REGRESSOR BASED ADAPTIVE INFINITE IMPULSE RESPONSE FILTERING

Emrah Acar

Dept. of Electrical and Computer Engineering
Carnegie Mellon University
Pittsburgh, PA 15213
emrah@ece.cmu.edu

Orhan Arıkan

Dept. of Electrical Engineering
Bilkent University
Ankara, Turkey 06533
oarikan@ee.bilkent.edu.tr

ABSTRACT

To take advantage of fast converging multi–channel recursive least squares algorithms, we propose an adaptive IIR system structure consisting of two parts: a two–channel FIR adaptive filter whose parameters are updated by rotation–based multi–channel least squares lattice (QR–MLSL) algorithm, and an adaptive regressor which provides more reliable estimates to the original system output based on previous values of the adaptive system output and noisy observation of the original system output. Two different regressors are investigated and robust ways of adaptation of the regressor parameters are proposed. Based on extensive set of simulations, it is shown that the proposed algorithms converge faster to more reliable parameter estimates than LMS type algorithms.

1. THE REGRESSOR BASED IIR ADAPTIVE FILTER STRUCTURE

As shown in Fig. 1, in a typical adaptive filtering application, input, $x(n)$, and noisy output, $d(n)$, of an unknown system are available for processing by an adaptive system to provide estimates, $y(n)$, to the output of the unknown system as time progresses. In our investigation, the unknown system or plant has an IIR model whose output can be compactly expressed as a function of its previous values and its input as:

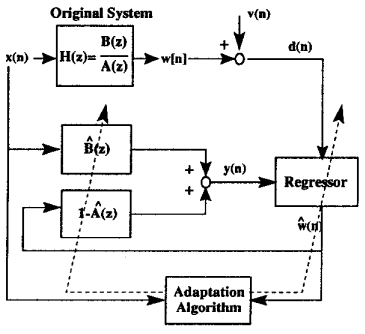

Figure 1: *Common framework of IIR–γ and IIR–Kalman*

$$w(n) = \sum_{j=1}^{N} a_j w(n-j) + \sum_{i=0}^{M} b_i x(n-i) = \underline{\theta}^T \underline{\phi}(n), \quad (1)$$

where $\underline{\theta}$ is the vector of direct form system parameters:

$$\underline{\theta} = \left[a_1 \cdots a_N \quad b_0 \cdots b_M \right]^T = \left[\mathbf{a}^T \quad \mathbf{b}^T \right]^T, \quad (2)$$

and $\underline{\phi}(n)$ is formed by the previous values of the output and the present and past values of the input:

$$\begin{aligned} \underline{\phi}(n) &= \left[w(n-1) \cdots w(n-N) \quad x(n) \cdots x(n-M) \right]^T \\ &= \left[\mathbf{w}(n)^T \quad \mathbf{x}(n)^T \right]^T. \end{aligned} \quad (3)$$

Unlike the FIR adaptation, in adaptive IIR filtering we are faced with the problem of deciding on the feedback signal used in the adaptation when we have noisy observations of the actual system output. Hence, as shown explicitly in Fig. 1, we use a regressor in the proposed structure which causally estimates the feedback signal based on the noisy output $d(n)$ and the output of the adaptive filter $y(n)$, which is obtained as:

$$y(n) = \hat{\underline{\theta}}^T(n) \hat{\underline{\phi}}(n) \quad (4)$$

where $\hat{\underline{\theta}}(n)$ is the vector of estimated system parameters ($\left[\hat{\mathbf{a}}(n)^T \quad \hat{\mathbf{b}}(n)^T \right]^T$) and $\hat{\underline{\phi}}(n)$ is the vector of the regressor output, $\hat{w}(n)$, and the system input, $x(n)$:

$$\hat{\underline{\phi}}(n) = \left[\hat{\mathbf{w}}(n)^T \quad \mathbf{x}(n)^T \right]^T. \quad (5)$$

The performance of the adaptive filter heavily depends on how well the regressor, $\hat{w}(n)$, provides estimates to the actual system output $w(n)$. The two well known formulations of adaptive IIR filtering, namely the output error (OE) and the equation error (EE) formulations, correspond to two different types of regressors. In the OE formulation, the signal vector $\hat{\underline{\phi}}_O(n)$ is described as: $\hat{\underline{\phi}}_O(n) = \left[\mathbf{y}(n)^T \quad \mathbf{x}(n)^T \right]^T$ which corresponds to a regressor whose output is the output of the adaptive filter. In the EE formulation the signal vector, $\hat{\underline{\phi}}_E(n)$ is given as: $\hat{\underline{\phi}}_E(n) = \left[\mathbf{d}(n)^T \quad \mathbf{x}(n)^T \right]^T$ which corresponds to a regressor whose

0-7803-4428-6/98 $10.00 © 1998 IEEE

output is the noisy observation of the system output, $d(n) = w(n) + v(n)$.

Since the least squares cost function of EE formulation is quadratic in terms of the parameter vector θ, fast converging recursive least squares techniques can be used in the adaptation. However, because of the additive measurement noise, $v(n)$, the converged parameter values are biased estimates of the actual system parameters [1]. In the OE formulation, the least squares cost function is not a quadratic function of the parameters. Hence, we are bound to use LMS type gradient descent techniques in the adaptation. When these LMS type adaptation algorithms converge to the global minima of the cost function, the obtained parameters are unbiased estimates of the cost function. Unfortunately, LMS type gradient adaptation methods not only converge slowly, but also may converge to a local minima of the cost surface.

In this work, we try to combine the desired features of both OE and EE formalism in one formulation where the cost function is kept as a quadratic function of the parameters in order to use fast RLS techniques. As suggested in Fig. 1, this is achieved by choosing the adaptive filter as a two–channel FIR filter with inputs $x(n)$ and $\hat{w}(n-1)$. Then, the corresponding weighted least squares cost function becomes:

$$J(\underline{\theta}, n) = \sum_{k=1}^{n} (d(k) - \underline{\theta}^T \underline{\hat{\phi}}(k))^2 \lambda^{n-k}, \qquad (6)$$

which is a quadratic function of θ, because $\underline{\hat{\phi}}(n)$ is a fixed sequence of vectors determined by the past parameter estimates $\underline{\hat{\theta}}(n-1), \underline{\hat{\theta}}(n-2), \cdots, \underline{\hat{\theta}}(0)$. Hence, efficient multi–channel FIR recursive least squares techniques can be used to obtain parameter estimates at time n, $\underline{\hat{\theta}}(n)$, as the minimizer of $J(\underline{\theta}, n)$.

In the following section, two different types of regressors will be investigated in detail and corresponding recursive least squares adaptation algorithms will be presented.

2. PROPOSED REGRESSORS

2.1. IIR–γ

The IIR–γ regressor estimates the actual system output as a convex combination of the noisy observations, $d(n)$ and the adaptive filter output, $y(n)$:

$$\hat{w}(n) = \gamma_n d(n) + (1-\gamma_n) y(n) \quad , \quad 0 \le \gamma_n \le 1 \qquad (7)$$

where γ_n is the regression coefficient. The proper choice of γ_n should be based on a measure of the reliability of the estimated system parameters. A significant deviation of $y(n)$ from $d(n)$ is an indication that the system parameters are not reliably estimated, and hence, γ_n should be set close to 1, so that equation error type adaptation should take place. On the contrary, if $y(n)$ closely follows $d(n)$, then to reflect our level of confidence to the estimated system parameters, γ_n should be set close to 0, so that output error type adaptation should be performed. We propose to base the measure of reliability of the estimated system parameters on the statistical significance of the observed deviation between $y(n)$ and $d(n)$ sequences. For this purpose, one way of choosing γ_n is based on weighted estimate of the expected energy of the error sequence $e(n) = d(n) - y(n)$:

$$L(n) = \frac{\sum_{i=0}^{n} \lambda_v^i e(n-i)^2}{\sum_{i=0}^{n} \lambda_v^i}, \qquad (8)$$

where λ_v is an exponential forgetting factor that can improve the performance of the estimator. In our investigation, we observed that the critical properties of the functional form between $L(n)$ and γ_n are the boundary values l_1 and l_2 such that $\gamma_n = 0$ if $L(n) < l_1$ and $\gamma_n = 1$ if $L(n) \ge l_2$. In order to determine which values for l_1 and l_2 should be used, we investigated the expected values of the $L(n)$ for the cases of $\gamma_n = 0$ and $\gamma_n = 1$, which correspond to output and equation error adaptation cases, respectively. Assuming that $\gamma_n = 0$ and the estimated parameters have converged to the actual ones, the observed error sequence, $e(n)$, will be equal to $v(n)$, the additive Gaussian observation noise. Hence, $\mathbf{E}\{L(n)\}$ will be σ_v^2, the variance of $v(n)$. Therefore, l_1 is chosen as σ_v^2. Likewise, when $\gamma_n = 1$, $\mathbf{E}\{L(n)\}$ is equal to the variance of $e(n)$ sequence for the EE formulation. Since the equation error, $e_E(n)$ is related to the output error, $e_O(n)$ as in [2]: $e_E(n) = e_O(n) - \hat{\mathbf{a}}^T(n)\mathbf{e}_O(n)$, when $v(n)$ is white noise, the variance of $e_E(n)$ can be written as: $\sigma_v^2 \left(1 + \sum_{i=1}^{N} \hat{a}_i^2(n)\right)$, at the time of convergence to true parameters. Hence, we propose to use:

$$l_1 = \sigma_v^2 \quad , \quad l_2 = U\sigma_v^2 (1 + \sum_{i=1}^{N} \hat{a}_i^2) \qquad (9)$$

where $U > 1$ is introduced to avoid the convergence point of the equation error adaptation. For computational efficiency, the functional relation between $L(n)$ and γ_n is chosen as:

$$\gamma_n = \begin{cases} 0 & L(n) < l_1 \\ \kappa \frac{(L(n)-l_1)^p}{(\frac{l_2-l_1}{2})^p} & l_1 \le L(n) < \frac{l_1+l_2}{2} \\ 1 - (1-\kappa)\frac{(l_2-L(n))^p}{(\frac{l_2-l_1}{2})^p} & \frac{l_1+l_2}{2} \le L(n) < l_2 \\ 1 & L(n) \ge l_2 \end{cases} \qquad (10)$$

where κ and p are two parameters providing some control of the actual shape of the curve in between two boundaries l_1 and l_2. Fortunately, we observed that the behavior of the algorithm is not so sensitive to these shape parameters. For each iteration, this regression algorithm requires $(N+11)$ multiplications which is $O(N)$.

2.2. IIR–Kalman

Output of the IIR–Kalman regressor is an estimate of the actual system output obtained by using the Kalman filter on the following state–space model of the original system [3]:

$$\mathbf{w}(n+1) = \mathcal{A}\mathbf{w}(n) + \mathcal{B}\mathbf{x}(n) \qquad (11)$$

$$d(n) = \mathcal{C}\mathbf{w}(n+1) + v(n) \qquad (12)$$

where $\mathcal{C} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$. Since the actual parameters are unknown in Eqn. (11), we can form $\hat{\mathcal{A}}$ and $\hat{\mathcal{B}}$ matrices by using the estimated parameters at time n. Then, we get

$$\mathbf{w}(n+1) = \hat{\mathcal{A}}(n)\mathbf{w}(n) + \hat{\mathcal{B}}(n)\mathbf{x}(n) + \mathbf{u}(n) \qquad (13)$$

$$
\begin{aligned}
\hat{\mathbf{w}}_K(n+1|n) &= \left[\; y(n) \quad \hat{w}_K(n-1) \quad \cdots \right. \\
&\qquad \left. \cdots \quad \hat{w}_K(n-N+1) \;\right]^T \\
\mathcal{P}(n+1|n) &= \hat{\mathcal{A}}(n)\mathcal{P}(n|n)\hat{\mathcal{A}}(n)^T + \mathcal{R}_u(n) \\
\mathcal{G}(n) &= \left[\; \frac{\mathcal{P}(n+1|n)}{\mathcal{P}(n+1|n)_{(1,1)}+\sigma_v^2} \quad 0 \quad \cdots \quad 0 \;\right]^T \\
\mathcal{P}(n+1|n+1) &= \left(\mathbf{I} - \mathcal{G}(n)\mathcal{C}\right)\mathcal{P}(n+1|n) \\
\hat{\mathbf{w}}_K(n+1) &= \hat{\mathbf{w}}_K(n+1|n) + \\
&\qquad \mathcal{G}(n)\left(d(n) - \hat{\mathbf{w}}_K(n+1|n)_{(1,1)}\right) \\
\hat{w}(n) &= \hat{\mathbf{w}}_K(n+1)_{(1,1)}
\end{aligned}
$$

Table 1: *Equations of IIR–Kalman State Estimator*

where $\mathbf{u}(n)$ is introduced as an additional noise term to the system dynamics to account for the approximations in $\mathcal{A}$ and $\mathcal{B}$ by $\hat{\mathcal{A}}(n)$ and $\hat{\mathcal{B}}(n)$, which are equal to:

$$
\hat{\mathcal{A}}(n) = \begin{bmatrix} -\hat{\mathbf{a}}(n)^T \\ \mathbf{I_{N-1}} \quad \mathbf{0} \end{bmatrix}, \quad \hat{\mathcal{B}}(n) = \begin{bmatrix} \hat{\mathbf{b}}(n)^T \\ \mathbf{0_{(N-1\times N)}} \end{bmatrix}. \tag{14}
$$

Since the approximations in $\mathcal{A}(n)$ and $\mathcal{B}(n)$ are only limited to the first row, the additional process noise $\mathbf{u}(n)$ is:

$$
\mathbf{u}(n) = \left[\; u(n) \quad 0 \quad \cdots \quad 0 \;\right]^T. \tag{15}
$$

Application of the Kalman filter on the approximated model requires the covariance matrices $\mathcal{R}_u(n)$ and $\mathcal{R}_v(n)$ of $u(n)$ and $v(n)$, as well as an initial estimate to the state vector $\mathbf{w}(0)$ and the variance of the initial system error $\mathcal{R}_u(0)$. The $\mathcal{R}_u(n)$ can be determined by the sample variance of $u(n)$ for which a robust way is presented in [4]. The steps of the corresponding Kalman estimator are given in Table 1, where $\hat{\mathcal{A}}(n), \hat{\mathcal{B}}(n), \hat{\mathbf{w}}(n)$ are defined in Eqns. (14), (5) and the notation of $\mathcal{T}_{(1,1)}$ is used to denote the first diagonal entry of the matrix $\mathcal{T}$. Note that the output of the regressor $\hat{w}(n)$ is the first entry in the estimated state vector $\hat{\mathbf{w}}_K(n+1)$ and also the a–priori state estimate $\hat{\mathbf{w}}_K(n+1|n)$ is obtained efficiently by using the output of the adaptive filter and the previous states of the Kalman filter. For each iteration, the Kalman regressor requires $(3N^2+2N)$ multiplications, hence it is $O(N^2)$.

The required two–channel FIR adaptation can be efficiently performed by using QR–MLSL algorithm which is a rotation–based multi–channel least squares lattice algorithm with many desired features [5]. For each update, this algorithm requires $O(4N)$ multiplications. The required direct form parameters for the Kalman regressor can be computed by using standard mapping rules between lattice and direct form parameters [5].

3. SIMULATION EXPERIMENTS

In the following simulations, the adaptive filters are "all–zero" initialized and reported system identification results are the ensemble average of 50 realizations. The proposed algorithms are compared with two descent type IIR adaptation algorithms CRA [6] and BRLE [2], as well as with the extended Kalman filter (EKF) algorithm (which is an $O(N^2)$ algorithm presented in [4]).

3.1. Simulation Example 1

The system to be identified is chosen as in [2]:

$$
H(z) = \frac{1}{1 - 1.7z^{-1} + 0.7225z^{-2}}. \tag{16}
$$

The input is a unit–variance white Gaussian process. The output noise, $v(n)$, is chosen as white Gaussian. σ_v is varied to investigate the sensitivity of the performance of the algorithms to the level of SNR.

In Fig. 2, the squared norms of the parameter error vectors, $\mathbf{e}_\theta(n) = \underline{\theta} - \underline{\hat{\theta}}(n)$ are plotted. σ_v is set to 0.5. The forgetting factor, λ of the QR–MLSL algorithm is chosen as 0.999, and the parameters of the regressor subsystem of Eqns. (9) and (10) are chosen as $\lambda_v = 0.9, p = 1, \kappa = 0.7, U = 2$. For the IIR–Kalman regression algorithm, the initial variance estimate, $\tilde{\sigma}_u^2(0)$ is chosen as unity. In order to better resolve the early convergence behaviors of the compared algorithms, a logarithmic time axis is used in Fig. 2. As seen in this figure, the proposed algorithms have converged to an error level of -10 dB earlier than the 1000^{th} sample, but the CRA and BRLE algorithms converge to the same error level at about 40000^{th} sample. The EKF algorithm, performing the best, converges to -20 dB at around 50000^{th} sample. Here, the same step–size of 0.0005 is used for the CRA and BRLE algorithms. As recommended in [2] and [6], the composition parameter γ for CRA is chosen as 0.9, and the remedier parameter of BRLE, $\tau(n)$ is chosen as $\min(\|\underline{\hat{\phi}}(n)\|/\|\mathbf{e}_O(n)\|, 1)$.

In this example, conventional equation and output error (EE and OE) adaptation converged to error levels of -7 dB and 5 dB respectively, which are significantly higher than those of compared algorithms here. Therefore, as initially expected, the performance of the regressor based RLS approaches can be better than both the EE and OE formulations.

We repeated this experiment at different noise levels and reported the obtained $\|\mathbf{e}_\theta(n)\|^2$ results in Table 2. In this experiment, the best performing algorithm is found as the EKF algorithm. However, EKF requires an order more multiplications than IIR$-\gamma$ algorithm. As seen from these results, at high SNR (low levels of σ_v), LMS type algorithms converge to lower error levels. However, as the SNR decreases (high values of σ_v) the proposed algorithms start providing closer or better results than LMS type algorithms, which is an important advantage in many practical applications. Note that, the tabulated results correspond to the error levels at the 5000^{th} sample for the proposed algorithms and 50000^{th} samples for the EKF, CRA and BRLE algorithms. Since, in many important applications, the speed of convergence is critical, the proposed algorithms provide a good trade–off between error levels and the speed of convergence even at high SNR. Also, IIR$-\gamma$ provides comparable results to IIR–Kalman although it requires an order less number of multiplications.

σ_v	IIR–γ	IIR–Kalman	EKF	CRA	BRLE
0.05	-74.57	-82.25	-74.55	-84.97	-77.38
0.10	-50.92	-56.29	-69.49	-62.50	-55.72
0.25	-24.19	-27.18	-38.88	-32.16	-28.38
0.50	-10.27	-10.90	-21.27	-11.31	-10.23
1.00	-1.40	0.40	-5.62	3.61	3.53

Table 2: *Squared norm of parameter error in dB at convergence for different noise levels (Example 1)*

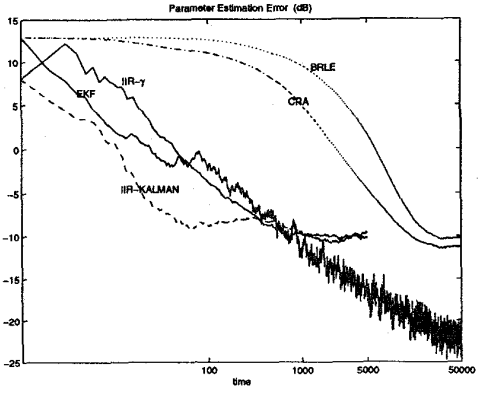

Figure 2: *Squared norm of parameter error (Example 1)*

3.2. Simulation Example 2

In this example, an abruptly changing system is selected with the time–varying transfer function:

$$H(z,n) = \begin{cases} \frac{0.2759+0.5121z^{-1}+0.5121z^{-2}+0.2759z^{-3}}{1-0.001z^{-1}+0.6546z^{-2}-0.0775z^{-3}} & n < 500 \\ \frac{0.7241+0.4879z^{-1}+0.4879z^{-2}+0.7241z^{-3}}{1+0.001z^{-1}+0.6546z^{-2}+0.0775z^{-3}} & n \geq 500 \end{cases} \quad (17)$$

The input sequence is chosen similar to the previous example. The output noise $v(n)$ is chosen as a zero–mean white Gaussian noise with a variance of 0.25. The step–size of CRA and BRLE algorithms is set to 0.01. The composition parameter, γ of CRA is set to 0.5, and the remedier parameter, $\tau(n)$ of BRLE is determined as in the first example. The forgetting factor of the proposed algorithms is set to 0.99 for a better tracking of the variations in the system parameters. For IIR–Kalman algorithm, initial variances are chosen as unity. The parameters of IIR–γ in Eqns. (9) and (10) are chosen as $\lambda_v = 0.95, p = 1, \kappa = 0.3, U = 5$. EKF is also initialized with all–unit variances. The squared norm of parameter errors $\|\mathbf{e}_\theta(n)\|^2$ is shown in Fig. 3. As seen from these results, both CRA and BRLE, whose performance are very close to each other, are outperformed by the proposed algorithms. IIR–γ and IIR–Kalman have the best performance where EKF algorithm has converged to a higher error level. Again, at an order less amount of multiplications, IIR–γ provides comparable results to IIR–Kalman.

4. CONCLUSION

In order to use fast recursive least squares adaptation algorithms in adaptive IIR filtering, a regressor based adaptive system structure is proposed. Two different regressor algorithms, requiring $O(N)$ and $O(N^2)$ number of multiplications respectively, are proposed to provide reliable estimates to the system output. Based on extensive set of simulations, it is found that for time–invariant systems, the proposed algorithms not only converge faster than LMS type algorithms, but also provide more reliable parameter estimates at low SNR. Additionally, in the simulation of the systems with abrupt changes, it is observed that the proposed regressor based adaptation algorithms establish faster convergence to lower error levels, outperforming BRLE, CRA and EKF.

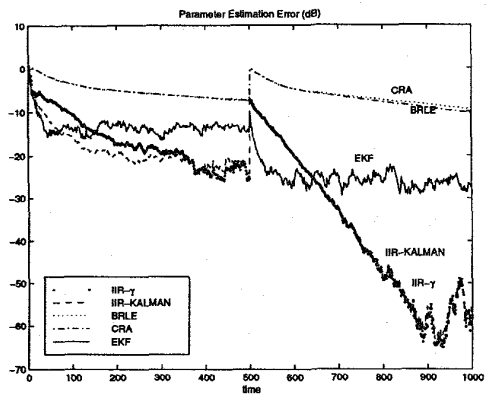

Figure 3: *Squared norm of parameter error (Example 2)*

5. REFERENCES

[1] T. Söderström and P. Stoica, *System Identification*. Englewood Cliffs, NJ: Prentice Hall, 1988.

[2] J. Lin and R. Unbehauen, "Bias-remedy least mean square equation error algorithm for IIR parameter recursive estimation", *IEEE Trans. on Signal Processing*, vol. 40, pp. 62–69, Jan. 1992.

[3] C. Chui and G. Chen, *Kalman Filtering with Real-Time Applications*. Berlin: Springer-Verlag, 1991.

[4] E. Acar, "Regressor Based Adaptive Infinite Impulse Response Filtering", Master's thesis, Dept. of Electrical Engineering, Bilkent University, Ankara, Turkey, 1997.

[5] B. Yang and J. Böhme, "Rotation-based RLS algorithms: Unified derivations, numerical properties, and parallel implementations", *IEEE Trans. on Signal Processing*, vol. 40, pp. 1151–1167, May 1992.

[6] J. Kenney and C. Rohrs, "The composite regressor algorithm for IIR adaptive systems", *IEEE Trans. on Signal Processing*, vol. 41, pp. 617–628, Feb. 1993.