A Global Optimization Method for Continuous-Time Adaptive Recursive Filters

William Edmonson, *Member, IEEE,* Juan Carlos Palacios, Chang An Lai, and Haniph Latchman, *Member, IEEE*

***Abstract*—A major drawback of recursive adaptive filters based on gradient methods is that convergence to global minimum is not always achieved. This is due to a nonconvex mean square error (MSE) performance surface. This letter develops a continuous-time least mean square algorithm that converges to the global minimum with probability one.**

***Index Terms*— Adaptive recursive filters, global optimization, stochastic approximation.**

I. Introduction

Many adaptive filtering applications require a recursive structure. Continuous-time adaptive filters are necessary with applications that require high-frequency signal processing and/or very fast initial convergence rates. Analog realizations also result in a smaller integrated circuit, lower power dissipation, and freedom from clocking and aliasing effects. An example of this is channel equalization for cellular telephony.

Unfortunately, the practical implementation of recursive filters often involves the solution of a nonconvex optimization problem. This implies that for gradient-based methods, convergence to a local minimum can occur, leading to suboptimal results. Guaranteed convergence to the global minimum is dependent on the initial conditions being within a finite convex region of the global minimum.

In this letter, we develop a continuous-time recursive adaptive filter based on the global least mean square (GLMS) method of Edmonson *et al.* [1]. The main characteristic of the GLMS algorithm is that it converges to the global minimum with probability one. This is accomplished by "smoothing" the nonconvex objective function through a convolution of the objective function $\xi(t,\theta)$ with a noise probability density function (pdf); we define its associated random variable as η. The types of pdf that can be used are Gaussian, Cauchy, or uniform [3]. The variance of the pdf at the start of the optimization procedure is large, which has the effect of "smoothing" the objective function so that it is convex. Then the variance is slowly reduced to zero, whereby the smoothed functional returns to the original objective function as the algorithm converges to the global minimum. As shown in [3], where $\xi(t,\theta)$ represents the true objective function and $\hat{\xi}(t,\theta)$ represents the smoothed version of $\xi(t,\theta)$, then

$$\lim_{\beta\to 0}\hat{\xi}(t,\beta)=\xi(t,\theta) \tag{1}$$

where β controls the variance of the pdf and thus controls the amount of smoothing and the rate at which the error tends to zero. Furthermore, as $\beta\to 0$, $\theta\to\theta^*$ where θ^* is the global minimum of $\xi(t,\theta)$.

II. Continuous-Time Global LMS Algorithm

The analog least mean square algorithm for which we will begin our development is based on minimizing the gradient of the instantaneous square error

$$\frac{d\theta_i}{dt}=\frac{\partial e^2(t,\theta)}{\partial\theta_i} \tag{2}$$

$$=\nabla_{\theta_i}e^2(t,\theta) \tag{3}$$

where θ_i are the parameters of the adaptive filter and $e(\cdot)$ is the error between the output of the unknown system and the output of the adaptive filter. Integrating both sides of (2) yields the continuous time least mean square (LMS) algorithm

$$\theta(t)=\int_0^t\nabla_{\theta_i}\xi(\tau,\theta)\,d\tau \tag{4}$$

where $\xi(\cdot)=e^2(t,\theta)$. Note that when $\theta(t)$ becomes constant at a minimum, the value of the input to the integrator is zero.

From the digital derivation of the GLMS algorithm [1], it can be easily shown that the development of the analog version begins with the gradient of $\hat{\xi}(\cdot)$. The "smoothed" gradient estimate, as suggested in [3], can be approximated for the continuous-time gradient as

$$\nabla_\theta\hat{\xi}(t,\beta)=\nabla_\theta\xi(t,\beta\eta)\approx\nabla_\theta\xi(t,\theta-\beta\eta). \tag{5}$$

To obtain an implementable expression for the smoothed gradient estimate, we expand (5) about θ using a Taylor series expansion. For practical purposes, it will be assumed that the performance function is twice continuously differentiable and converges to its Taylor series in some interval (θ_i-h,θ_i+h) when it is centered at a particular point θ_i. The fact that the Taylor series expansion is only valid "locally" does not represent a problem since the gradient is an operator that characterizes the local features of the function.

Given that $\xi(\theta-\beta\eta)$ is an N variable function, the following procedure will be repeated for every independent variable, θ_i.

Manuscript received July 28, 1998. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. S. C. Douglas.

W. Edmonson, C. A. Lai, and H. Latchman are with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611-6130 USA.

J. C. Palacios is with Motorola, Inc., Plantation, FL 33322 USA (e-mail: edmonson@dsp.ufl.edu).

Publisher Item Identifier S 1070-9908(99)05797-1.

Taking θ_i as a fixed point and η as a random variable, the following series expansion about θ_i can be obtained:

$$\nabla_{\theta_i}\xi(t, \theta_i - \beta\eta) = \left[\xi'(t, \theta_i) + \beta\eta\xi''(t, \theta_i) + \frac{(\beta\eta)^2}{2!}\xi'''(t, \theta_i) + \cdots\right]. \quad (6)$$

From this equation, it can be seen that the second-order and higher-order derivatives are multiplied either by a random variable or by a power of this variable. This results in every second and higher order term being a random variable. Finally, applying the central limit theorem and using a vector representation, the following expression is obtained for the gradient

$$\nabla_\theta\xi(t, \theta - \beta\eta) \approx \nabla_\theta\xi(t, \theta) + \{\beta\eta\} \quad (7)$$

where the term $\{\beta\eta\}$ represents the second and higher-order derivative terms of (6). We will assume that η is a noise vector containing as each of its elements, a different random process realization. As described in [1], (6) represents the computation of the gradient at a random location in the neighborhood of the operating point. This implies that the purpose of this constant random number perturbation $\beta\eta$ is to avoid local minima during the early stages of adaptation rather than to enhance convergence. Therefore, the speed of convergence is not improved. The gradient approximation of (7) can also be described as computing the gradient of the MSE around θ with only the addition of a random number η. The variance of this random number will be decreased over time following an experimentally determined rate schedule, i.e., β [1], [4], thereby decreasing the chance of diverging away from the global minimum during the last phases of adaptation.

The continuous-time GLMS algorithm (CT-GLMS) can now be derived by replacing the gradient term of (4) with the convolutionally smoothed gradient of (7). This leads to the following parameter update equation:

$$\theta(t) = -\mu \int_0^t [\nabla_{\theta_i}\xi(\tau, \theta) + \beta\eta]\, d\tau. \quad (8)$$

From the above developments, we note as with the discrete version of the GLMS, that the smoothing of the objective function results in the noise term, $\beta\eta$, being added to the gradient, $\xi(\cdot)$, allowing the LMS algorithm to have the ability to jump out of local minima. The gradient term of (7) is the same as for the recursive LMS algorithm, i.e.,

$$\nabla_{\theta_i}\xi(t, \theta) = e(t)\,\frac{\partial y(t)}{\partial\theta_i}. \quad (9)$$

We will utilize the state space implementation of Johns *et al.* [2] to compute the gradient approximation of (9). State space filters represent the most general form of a linear system. The computation of $\partial y(t)/\partial\theta_i$ is derived from sensitivity analysis

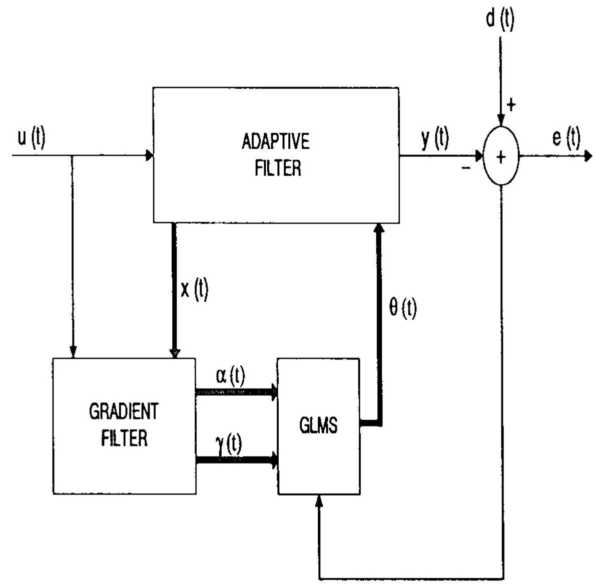

Fig. 1. CT-GLMS adaptive SS filter.

of the transfer function. The SS model is described as

$$s\mathbf{X}(s) = \mathbf{A}\mathbf{X}(s) + \mathbf{b}u(s)$$
$$Y(s) = \mathbf{c}^\mathbf{T}X(s) + du(s) \quad (10)$$

where θ in the above development corresponds to $\{A, b, c, d\}$. This leads to the partial derivative of (9) being computed as a convolution of the input and states with the intermediate transfer function of $\mathbf{G}(\mathbf{s}) = \mathbf{c}^\mathbf{T}(\mathbf{sI} - \mathbf{A})^{-1}$, i.e.,

$$\alpha_{ij} = g_i(t) \otimes x_i(t)$$
$$\gamma_i = g_i(t) \otimes u(t)$$

where $\otimes$ represents the convolution operation. This results in the following coefficient-update equations for determining the global minimum

$$A_{ij}(t) = 2\mu \int_0^t [e(\tau)\alpha_{ij}(\tau) + \beta\eta]\, d\tau \quad (11)$$

$$b_i(t) = 2\mu \int_0^t [e(\tau)\gamma_i(\tau) + \beta\eta]\, d\tau \quad (12)$$

$$c_i(t) = 2\mu \int_0^t [e(\tau)x_i(\tau) + \beta\eta]\, d\tau \quad (13)$$

$$d(t) = 2\mu \int_0^t [e(\tau)u(\tau) + \beta\eta]\, d\tau. \quad (14)$$

We define the above algorithm, (11)–(14), as the CT-GLMS. A block diagram of this algorithm is shown in Fig. 1.

III. Simulations

In this section, we present simulation results using a system identification formulation of the adaptive filter. We will identify the following unknown system:

$$H(s) = \frac{0.2s - 0.4}{0.1s^2 + 0.2s + 0.1} \quad (15)$$

TABLE I
RESULTS OF 100 MONTE CARLO RUNS

	a	b
True	-0.272	-1.23
Mean	-0.2609	-1.2160
Std	0.0992×10^{-3}	0.3978×10^{-3}

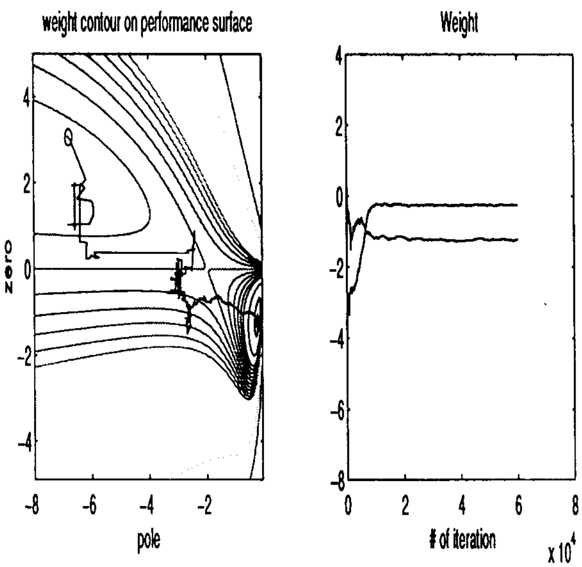

Fig. 2. Simulation result for the system $\hat{H}(s) = b/(s-a)$.

by the following underparameterized system:

$$\hat{H}(s) = \frac{b}{s-a}. \tag{16}$$

The input is chosen to be random Gaussian noise with a mean of zero and a variance of one. The constant step size $\mu = 0.001$. The system identification of (15) by that of (16) yields a nonconvex MSE performance surface where there exist a local minimum at $\{a, b\} = \{-8.78, 2.49\}$ and a global minimum at $\{a, b\} = \{-0.272, -1.2\}$.

The noise source used for $\eta(t)$ is a Cauchy pdf. The rate schedule equation can be described as

$$\beta = \frac{10^3}{t+2} \tag{17}$$

which is experimentally chosen based upon how $\beta \to 0$ and the initial value of the variance such that proper amount of "smoothing" occurs. For this simulation, 100 Monte Carlo runs were performed with different randomly chosen initial conditions located in the valley containing the local minimum, i.e., $a \in [-4, -8]$ and $b \in [0, 4]$. The results are given in Table I, and shows that convergence to the global minimum is achieved with small standard deviations. Convergence to the global minimum was achieved for each Monte Carlo run. An example of one of the runs is shown in Fig. 2 and clearly shows that the algorithm was able to escape from the local minimum.

IV. CONCLUSION

We have developed a continuous-time adaptive infinite impulse response (IIR) filter that converges to the global minimum. This proposed algorithm is implemented as a state space formulation for the coefficient-update equations. Research will be continuing on the CT-GLMS algorithm in the areas of convergence and error analysis.

REFERENCES

[1] W. Edmonson, J. Principe, K. Srinivasan, and C. Wang, "A global least mean square algorithm for adaptive IIR filtering," *IEEE Trans. Circuits Syst. II,* vol. 45, pp. 379–384, Mar. 1998.
[2] D. Johns, M. Snelgrove, and A. Sedra, "Continuous-time LMS adaptive recursive filters," *IEEE Trans. Circuits Syst.,* vol. 38, pp. 769–778, July 1991.
[3] M. Styblinski and T. Tang, "Experiments in nonconvex optimization: Stochastic approximation with function smoothing and simulated annealing," *Neural Networks,* vol. 3, pp. 467–483, 1990.
[4] C. Darken and J. Moody, "Note on learning rate schedules for stochastic optimization," in *Advances in Neural Information Processing Systems 3,* J. Moody, S Hanson, and R. Lipmann, Eds. San Mateo, CA: Morgan-Kaufmann, 1990.