &- 3(M + 4) \frac{4 - 60\mu\sigma_0^2\|\mathbf{X}\|^2 + 225\mu^2\sigma_0^4\|\mathbf{X}\|^4}{20M(M + 2)} = \\ &= 1 - \frac{3}{5} \frac{M + 4}{M(M + 2)} - \\ &- \frac{60(2 - 5\beta\psi_0^4/\sigma_0^4)\beta(M + 2) + (M + 4)\beta(675\beta - 180)}{20M(M + 2)} \geq \\ &\geq 1 - \frac{3}{5} \frac{M + 4}{M(M + 2)} - \frac{(75M + 420)\beta - (12M + 96)\beta}{4M(M + 2)}. \end{aligned}$$

Agora demonstra-se 3). D_{\min} pode ser negativo apenas se $(75M+420)\mu\sigma_0^2\|\mathbf{X}\|^2 - (12M+96) > 0$. Nesta condição, y_{\min} é limitado por

$$\begin{aligned} y_{\min} &< \frac{M+4}{10\mu\|\mathbf{X}\|^4} \left(2 - 15 \frac{12M+96}{75M+420} \right) = \\ &= -\frac{M+4}{10\mu\|\mathbf{X}\|^4} \frac{2M+40}{5M+28} < 0. \end{aligned}$$

Isto conclui a demonstração.

REFERÊNCIAS

- 1 WALACH, E.; WIDROW, B. The least mean fourth (LMF) adaptive algorithm and its family. *IEEE Trans. Inform. Theory*, IT-30, n. 2, p. 275–283, mar. 1984.
- 2 SETHARES, W. A. Adaptive algorithms with nonlinear data and error functions. *IEEE Trans. Signal Processing*, v. 40, n. 9, p. 2199–2206, set. 1992.
- 3 YOUSEF, N.; SAYED, A. H. Tracking analysis of the LMF and LMMN adaptive algorithms. In: *Conference Record of the Thirty-Third Asilomar Conference on Signals, Systems, and Computers*. Pacific Grove, EUA: [s.n.], 1999. v. 1, p. 786–790.
- 4 KOIKE, S. Stability conditions for adaptive algorithms with non-quadratic error criteria. In: *EUSIPCO2000*. Tampere, Finland: [s.n.], 2000. p. 131–134.
- 5 AL-NAFFOURI, T. Y.; SAYED, A. H. Adaptive filters with error nonlinearities: mean-square analysis and optimum design. In: *EURASIP Journal on Applied Signal Processing*. [S.l.: s.n.], 2001. p. 192–205.
- 6 KOIKE, S. Analysis of the least mean fourth algorithm based on gaussian distributed tap weights. In: IEEE. *IEEE International Symposium on Intelligent Signal Processing and Communication Systems, ISPACS 2001*. Nashville, USA, 2001.
- 7 AL-NAFFOURI, T. Y.; SAYED, A. H. Transient analysis of adaptive filters with error nonlinearities. *IEEE Trans. Signal Processing*, v. 51, n. 3, p. 653–663, mar. 2003.
- 8 SAYED, A. H.; AL-NAFFOURI, T. Y.; NASCIMENTO, V. H. Energy conservation in adaptive filtering. In: BARNER, K. E.; ARCE, G. R. (Ed.). *Nonlinear Signal and Image Processing: Theory, Methods, and Applications*. [S.l.]: CRC Press, 2003, (The Electrical Engineering and Applied Signal Processing Series). cap. 1, p. 223–255.
- 9 HÜBSCHER, P. I.; BERMUDEZ, J. C. M. An improved statistical analysis of the least mean fourth (LMF) adaptive algorithm. *IEEE Trans. Signal Processing*, v. 51, n. 3, p. 664–671, mar. 2003.
- 10 HÜBSCHER, P. I.; NASCIMENTO, V. H.; BERMUDEZ, J. C. M. New results on the stability analysis of the LMF (least mean fourth) adaptive algorithm. In: *Proceedings of the 2003 IEEE International Conference on Acoustics, Speech and Signal Processing*. [S.l.: s.n.], 2003. VI, p. VI 369–VI 372.

- 11 DINIZ, P. S. R. *Adaptive Filtering: Algorithms and Practical Implementation*. 2nd. ed. [S.l.]: Kluwer Academic Publishers, 2002.
- 12 SAYED, A. H. *Fundamentals of Adaptive Filtering*. [S.l.]: Wiley-Interscience, 2003.
- 13 HAYKIN, S. *Adaptive Filter Theory*. 4th. ed. [S.l.]: Prentice Hall, 2001.
- 14 HANSEN, C. H.; SNYDER, S. D. *Active Control of Noise and Vibration*. [S.l.]: E & FN Spon, 1997.
- 15 MESSERSCHMITT, D. G. Echo cancellation in speech and data transmission. *IEEE J. Select. Areas Commun.*, SAC-2, n. 2, p. 283–297, mar. 1984.
- 16 LJUNG, L. *System Identification: Theory for the User*. [S.l.]: Prentice Hall, 1987. (Information and System Sciences Series).
- 17 GITLIN, R. D.; HAYES, J. S.; WEINSTEIN, S. B. *Data Communications Principles*. NY: Plenum Press, 1992.
- 18 CHEN, J.-T.; TSAI, H.-S.; CHEN, Y.-K. Adaptive joint linearisation/equalisation with delay alignments for a wideband power amplifier. *Electronics Letters*, v. 34, n. 6, p. 514–515, mar. 1998.
- 19 KUO, S. M.; MORGAN, D. R. Active noise control: a tutorial review. *Proceedings of the IEEE*, v. 87, n. 6, p. 943–973, jun. 1999.
- 20 PROAKIS, J. G. *Digital Communications*. [S.l.]: McGraw-Hill, 2001.
- 21 NASCIMENTO, V. H.; BERMUDEZ, J. C. M. When is the least-mean fourth algorithm mean-square stable? In: *Proceedings of the 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing*. [S.l.: s.n.], 2005. IV, p. 341–344.
- 22 _____. On the stability of the least-mean fourth (LMF) algorithm. In: *Anais do XXI Simpósio Brasileiro de Telecomunicações*. [S.l.: s.n.], 2004. p. 1–5. [Http://www.lps.usp.br/~vitor/sbrt2004.pdf](http://www.lps.usp.br/~vitor/sbrt2004.pdf).
- 23 _____. Probability of divergence for the least-mean fourth (LMF) algorithm. Accepted for publication in the *IEEE Transactions on Signal Processing*. 2005.
- 24 TAN, L.; JIANG, J. Adaptive Volterra filters for active control of nonlinear noise processes. *IEEE Trans. Signal Processing*, v. 49, n. 8, p. 1667–1676, ago. 2001.
- 25 SICURANZA, G. L.; CARINI, A.; FERMO, A. Nonlinear adaptive filters for acoustic echo cancellation in mobile terminals. In: BARNER, K. E.; ARCE, G. R. (Ed.). *Nonlinear Signal and Image Processing: Theory, Methods, and Applications*. [S.l.]: CRC Press, 2003, (The Electrical Engineering and Applied Signal Processing Series). cap. 7, p. 223–255.

- 26 GODARD, D. N. Self-recovering equalization and carrier tracking in two-dimensional data communication systems. *IEEE Trans. Commun.*, v. 28, n. 11, p. 1867–1875, nov. 1980.
- 27 FIJALKOW, I.; MANLOVE, C. E.; Johnson Jr., C. R. Adaptive fractionally spaced blind CMA equalization: excess MSE. *IEEE Trans. Signal Processing*, v. 46, n. 1, p. 227–231, jan. 1998.
- 28 SOLO, V.; KONG, X. *Adaptive Signal Processing Algorithms*. Englewood Cliffs, NJ: Prentice Hall, 1995.
- 29 MOUSTAKIDES, G. V. Exponential convergence of products of random matrices: application to adaptive algorithms. *International Journal of Adaptive Control and Signal Processing*, v. 12, n. 7, p. 579–597, nov. 1998.
- 30 EWEDA, E. Comparison of RLS, LMS, and sign algorithms for tracking randomly time-varying channels. *IEEE Trans. Signal Processing*, v. 42, n. 11, p. 2937–2944, nov. 1994.
- 31 EWEDA, E.; MACCHI, O. Tracking error bounds of adaptive nonstationary filtering. *Automatica*, p. 296–302, 1985.
- 32 SLOCK, D. T. M. On the convergence behavior of the LMS and the normalized LMS algorithms. *IEEE Trans. Signal Processing*, v. 41, n. 9, p. 2811–2825, set. 1993.
- 33 MIRANDA, M. D.; GERKEN, M. A hybrid least squares QR-lattice algorithm using a priori errors. *IEEE Trans. Signal Processing*, v. 45, n. 12, p. 2900–2911, dez. 1997.
- 34 REGALIA, P. A. Numerical stability issues in fast least-squares adaptation algorithms. *Optical Engineering*, v. 31, n. 6, p. 1144–1152, jun. 1992.
- 35 BREINING, C. et al. Acoustic echo control – an application of very-high-order adaptive filters. *IEEE Signal Processing Magazine*, p. 42–69, jul. 1999.
- 36 KWONG, R. H.; JOHNSTON, E. W. A variable step size LMS algorithm. *IEEE Trans. Signal Processing*, v. 40, n. 7, p. 1633–1642, jul. 1992.
- 37 ABOULNASR, T.; MAYYAS, K. A robust variable step-size LMS-type algorithm. *IEEE Trans. Signal Processing*, v. 45, n. 3, p. 631–639, mar. 1997.
- 38 MOUSTAKIDES, G. V.; THEODORIDES, S. Fast Newton transversal filters — a new class of adaptive estimation algorithms. *IEEE Trans. Signal Processing*, v. 39, n. 10, p. 2184–2193, out. 1991.
- 39 GAY, S. L.; TAVATHIA, S. The fast affine projection algorithm. In: *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing*. [S.l.: s.n.], 1995. v. 5, p. 3023–3026.

- 40 ALMEIDA, S. J. M. de et al. A stochastic model for the convergence behavior of the affine projection algorithm for gaussian inputs. In: *Proceedings of the 2003 IEEE International Conference on Acoustics, Speech and Signal Processing (CDROM)*. [S.l.: s.n.], 2003. VI, p. VI 313–VI 316.
- 41 AHLÉN, A.; LINDBOM, L.; SERNAD, M. Analysis of stability and performance of adaptation algorithms with time-invariant gains. *IEEE Trans. Signal Processing*, v. 52, n. 1, p. 103–116, jan. 2004.
- 42 BENESTY, J.; GAY, S. L. An improved PNLMS algorithm. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*. [S.l.], 2002. v. 2, p. 1881–1884.
- 43 MARTIN, R. K. et al. Exploiting sparsity in adaptive filters. *IEEE Trans. Signal Processing*, v. 50, n. 8, p. 1883–1894, ago. 2002.
- 44 NASCIMENTO, V. H. Improving the initial convergence of adaptive filters: the variable-length LMS filter. In: *IEEE/EURASIP. Proceedings of the 14th IEEE International Conference on Digital Signal Processing (DSP2002), Santorini, Greece*. [S.l.], 2002. v. 2, p. 667–670.
- 45 MERCHED, R.; PETRAGLIA, M.; DINIZ, P. S. R. A delayless alias-free subband adaptive filter structure. *IEEE Trans. Signal Processing*, v. 47, n. 6, p. 1580–1591, jun. 1999.
- 46 WEISS, S.; STEWART, R. W.; RABENSTEIN, R. Steady-state performance limitations of subband adaptive filters. *IEEE Trans. Signal Processing*, v. 49, n. 9, p. 1982–1991, set. 2001.
- 47 NGIA, L. S. H. Recursive identification of acoustic echo systems using orthonormal basis functions. *IEEE Transactions on Speech and Audio Processing*, v. 11, n. 3, p. 278–293, maio 2003.
- 48 MERCHED, R. Extended RLS lattice adaptive filters. *IEEE Trans. Signal Processing*, v. 51, n. 9, p. 2294–2309, set. 2003.
- 49 Jojoa Gómez, P. E. *Análise de Algoritmo Adaptativo que Ajusta a Diferença de Segunda Ordem dos Parâmetros*. Dissertação (Mestrado) — Escola Politécnica da Universidade de São Paulo, 1999.
- 50 Jojoa Gómez, P. E. *Um Algoritmo Acelerador de Parâmetros*. Tese (Doutorado) — Escola Politécnica da Universidade de São Paulo, 2003.
- 51 Jojoa Gómez, P. E.; GERKEN, M.; PAIT, F. M. Sobre o algoritmo acelerador para filtragem adaptativa. In: *Anais do 19º Simpósio Brasileiro de Telecomunicações*. Fortaleza, CE: [s.n.], 2001.
- 52 _____. The accelerating adaptive filtering algorithm. In: *Workshop on Adaptation and Learning in Control and Signal Processing (ALCOSP 2001)*. Cernobbio-Como: [s.n.], 2001.

- 53 GERKEN, M.; PAIT, F. M.; Jojoa Gómez, P. E. An adaptive algorithm with parameter acceleration. In: *Proceedings of the 2000 IEEE International Conference on Acoustics, Speech and Signal Processing (CDROM)*. Istanbul: [s.n.], 2000. p. 1–4.
- 54 KUSHNER, H. J.; YIN, G. G. *Stochastic Approximation and Recursive Algorithms and Applications*. [S.l.]: Springer, 2003. (Applications of Mathematics, v. 35).
- 55 FLORIAN, S.; FEUER, A. Performance analysis of the LMS algorithm with a tapped delay line (two-dimensional case). *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-34, n. 6, p. 1542–1549, dez. 1986.
- 56 DOUGLAS, S. C.; PAN, W. Exact expectation analysis of the LMS adaptive filter. *IEEE Trans. Signal Processing*, v. 43, n. 12, p. 2863–2871, dez. 1995.
- 57 NASCIMENTO, V. H. *Stability and Performance of Adaptive Filters without Slow Adaptation Approximations*. Tese (Doutorado) — University of California, Los Angeles, 1999.
- 58 GUO, L.; LJUNG, L.; WANG, G.-J. Necessary and sufficient conditions for stability of LMS. *IEEE Trans. Automat. Contr.*, v. 42, n. 6, p. 761–770, jun. 1997.
- 59 SOLO, V. The stability of LMS. *IEEE Trans. Signal Processing*, v. 45, n. 12, p. 3017–3026, dez. 1997.
- 60 IOANNOU, P. A.; SUN, J. *Robust Adaptive Control*. NJ: PTR Prentice Hall, 1996.
- 61 NASCIMENTO, V. H.; SAYED, A. H. Unbiased and stable leakage-based adaptive filters. *IEEE Trans. Signal Processing*, v. 47, n. 12, p. 3261–3276, dez. 1999.
- 62 BITMEAD, R. R.; ANDERSON, B. D. O. Adaptive frequency sampling filters. *IEEE Trans. Circuits Syst.*, CAS-28, n. 6, p. 524–533, jun. 1981.
- 63 NASCIMENTO, V. H.; SAYED, A. H. On the learning mechanism of adaptive filters. *IEEE Trans. Signal Processing*, v. 48, n. 6, p. 1609–1625, jun. 2000.
- 64 CHAMBERS, J. A.; TANRIKULU, O.; CONSTANTINIDES, A. G. Least mean mixed-norm adaptive filtering. *Electronics Letters*, v. 30, n. 19, set. 15 1994.
- 65 DOUGLAS, S. C.; MENG, T. H.-Y. Stochastic gradient adaptation under general error criteria. *IEEE Trans. Signal Processing*, v. 42, n. 6, p. 1335–1351, jun. 1994.
- 66 SHARMA, R.; SETHARES, W. A.; BUCKLEW, J. A. Asymptotic analysis of stochastic gradient-based adaptive filtering algorithms with general cost functions. *IEEE Trans. Signal Processing*, v. 44, n. 9, p. 2186–2194, set. 1996.

- 67 TANRIKULU, O.; CHAMBERS, J. A. Convergence and steady-state properties of the least-mean mixed-norm (LMMN) adaptive algorithm. *IEEE Trans. Signal Processing*, v. 143, n. 3, p. 137–142, jun. 1996.
- 68 SETHARES, W. A. The least mean square family. In: KALOUPSIDIS, N.; THEODORIDIS, S. (Ed.). *Adaptive System Identification and Signal Processing Algorithms*. [S.l.]: Prentice Hall, 1993.
- 69 YOUSEF, N.; SAYED, A. H. A unified approach to the steady-state and tracking analysis of adaptive filters. *IEEE Trans. Signal Processing*, v. 49, n. 2, p. 314–324, fev. 2001.
- 70 RUPP, M.; SAYED, A. H. A time-domain feedback analysis of filtered-error adaptive gradient algorithms. *IEEE Trans. Signal Processing*, v. 44, p. 1428–1439, 1996.
- 71 SAYED, A. H.; RUPP, M. Error-energy bounds for adaptive gradient algorithms. *IEEE Trans. Signal Processing*, v. 44, n. 8, p. 1982–1989, ago. 1996.
- 72 SAYED, A. H.; NASCIMENTO, V. H. Energy conservation and the learning ability of LMS adaptive filters. In: HAYKIN, S.; WIDROW, B. (Ed.). *Least-Mean-Square Adaptive Filters*. [S.l.]: Wiley, 2003.
- 73 GRADSHTEYN, I. S.; RYZHIK, I. M.; JEFFREY, A. *Table of Integrals, Series, and Products — Corrected and Enlarged Edition*. [S.l.]: Academic Press, 1980.