

Modelo Estocástico do Algoritmo LMS de Passo Variável Baseado no Gradiente do Sinal de Erro Quadrático

José Gil F. Zipf, Orlando J. Tobias e Rui Seara

Resumo—Este artigo discute o algoritmo LMS de passo de adaptação variável baseado no gradiente proposto por Okello. Um modelo estocástico para o comportamento médio dos pesos e para a curva de aprendizagem do algoritmo em questão é apresentado. A principal característica do algoritmo de Okello é o uso da potência média do gradiente estimado para a atualização do passo de adaptação. Uma análise em regime permanente é também considerada. Simulações numéricas, comparando os resultados obtidos a partir do método de Monte Carlo e através do modelo proposto, permitem avaliar a qualidade das previsões obtidas.

Palavras-chave—Algoritmo de Okello, algoritmo de passo variável baseado no gradiente, comportamento médio dos pesos, curva de aprendizagem.

Abstract—This paper discusses the gradient-based variable step-size LMS algorithm proposed by Okello. A stochastic model for the mean weight behavior and the learning curve of the referred algorithm is presented. The main characteristic of Okello's algorithm is the use of the average power of the gradient estimate for updating the variable step size. A steady-state analysis is also carried out. Numerical simulations comparing the results obtained via Monte Carlo and the proposed model allow assessing the accuracy of the achieved model predictions.

Keywords—Okello's algorithm, gradient-based variable step-size algorithm, mean weight behavior, learning curve.

I. INTRODUÇÃO

Nas duas últimas décadas, o algoritmo LMS (*least-mean-square*) vem sendo amplamente utilizado em um número considerável de aplicações de filtragem adaptativa devido essencialmente à sua simplicidade e robustez [1], [2]. Nesse algoritmo adaptativo, o parâmetro que controla a velocidade de convergência tem um grande impacto sobre o erro em excesso obtido. Assim, de forma geral, quanto maior

for o valor do passo de adaptação, maior será a velocidade de convergência. Em contrapartida, o erro em excesso também aumentará proporcionalmente, o que é indesejável. Em muitas aplicações, tais como identificação de sistemas [2], cancelamento de ruído [2], cancelamento de eco [2], equalização de canal [2], dentre outras, é desejável que o filtro adaptativo apresente simultaneamente uma alta velocidade de convergência associada a um pequeno desajuste em regime permanente. É nesse contexto de aplicações que os algoritmos LMS de passo variável (*variable step-size LMS* – VSSLMS) são requeridos, visando ajustar dinamicamente, durante o processo de convergência, o passo de adaptação. Então, várias estratégias de ajuste de passo vêm sendo propostas na literatura, as quais são baseadas em diferentes princípios, tais como no gradiente do erro quadrático [3]-[5], no erro quadrático instantâneo [6], na autocorrelação do erro [7], no valor do erro absoluto [8], na normalização do vetor de erros [9] e em outros critérios aqui não nomeados [10]-[11].

Em [4], é proposto um algoritmo VSSLMS bastante simples e eficiente, baseado na estimativa do gradiente do erro quadrático, aqui denominado algoritmo de Okello. Esse algoritmo exibe uma alta velocidade de convergência e uma muito boa imunidade ao ruído aditivo. O comportamento do algoritmo, discutido em [4], é avaliado através de resultados de simulação, não sendo apresentado qualquer modelo analítico que possibilite prever o seu desempenho. Por outro lado, com o modelo analítico do algoritmo em mãos, pode-se entender melhor o seu comportamento assim como propor modificações que melhorem o seu desempenho para certas aplicações. Neste trabalho, o algoritmo VSSLMS proposto em [4] é revisitado e um modelo estocástico para o comportamento médio do vetor de coeficientes como também para a curva de aprendizagem é proposto. Além disso, uma breve análise do comportamento do algoritmo em regime permanente é realizada. A precisão do modelo aqui proposto é verificada através de simulações numéricas.

II. ALGORITMO DE OKELLO

O algoritmo de Okello é discutido nesta seção considerando um problema de identificação de sistemas, ilustrado pela Fig. 1.

José Gil F. Zipf, Orlando J. Tobias e Rui Seara, LINSE – Laboratório de Circuitos e Processamento de Sinais, Depto. de Eng. Elétrica, Universidade Federal de Santa Catarina, Florianópolis, SC, E-mails: {gil, orlando, seara}@linse.ufsc.br.

José Gil F. Zipf e Orlando J. Tobias são também Professores do Departamento de Engenharia Elétrica e Telecomunicações da Universidade Regional de Blumenau (FURB), Blumenau, SC, Brasil.

Este trabalho foi parcialmente financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

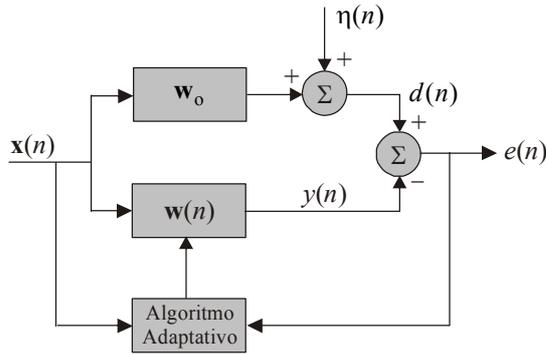


Fig. 1. Diagrama em blocos de um esquema de identificação de sistema.

A saída do sistema desconhecido, corrompida por um ruído aditivo $\eta(n)$ (ver Fig. 1), é dada por

$$d(n) = \mathbf{w}_0^T \mathbf{x}(n) + \eta(n) \quad (1)$$

onde $\mathbf{x}(n) = [x(n) \ x(n-1) \ \dots \ x(n-N+1)]^T$ denota o vetor de entrada, sendo $\{x(n)\}$ um processo gaussiano de média zero com variância σ_x^2 , $\eta(n)$ é um ruído de medição gaussiano com média zero e variância σ_η^2 e $\mathbf{w}_0 = [w_{0,1}(n) \ w_{0,2}(n-1) \ \dots \ w_{0,N-1}(n-N+1)]^T$ caracteriza a resposta ao impulso do sistema a ser identificado. O sinal de erro é dado por

$$e(n) = d(n) - y(n) \quad (2)$$

onde $y(n) = \mathbf{w}^T(n) \mathbf{x}(n)$ é a saída do filtro adaptativo e $\mathbf{w}(n) = [w(n) \ w(n-1) \ \dots \ w(n-N+1)]^T$, o vetor de coeficientes do filtro, cuja equação de atualização é [4]

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu(n)e(n)\mathbf{x}(n). \quad (3)$$

Em (3), $\mu(n)$ representa o passo de adaptação variável. O algoritmo de Okello realiza uma estimativa da correlação cruzada entre o sinal de erro e cada componente do vetor de entrada, a qual é dada por

$$g_i(n) = \beta g_i(n-1) + e(n)x(n-i) \quad (4)$$

onde $0 < \beta < 1$ é o parâmetro de controle do processo de estimação. O valor do passo variável é então ajustado recursivamente através da seguinte expressão [4]:

$$\mu(n) = \alpha \mu(n-1) + \gamma \sum_{i=0}^N g_i^2(n), \quad 0 < \alpha < 1, \gamma > 0. \quad (5)$$

De modo a garantir uma operação estável do algoritmo, os limites de variação do passo de adaptação são fixados pela seguinte inequação [2]:

$$0 < \mu(n) < \frac{2}{3\text{tr}(\mathbf{R})} \quad (6)$$

onde $\mathbf{R} = E[\mathbf{x}(n)\mathbf{x}^T(n)]$ representa a matriz de autocorrelação do vetor de entrada.

III. MODELO PARA O COMPORTAMENTO MÉDIO DO VETOR DE COEFICIENTES

O modelo aqui proposto é derivado à luz da Teoria de Independência (TI) [2], levando em conta também as seguintes hipóteses simplificativas:

- i) O passo variável de adaptação $\mu(n)$ é independente e descorrelacionado de qualquer outro sinal no sistema.
- ii) $E[\mu^2(n)] \approx E^2[\mu(n)]$.

As hipóteses (i) e (ii) são válidas sob condições de adaptação lenta.

O modelo para o comportamento médio do vetor de coeficientes é obtido tomando o valor esperado de (3), levando em conta a hipótese (i). Dessa forma, tem-se

$$E[\mathbf{w}(n+1)] = E[\mathbf{w}(n)] + E[\mu(n)]\{\mathbf{p} - \mathbf{R}E[\mathbf{w}(n)]\} \quad (7)$$

onde $\mathbf{p} = E[d(n)\mathbf{x}(n)]$ é o vetor de correlação cruzada entre o sinal desejado e o vetor de entrada. Para se obter o modelo desejado, $E[\mu(n)]$ deve ser determinado em (7).

O ajuste do passo variável, em (5), pode ser expresso em termos do vetor gradiente $\mathbf{g}(n)$ como

$$\mu(n) = \alpha \mu(n-1) + \gamma \mathbf{g}^T(n) \mathbf{g}(n) \quad (8)$$

com o vetor gradiente obtido por

$$\mathbf{g}(n) = \beta \mathbf{g}(n-1) + e(n)\mathbf{x}(n). \quad (9)$$

Agora, tomando o valor esperado de ambos os lados de (8), pode-se então escrever

$$E[\mu(n)] = \alpha E[\mu(n-1)] + \gamma E[\mathbf{g}^T(n) \mathbf{g}(n)]. \quad (10)$$

Para resolver (10), o segundo termo no lado direito deve ser determinado. Assim, utilizando (9), obtêm-se

$$\begin{aligned} \mathbf{g}^T(n) \mathbf{g}(n) &= \beta^2 \mathbf{g}^T(n-1) \mathbf{g}(n-1) + \beta e(n) \mathbf{g}^T(n-1) \mathbf{x}(n) \\ &\quad + \beta e(n) \mathbf{x}^T(n) \mathbf{g}(n-1) + e^2(n) \mathbf{x}^T(n) \mathbf{x}(n). \end{aligned} \quad (11)$$

Então, rearranjando (11), tem-se

$$\begin{aligned} \mathbf{g}^T(n) \mathbf{g}(n) &= \beta^2 \mathbf{g}^T(n-1) \mathbf{g}(n-1) + 2\beta e(n) \mathbf{g}^T(n-1) \mathbf{x}(n) \\ &\quad + e^2(n) \mathbf{x}^T(n) \mathbf{x}(n). \end{aligned} \quad (12)$$

Finalmente, determinando o valor esperado de ambos os lados de (12), $E[\mathbf{g}^T(n) \mathbf{g}(n)]$ pode ser obtido de forma recursiva por

$$\begin{aligned} E[\mathbf{g}^T(n) \mathbf{g}(n)] &= \beta^2 E[\mathbf{g}^T(n-1) \mathbf{g}(n-1)] \\ &\quad + 2\beta E[e(n) \mathbf{x}^T(n) \mathbf{g}(n-1)] + E[e^2(n) \mathbf{x}^T(n) \mathbf{x}(n)]. \end{aligned} \quad (13)$$

Em (13), o termo $E[e^2(n) \mathbf{x}^T(n) \mathbf{x}(n)]$ pode ser aproximado por

$$\begin{aligned} E[e^2(n) \mathbf{x}^T(n) \mathbf{x}(n)] &\approx E[e^2(n)] E[\mathbf{x}^T(n) \mathbf{x}(n)] \\ &\approx J(n) \text{tr}(\mathbf{R}) \end{aligned} \quad (14)$$

onde $J(n)$ representa a curva de aprendizagem do algoritmo.

O termo $E[e(n)\mathbf{x}^T(n)\mathbf{g}(n-1)]$ é obtido sob a condição de adaptação lenta, resultando na seguinte aproximação:

$$\begin{aligned} E[e(n)\mathbf{x}^T(n)\mathbf{g}(n-1)] &\approx E[e(n)\mathbf{x}^T(n)]E[\mathbf{g}(n-1)] \\ &\approx \{\mathbf{p} - \mathbf{R}E[\mathbf{w}(n)]\}^T E[\mathbf{g}(n-1)]. \end{aligned} \quad (15)$$

Então, para finalizar, $E[\mathbf{g}(n-1)]$ pode ser obtido, de forma recursiva, tomando o valor esperado de ambos os lados de (9). Assim,

$$E[\mathbf{g}(n)] = \beta E[\mathbf{g}(n-1)] + \{\mathbf{p} - \mathbf{R}E[\mathbf{w}(n)]\}. \quad (16)$$

IV. CURVA DE APRENDIZAGEM

Nesta seção, é derivado um modelo para prever a curva de aprendizagem do algoritmo de Okello. A partir desse modelo, tanto o desajuste do algoritmo quanto o valor máximo do passo de adaptação podem ser determinados, considerando que o algoritmo opere na região de estabilidade.

Então, definindo o vetor de erro nos coeficientes como $\mathbf{v}(n) = \mathbf{w}(n) - \mathbf{w}_o$ e usando (3), obtém-se

$$\mathbf{v}(n+1) = \mathbf{v}(n) + \mu(n)e(n)\mathbf{x}(n). \quad (17)$$

Representando agora o sinal de erro como uma função do vetor $\mathbf{v}(n)$, tem-se

$$\begin{aligned} e(n) &= d(n) - [\mathbf{x}^T(n)\mathbf{w}(n)] \\ &= d(n) - \mathbf{x}^T(n)\mathbf{v}(n) - \mathbf{x}^T(n)\mathbf{w}_o. \end{aligned} \quad (18)$$

Substituindo então (18) em (17), obtém-se

$$\begin{aligned} \mathbf{v}(n+1) &= \mathbf{v}(n) + \mu(n)[d(n)\mathbf{x}(n) \\ &\quad - \mathbf{x}(n)\mathbf{x}^T(n)\mathbf{v}(n) - \mathbf{x}(n)\mathbf{x}^T(n)\mathbf{w}_o]. \end{aligned} \quad (19)$$

Como

$$d(n) - \mathbf{x}^T(n)\mathbf{w}_o = \eta(n) \quad (20)$$

pode-se então escrever

$$\mathbf{v}(n+1) = \mathbf{v}(n) + \mu(n)[\eta(n)\mathbf{x}(n) - \mathbf{x}(n)\mathbf{x}^T(n)\mathbf{v}(n)]. \quad (21)$$

Agora, definindo a matriz de covariância do vetor de erro nos coeficientes como $\mathbf{K}(n) = E[\mathbf{v}(n)\mathbf{v}^T(n)]$, considerando as hipóteses (i) e (ii) e também o teorema de fatoração de momentos para sinais gaussianos [2], obtém-se

$$\begin{aligned} \mathbf{K}(n+1) &= \mathbf{K}(n) - E[\mu(n)][\mathbf{R}\mathbf{K}(n) + \mathbf{K}(n)\mathbf{R}] \\ &\quad + E^2[\mu(n)]\{\mathbf{R}\text{tr}[\mathbf{R}\mathbf{K}(n)] + 2\mathbf{R}\mathbf{K}(n)\mathbf{R}\} \\ &\quad + E^2[\mu(n)]\mathbf{R}\sigma_\eta^2. \end{aligned} \quad (22)$$

Finalmente, a expressão para a curva de aprendizagem é dada por

$$J(n) = E[e^2(n)] = \sigma_\eta^2 + \text{tr}[\mathbf{R}\mathbf{K}(n)]. \quad (23)$$

V. ANÁLISE EM REGIME PERMANENTE

Assumindo que o filtro adaptativo converge, apresenta-se aqui uma análise em regime permanente do vetor gradiente estimado $\mathbf{g}(n)$, do passo variável $\mu(n)$, do vetor de coeficientes $\mathbf{w}(n)$ e do erro quadrático médio (EQM).

A. Vetor Gradiente Estimado

Considerando (16) para $n \rightarrow \infty$, obtém-se

$$\begin{aligned} \mathbf{g}(\infty) &= \beta\mathbf{g}(\infty) + \{\mathbf{p} - \mathbf{R}E[\mathbf{w}(\infty)]\} \\ &= \frac{\mathbf{p} - \mathbf{R}E[\mathbf{w}(\infty)]}{(1-\beta)} \end{aligned} \quad (24)$$

onde $\mathbf{w}(\infty)$ é o valor em regime permanente do vetor de coeficientes. Tomando agora (13) para $n \rightarrow \infty$, obtém-se a seguinte expressão aproximada:

$$\mathbf{g}^T(\infty)\mathbf{g}(\infty) \approx \frac{2\beta J(\infty)\text{tr}(\mathbf{R})}{(1-\beta^2)}. \quad (25)$$

B. Passo de Adaptação Variável

O valor em regime permanente do passo de adaptação, obtido de (10), é dado por

$$\mu(\infty) = \frac{2\beta\gamma J(\infty)\text{tr}(\mathbf{R})}{(1-\beta^2)(1-\alpha)}. \quad (26)$$

C. Vetor de Coeficientes

A partir de (7), o valor em regime permanente do vetor $\mathbf{w}(n)$ é

$$\mathbf{w}(\infty) = \mathbf{w}(\infty) + \mu(\infty)[\mathbf{p} - \mathbf{R}\mathbf{w}(\infty)] \quad (27)$$

levando a

$$\mu(\infty)[\mathbf{p} - \mathbf{R}\mathbf{w}(\infty)] = 0. \quad (28)$$

Agora, dado que $[\mathbf{p} - \mathbf{R}\mathbf{w}(\infty)] = 0$, obtém-se então

$$\mathbf{w}(\infty) = \mathbf{R}^{-1}\mathbf{p}. \quad (29)$$

D. EQM em Excesso

Assumindo que o passo de adaptação (variável) $\mu(n)$ converge para um valor constante $\mu(\infty)$ quando $n \rightarrow \infty$, pode-se utilizar a mesma expressão para cálculo do erro em excesso do algoritmo LMS convencional. Assim, observando a condição de adaptação lenta, tem-se

$$J_{\text{ex}}(\infty) = \frac{\mu(\infty)J_{\text{min}}}{2} \sum_{k=1}^N \lambda_k = \frac{\beta\gamma J(\infty)}{2(1-\beta^2)(1-\alpha)} \sigma_\eta^2 [\text{tr}(\mathbf{R})]^2. \quad (30)$$

Como $J_{\text{ex}}(\infty) = J(\infty) - J_{\text{min}}$, obtém-se

$$J(\infty) = \frac{\sigma_\eta^2}{1 - \frac{\beta\gamma}{2(1-\beta^2)(1-\alpha)} \sigma_\eta^2 [\text{tr}(\mathbf{R})]^2}. \quad (31)$$

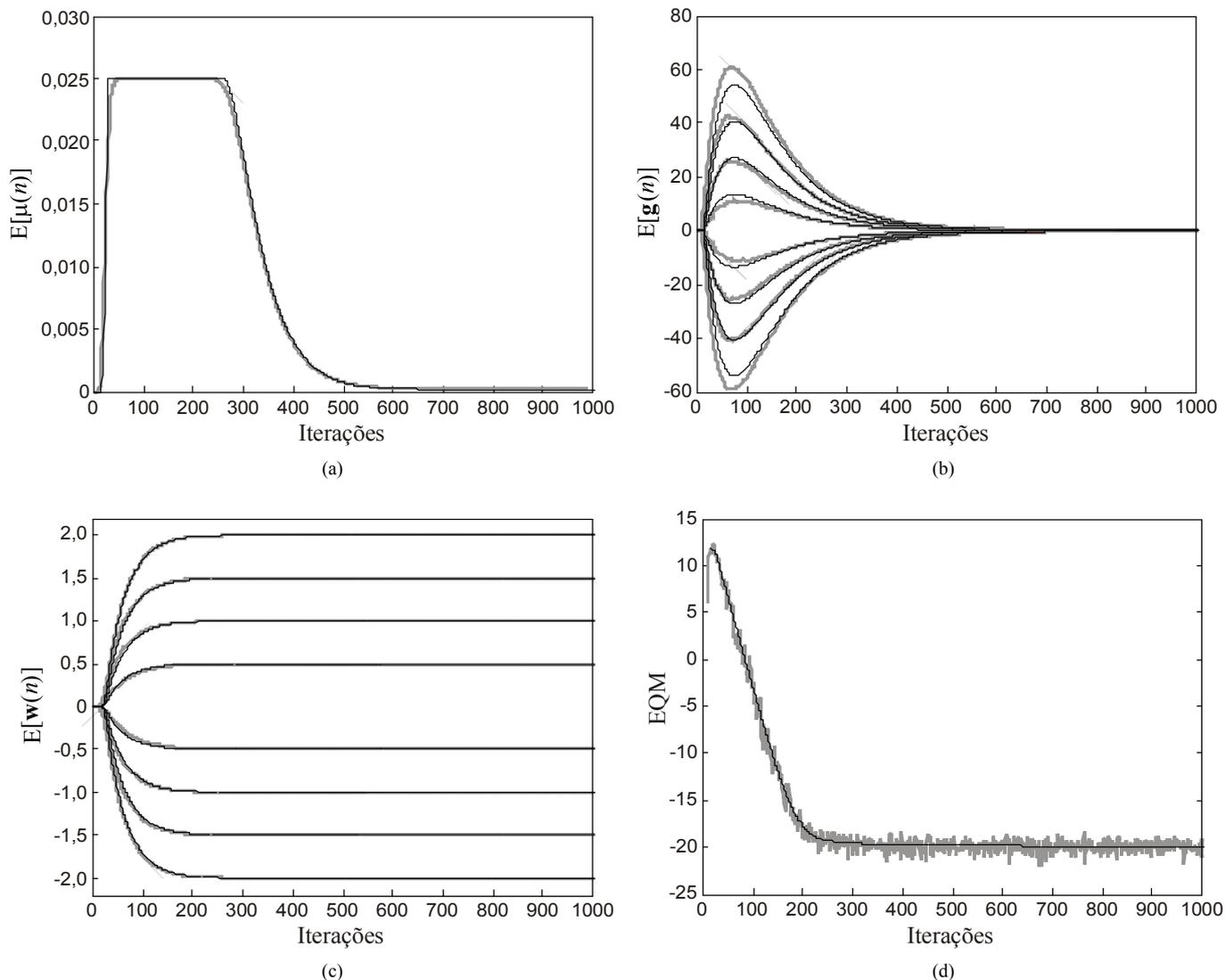


Fig. 2. Exemplo 1. Sinal de entrada branco. (a) Evolução do passo de adaptação variável: simulação (linha cinza), modelo (linha preta). (b) Vetor gradiente estimado: simulação (linhas cinzas), modelo (linhas pretas). (c) Comportamento médio do vetor de coeficientes: simulação (linhas cinzas), modelo (linhas pretas). (d) Curva do EQM: simulação (linha cinza irregular), modelo (linha preta).

VI. RESULTADOS DE SIMULAÇÃO

Nesta seção, são apresentadas simulações numéricas que comprovam a precisão do modelo aqui desenvolvido. Para tal, utiliza-se um problema de identificação de sistemas com a planta desconhecida dada pelo vetor $\mathbf{w}_o = [2, 0 \ -2, 0 \ 1, 5 \ -1, 5 \ 1, 0 \ -1, 0 \ 0, 5 \ -0, 5]^T$. Para o sinal de entrada, são utilizados tanto um ruído branco quanto um sinal colorido, ambos gaussianos com média zero e variância $\sigma_x^2 = 1$. O sinal colorido é obtido de um processo AR(1), dado por $x(n) = ax(n-1) + u(n)$ com $a = 0,6$ e $u(n)$ sendo um ruído branco gaussiano de média zero e variância $\sigma_u^2 = 0,65$. A dispersão dos autovalores da matriz \mathbf{R} do sinal de entrada colorido é $\chi = 13,2$. Além disso, um ruído aditivo branco gaussiano $\eta(n)$ com variância $\sigma_\eta^2 = 0,01$ é adicionado à

saída da planta. Os parâmetros do algoritmo utilizados nas simulações são ajustados conforme indicado na Tabela I.

A. Exemplo 1: Resultados do Modelo Proposto

Neste exemplo, os resultados de simulação Monte Carlo (média de 200 rodadas independentes) e os obtidos com modelo proposto são comparados considerando tanto um sinal de entrada branco quanto um sinal colorido. As Figs. 2 e 3 ilustram, respectivamente, esses resultados. As Figs. 2(a) e 3(a) mostram a evolução do passo de adaptação. As Figs. 2(b) e 3(b) apresentam as curvas do vetor gradiente estimado. Nas Figs. 2(c) e 3(c), o comportamento médio do vetor de coeficientes é mostrado e nas Figs. 2(d) e 3(d) são apresentadas as curvas de EQM. A partir dessas figuras, constata-se o bom casamento entre os resultados de simulação e aqueles obtidos através do modelo proposto.

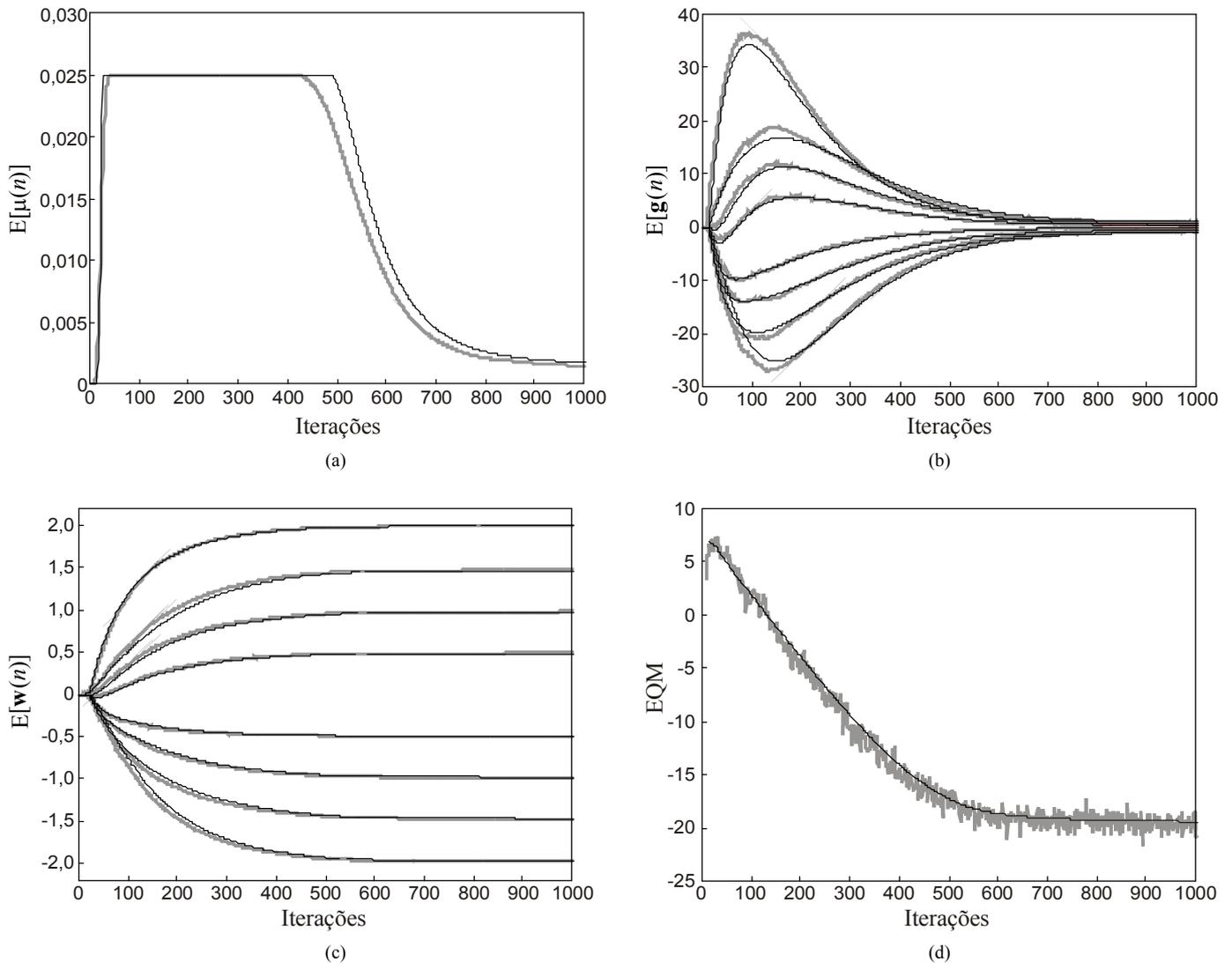


Fig. 3. Exemplo 2. Sinal de entrada colorido. (a) Evolução do passo de adaptação variável: simulação (linha cinza), modelo (linha preta). (b) Vetor gradiente estimado: simulação (linhas cinzas), modelo (linhas pretas). (c) Comportamento médio do vetor de coeficientes: simulação (linhas cinzas), modelo (linhas pretas). (d) Curva do EQM: simulação (linha cinza irregular), modelo (linha preta).

B. Exemplo 2: Validação da Suposição (ii)

Dado que a suposição (ii), considerada na Seção III, facilita sobremaneira os cálculos do modelo proposto, ela é então verificada através de simulação nesta seção. A Fig. 4 mostra a evolução de $E[\mu^2(n)]$ e $E^2[\mu(n)]$, obtida por simulação MC, para o caso em que é utilizado um sinal de entrada colorido. Observa-se um casamento muito bom entre as curvas, atestando que a suposição (ii) pode ser eficazmente adotada na derivação do modelo proposto.

VII. CONCLUSÕES

Neste trabalho, expressões analíticas para o momento de primeira ordem do vetor de coeficientes e para a curva de aprendizagem (EQM) são apresentadas. As expressões do modelo são obtidas considerando algumas hipóteses

simplificativas, as quais são devidamente comprovadas através de simulações numéricas. Observa-se um ótimo casamento entre as simulações MC e o modelo de predição desenvolvido tanto na fase transiente quanto em regime permanente para sinais de entrada branco e colorido.

TABELA I
PARÂMETROS UTILIZADOS NAS SIMULAÇÕES

| Entrada | Parâmetros |
|----------------|-----------------------------|
| Ruído Branco | $\alpha = 0,97$ |
| | $\beta = 0,99$ |
| | $\gamma = 1 \times 10^{-6}$ |
| Sinal Colorido | $\alpha = 0,97$ |
| | $\beta = 0,99$ |
| | $\gamma = 6 \times 10^{-6}$ |

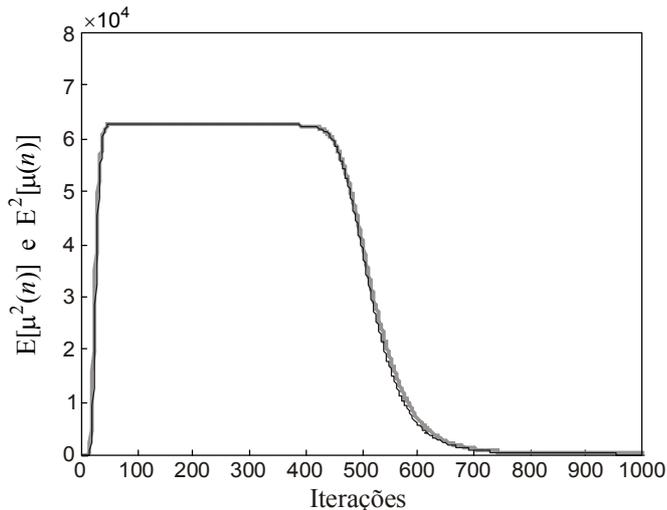


Fig. 4. Validação da suposição (ii): evolução de $E[\mu^2(n)]$ e $E^2[\mu(n)]$.

REFERÊNCIAS

- [1] B. Widrow and M. Hoff, "Adaptive switching circuits," in *Proc. IRE Western Electronic Show and Convention*, New York, USA, Part 4, Aug. 1960, pp. 96-104.
- [2] S. Haykin, *Adaptive Filter Theory*, 4th ed., Upper Saddle River, NJ: Prentice Hall, 2002.
- [3] V. J. Mathews and Z. Xie, "A stochastic gradient adaptive filter with gradient adaptive step size," *IEEE Trans. Signal Process.*, vol. 41, no. 6, pp. 2075-2087, June 1993.
- [4] J. Okello, Y. Itoh, Y. Fukui, I. Nakanishi, and M. Kobayashi, "A new modified variable step size for the LMS algorithm," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS)*, Monterey, USA, vol. 5, June 1998, pp. 170-173.
- [5] W-P. Ang and B. Farhang-Boroujeny, "A new class of gradient adaptive step-size LMS algorithms," *IEEE Trans. Signal Process.*, vol. 49, no. 4, pp. 805-810, Apr. 2001.
- [6] R. H. Kwong and E. W. Johnston, "A variable step size LMS algorithm," *IEEE Trans. Signal Process.*, vol. 40, no. 7, pp. 1633-1642, July 1992.
- [7] T. Aboulnasr and K. Mayyas, "A robust variable step-size LMS-type algorithm: Analysis and simulations," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 631-639, Mar. 1997.
- [8] D. W. Kim, J. H. Hoi, Y. S. Choi, C. H. Jeon, and H. Y. Ko, "A VS-LMS algorithm using normalized absolute estimation error," in *Proc. IEEE Digital Signal Processing Applications (TENCON)*, Perth, Australia, vol. 2, Nov. 2006, pp. 692-697.
- [9] Z. Ramadan and A. Poularikas, "A robust variable step-size LMS algorithm using error-data normalization," in *Proc. IEEE Southeastcon*, Huntsville, USA, Apr. 2005, pp. 219-224.
- [10] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Orlando, USA, May 2002, pp. 1881-1884.
- [11] Y. Wei and S. B. Gelfand, "Noise-constrained least mean squares algorithm," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1961-1970, Sept. 2001.