

IMPROVED CONVERGENCE OF GRADIENT ALGORITHMS FOR ADAPTIVE IIR FILTERS

José A. Rodríguez Fonollosa and Enrique Masgrau

D. Teoria del Senyal i Comunicacions. UPC. Spain

ABSTRACT

The introduction of a simple correction term in gradient algorithms for adaptive IIR filtering is shown to improve their convergence and robustness. The error surface, i.e., the mean squared value of the output error as a function of adaptive coefficients, is quadratic in adaptive FIR filtering and thus a simple gradient search works well. This is not the case in adaptive IIR filtering where the error surface is far from being quadratic. Nevertheless, a simple modification of the updating equation of lattice-form algorithms can reduce this problem thus obtaining faster adaptive IIR filters.

INTRODUCTION

To date, the majority of adaptive filtering problems have been solved with FIR filters because they are well behaved, i.e., the unconditionally stable and their error surface is quadratic. However, IIR filters present some advantages over FIR filters. IIR filters usually require, for the same performance, much less computation than FIR filters and can be a good alternative in applications which require an extremely large number of FIR taps as echo cancellation. Adaptive IIR filtering has also been shown to improve the performance of predictors in differential pulse code modulation (DPCM) systems.

Despite such practical utility, adaptive IIR filter algorithms are a relatively recent development and several problems still remain unsolved. Unlike adaptive FIR filtering, the error surfaces for adaptive IIR filters may have some local minima. Their convergence properties are not well understood, and the

poles may move outside the unit circle during adaptation. Analysis of this problem is difficult because the filter output is a nonlinear function of the coefficients.

Most of adaptive filters use a steepest descent method to adapt the weights of the filters in the form

$$W_i(n+1) = W_i(n) - \mu \left[\frac{\partial J}{\partial W_i(n)} \right] \quad (1)$$

where J is the cost function to minimize.

In adaptive FIR filtering the excess squared error ($J - J_{\min}$) is a quadratic function of the difference between the actual weight vector W and the optimal one W_0 , and gradient descent techniques performs well because there is a linear relation between this weight vector error and the gradient.

$$J - J_{\min} = (W - W_0)^T R (W - W_0) \quad (2a)$$

$$J'w = -2R (W - W_0) \quad (2b)$$

$$W_0 = W - \frac{1}{2} R^{-1} [J'w] \quad (2c)$$

where R is the autocorrelation matrix and $J'w$ the gradient of J with respect to the elements of W .

However, in adaptive IIR filtering the error surfaces are very different and this fact causes problems in the convergence when this kind of methods are used to adapt the coefficients of the filter. In our work, the gradient descent procedure has been modified

with a correction term that takes into account the form of the error surface to improve the convergence and stability of adaptive IIR filters. Some attempts have been made to study the behaviour of error surfaces for the "System Identification Mode" of IIR filtering [1] but, in general, the knowledge about these error surfaces is quite limited. However, we know that the cost function J tends to infinite when the backward coefficients of the adaptive filter are close to the limits of the stability region. In these limits, the gradient also tends to infinite and this fact can cause a great oscillation in (1). To reduce this oscillation, a correction term that results from the study of a simple System Identification case has been added to (1). As the limits of the stability region are easy to express when we work with the IIR lattice structure, this correction term can be easily extended to the IIR lattice algorithms of any order.

The rest of paper is organized as follows: first, the error surface of a simple system identification problem is studied, the results are discussed and a adaptation algorithm is derived. Then, some possible generalizations of the algorithm for higher order adaptive systems are proposed. The last part of the paper is dedicated to show the results of the simulations and the conclusions.

ERROR SURFACE

In the identification of a system with one pole we have the following signals:

$$u(n) = V x(n) - K u(n-1) \quad (3a)$$

$$d(n) = u(n) + w(n) \quad (3b)$$

$$\hat{u}(n) = \nu x(n) - k \hat{u}(n-1) \quad (3c)$$

$$e(n) = d(n) - \hat{u}(n) \quad (3d)$$

where $x(n)$ is the input, $u(n)$ the output of the system to be identified, $w(n)$ the disturbance, $\hat{u}(n)$ the output of the adaptive filter and $e(n)$ the error. The cost function to minimize is defined as

$$J = E\{e^2(n)\} \\ = E\{d^2(n)\} + E\{\hat{u}^2(n)\} - 2 E\{d(n)\hat{u}(n)\} \quad (4)$$

If the noise $w(n)$ is uncorrelated with $x(n)$ and if $x(n)$ is a white sequence, then the expression for J as a function of ν and k , (error surface), can be easily derived.

$$\sigma^2 = E\{x^2(n)\} \quad (5a)$$

$$n = E\{w^2(n)\} \quad (5b)$$

$$J = E\{u^2(n)\} + E\{\hat{u}^2(n)\} - 2E\{u(n)\hat{u}(n)\} + n \quad (6)$$

$$J = \sigma^2 \left(\frac{V^2}{1-K^2} + \frac{\nu^2}{1-k^2} - \frac{2\nu V}{1-kK} \right) + n \quad (7)$$

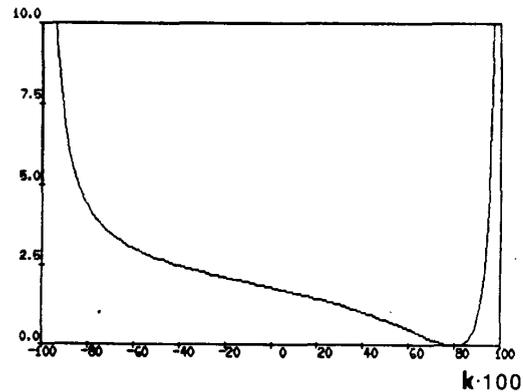


Figure 1. "Error Surface" J as a function of k , for $K=0.8$ and $V=\nu=1$.

Then we can take the gradient of J with respect to ν and k .

$$\frac{\partial J}{\partial \nu} = \sigma^2 \left(\frac{2\nu}{1-k^2} - \frac{2V}{1-kK} \right) \quad (8)$$

$$\frac{\partial J}{\partial k} = \sigma^2 \left(\frac{2\nu^2 k}{(1-k^2)^2} - \frac{2\nu V K}{(1-kK)^2} \right) \quad (9)$$

As it was pointed out in the introduction, we can observe that both the cost function J and its gradient tend to infinite when k tends to 1.

We are now interested in study the adaptation of the feedback coefficient k . To obtain a useful result we can suppose that the forward coefficient (FIR part of the filter) has converged

$$\frac{\partial J}{\partial u} = 0 \quad (10)$$

$$\frac{u}{1-k^2} = \frac{V}{1-kK} \quad (11)$$

and then the derivative of J with respect to the feedback coefficient k can be expressed as

$$\begin{aligned} \frac{\partial J}{\partial k} &= \frac{2\sigma^2 u^2}{1-k^2} \left(\frac{k}{1-k^2} - \frac{K}{1-kK} \right) \\ &= \frac{2\sigma^2 u^2}{1-k^2} (k-K) \frac{1}{(1-k^2)(1-kK)} \end{aligned} \quad (12)$$

Now, we can rewrite the above equation to obtain

$$\Delta_k = -\frac{1}{2\sigma^2 u^2} \left[\frac{\partial J}{\partial k} \right] (1-k^2)^2 (1-kK) \quad (13)$$

where $\Delta_k = K-k$. To simplify the notation we will define H as

$$H = -\frac{1}{2\sigma^2 u^2} \left[\frac{\partial J}{\partial k} \right] (1-k^2)^3 \quad (14)$$

and then Δ_k can be expressed as

$$\begin{aligned} \Delta_k &= H \frac{1-kK}{1-k^2} \\ &= H - H \frac{k}{1-k^2} \Delta_k \end{aligned} \quad (15)$$

$$\Delta_k = H \frac{1}{1+H \frac{k}{1-k^2}} \quad (16)$$

Observe that, if the mentioned conditions were true and we had the exact gradient we will obtain a one-step convergence. As we will have only an estimation, the proposed updating equation will be

$$k(n+1) = k(n) - \mu \Delta_k \quad (17)$$

where Δ_k is calculated from (16).

With the approximation $\Delta_k = H$ we can simplify the above equation to

$$k(n+1) = k(n) - \mu_n \left\{ \frac{\partial J}{\partial k(n)} \right\} (1-k(n)^2)^2 \quad (18)$$

where μ_n is a normalized step-size

$$\mu_n = \frac{\mu}{2E\{\hat{u}^2(n)\}} = \frac{\mu}{2\frac{\sigma^2 u^2}{1-k^2}} \quad (19)$$

It is clear from the observation of (12) and (18) that the effective step-size is reduced as the backward coefficient approaches the limits of the stability region. This reduction compensates the increase in the derivative and prevents the filter for becoming unstable.

IIR LATTICE ALGORITHMS

The study of the error surface and its gradient for systems with 2 or more poles is complicated. However, the IIR lattice structure allows an easy extension of the above results. The lattice realization is an important alternative to the direct form in adaptive IIR filtering because the stability can be controlled with essentially no computation [2] with the following conditions:

$$|k_m(n)| < 1, \quad 1 \leq m \leq N \quad (20)$$

The literature about IIR lattice algorithms is not very extensive. In [3] the expression for calculate the gradient is derived and a updating equation of the form (1) is applied to both the forward and backward coefficients. In a recent paper [4] some adaptive algorithms based on the Gauss-Newton method can also be found. Our goal is to use the derived results for the simple case to improve these algorithms. Thanks to the condition (20), the updating equation derived for the case of one pole can be extended to the general case with a clear meaning, i.e., the reduction of the step-size as the filter approaches the unstable region. Therefore we propose the updating equation

$$k_i(n+1) = k_i(n) - \mu_n \left\{ \frac{\partial J}{\partial k_i(n)} \right\} (1-k_i(n)^2)^\gamma \quad (21)$$

The choice of exponent γ will depend on the normalization of μ_n and on the specific

problem. An exponent greater than the given by (14) will reduce the misadjustment and increase the stability. The price will be reduction of the tracking capacity of the filter.

COMPUTER SIMULATIONS

We have compared the convergence properties of the IIR lattice algorithm [3] using the originally proposed updating equation (1) and with the modification (21) with $\gamma=2$. We have also compared this modified gradient procedure with the well-known SHARF algorithm. We consider a system identification problem where the unknown system is the one used in [5] to test the SHARF subroutine. Its lattice coefficients V_i and k_i are as follows [2]:

$$\begin{aligned} V_0 &= 0.56, & V_1 &= 0.30, & V_2 &= -0.50 \\ K_1 &= -0.875, & K_2 &= 0.60 \end{aligned}$$

The comparison is not easy because the convergence depends on the step-size. We decided to try different step-sizes in each algorithms and to use the one that give the fastest convergence without instability. For the algorithm [3] the step-size was $\mu=0.01$, for the modified algorithm $\mu=0.12$ and for the SHARF algorithm the same used in [5]. The

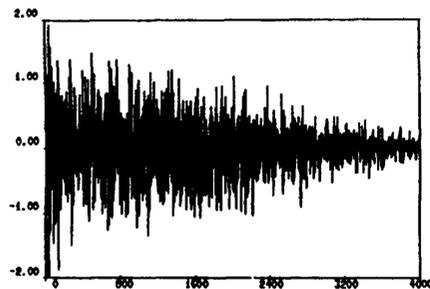


Figure 2. Output error with the Parik algorithm [2].

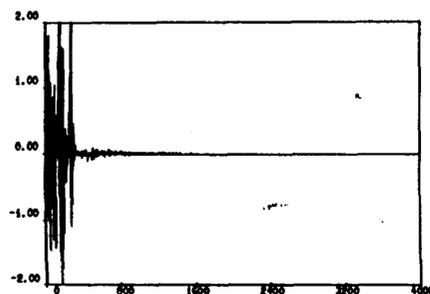


Figure 3. Output error with the modified Parik algorithm.

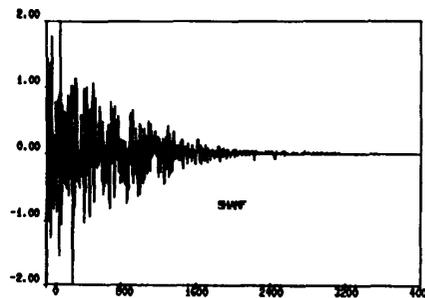


Figure 4. Output error with the SHARF algorithm.

results can be shown in the figures 2, 3, and 4 and they are in concordance with the predicted increase in stability and robustness.

CONCLUSIONS

It has been shown that the study of the error surfaces and their gradients is important to understand the convergence and misadjustment of IIR adaptive algorithms. A modification of the usual updating equation has been proposed to reduce the step-size as the coefficients approach the limits of the stable region. The result is a great improvement in both convergence and robustness against different step-sizes.

REFERENCES

- [1] S. D. Stearns, "Error surfaces of Recursive Adaptive Filters". IEEE Trans. on ASSP, Vol-29, No.3, June 1981.
- [2] A. H. Gray, Jr., and J. D. Markel, "Digital lattice and ladder filter synthesis", IEEE Trans. Audio Electroacoust., Vol AU-21, no. 6, pp. 491-500, Dec. 1973.
- [3] D. Parikh, N. Ahmed and S. D. Stearns, "An adaptive lattice algorithm for recursive filters". IEEE Trans. on ASSP. Vol ASSP-28, No. 1, pp 110-111. Feb.1980.
- [4] J. J. Shynk, "On lattice-form algorithms for adaptive IIR filtering". Proc. ICASSP 1988, pp 1554-1557.
- [5] J. R. Treichler, C. R. Johnson and M. G. Larimore. *Theory and design of Adaptive Filters*. Wiley-Interscience 1987.