Non-causal adaptive IIR filtering

Phillip M. S. Burt

***Abstract*— A procedure to approximate non-causal adaptive IIR filtering is described. Such structures could be useful, for instance, in the equalization of non-minimum phase communication channels. The procedure is based on backward calculations within blocks of samples and we show that for adaptive filtering the overlap-save method is more suitable than the overlap-add method. Moreover, a temporary freeze of the adapted parameters at the beginning of the backward calculations for each block is also required. Finally, applying a previously presented method of convergence speed analysis, we show that in the inverse identification of a causal maximum-phase system with an anti-causal adaptive filter, one of the conditions for faster convergence is always satisfied, regardless of the particular response of the system. A comparison with FIR adaptive filters is presented, showing that the proposed algorithm can achieve a better trade-off bewtween convergence speed and computational complexity.**

***Keywords*— Non-causal filters, adaptive IIR filtering, convergence speed.**

I. INTRODUCTION

The approximation of non-causal fixed IIR filters (see [1] and the references therein) is motivated by the prospect of approximating zero-phase or other particular phase responses, such as that of a Hilbert transformer. For this, samples are segmented in blocks within which calculations proceed in forward and backward manner. In this way, one can take advantage of the relatively small number of coefficients in typical IIR structures, even though the memory requirements are not likewise small. In the case of adaptive IIR filtering, the approximation of non-causal structures could be of interest, for instance, for the equalization of non-minimum phase communication channels. The reduced computational complexity, even considering the memory requirement, could be specially attractive in high-speed applications.

II. FIXED NON-CAUSAL IIR FILTERS

The impulse response $h(n)$ of a discrete-time non-causal LTI (linear and time invariant) system can be written as the convolution $h(n) = h_f(n) * h_b(n)$ of a causal response $h_f(n)$, where $h_f(n) = 0$ for $n < 0$, and an anti-causal response $h_b(n)$, where $h_b(n) = 0$ for $n > 0$. Alternatively, it could be written as the sum of a causal response and a strictly anti-causal (equal to 0 for $n \leq 0$) response, but we do not employ this form here.

A. Backward solution of difference equations

For a rational discrete-time system, the input $u(n)$ and output $y(n)$ satisfy a difference equation, as, for instance,

$$b_0u(n) + b_1u(n-1) + b_2u(n-2) = \\ = a_0y(n) + a_1y(n-1) + a_2y(n-2).$$

Dept. of Telecommunications and Control Engineering, Escola Politécnica, University of São Paulo, E-mail: phillip@lcs.poli.usp.br. This work was supported by FAPESP (Proc. 06/01113-0).

For a causal system and assuming $u(n) = y(n) = 0$ for $n < 0$, successive outputs are calculated then as

$$\begin{aligned} y(0) &= b_0u(0)/a_0, \\ y(1) &= [b_0u(1) + b_1u(0) - a_1y(0)]/a_0, \\ y(2) &= [b_0u(2) + b_1u(1) + b_2u(0) - a_1y(1) - \\ &\quad -a_2y(0)]/a_0. \end{aligned}$$

In the anti-causal case, rewriting the difference equation as

$$b_2u(n) + b_1u(n+1) + b_0u(n+2) = \\ = a_2y(n) + a_1y(n+1) + a_0y(n+2)$$

and assuming $u(n) = y(n) = 0$ for $n > 0$, successive outputs are backward calculated as

$$\begin{aligned} y(0) &= b_2u(0)/a_2, \\ y(-1) &= [b_2u(-1) + b_1u(0) - a_1y(0)]/a_2, \\ y(-2) &= [b_2u(-2) + b_1u(-1) + b_0u(0) - a_1y(-1) \\ &\quad -a_0y(0)]/a_2. \end{aligned}$$

We can see that in the anti-causal case the roles of coefficients b_k and a_k are analogous to the roles of coefficients b_{N-k} and a_{N-k} in the causal case, where N is the order of system ($N = 2$ in this example).

B. Block implementation

For an infinite length input signal, the delayed output of a stable anti-causal system $h_b(n)$ can be approximated by segmenting the input in blocks of length $L > R$, where $h_b(n) \approx 0$ for $n < -R$ [1]. If the input blocks don't overlap, then the backward calculated outputs (with zero initial conditions) of adjacent blocks effectively overlap for R samples, during which the corresponding output samples from each block must be added. This "overlap-add" method is represented in Figure 1. Alternatively, if the input blocks are made to overlap for R samples, then the corresponding output samples of one block are replaced by the output samples of the next block. This "overlap-save" method is represented in Figure 2. In either case, the R sample overlap represents a computational overhead. As discussed in the next section, the overlap-save method is more suitable for adaptive filtering.

From the condition $L > R$ it follows that the choice of R involves a trade-off between precision on one side and memory requirement and delay on the other side. Moreover, since the average (per sample) computational complexity overhead is small when $L \gg R$, the choice of L also involves a trade-off between memory requirement and computational complexity. As an example, let us consider the anti-causal all-pass system $H_b(z)$ with poles at $0.7\angle \pm 60$ and $0.7\angle \pm 100$ (all angles are in degrees and we adopt z for the unit delay, so the poles are *inside* the unit circle). The truncation error $\sum_{n=-\infty}^{-R} h_b^2(n)$ is shown in Figure 3 and serves as a guide for choosing R.

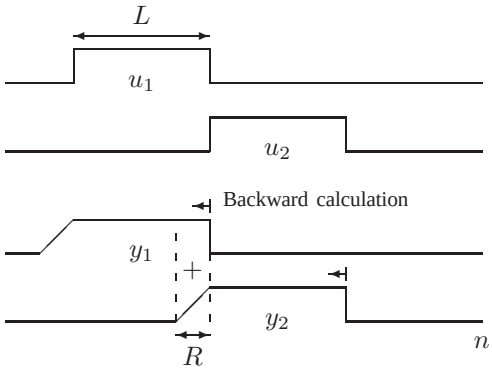

Fig. 1. Overlap-add method: blocks of input (u_j) and output (y_j) samples.

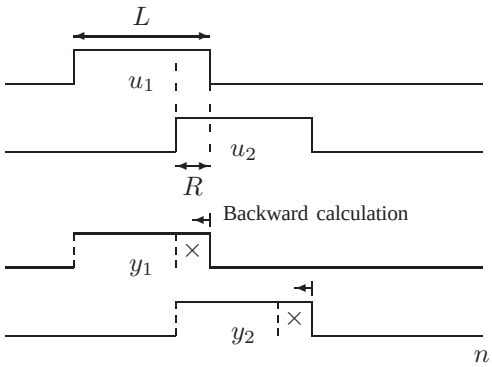

Fig. 2. Overlap-save method: blocks of input (u_j) and output (y_j) samples. Discarded samples are indicated by "×".

The square error between an overlap-save approximation $\widehat{y}_b(n)$ of the output and its true value (delayed by L samples), for an input of $2500 + R$ samples of white gaussian noise, is shown in Figure 4, for $L - R = 500$, $R = 20$ and also $R = 50$. The output blocks are backward calculated starting from $n = 500 + R$, $1000 + R$, $1500 + R$, etc. and ending at $n = 0$, 500, 1000, etc. respectively. As expected from the truncation error in Figure 3, the square error reaches a peak at these ending points and the value of this peak decreases as the overlap between blocks (given by R) increases.

III. Adaptive algorithm

We consider now that the backward calculation of $\widehat{y}_b(n)$ uses direct-form time-varying parameters $b_k(n)$ and $a_k(n)$, $k = 0, 1, \ldots, M$, which are adapted to minimize the mean-square error $E\{e^2(n)\}$, $e(n) = y(n) - \widehat{y}_b(n)$, given the input and desired signals $u(n)$ and $y(n)$, respectively.

A. *Overlap-add* vs. *overlap-save*

As the calculation of the output $\widehat{y}_b(n)$, the adaptation must also be carried out in a backward manner within each block of samples. In figures 1 and 2 this implies that, in the overlap region, the parameters used to calculate the samples of block y_2 are probably closer to their optimum values than those used to calculated the samples of block y_1. Adding these samples, as in the overlap-add method, would disturb the adaptation, whereas replacing the samples of block y_1 by those of block y_2 has no such effect. Therefore, the overlap-save method is more suitable for adaptive filtering.

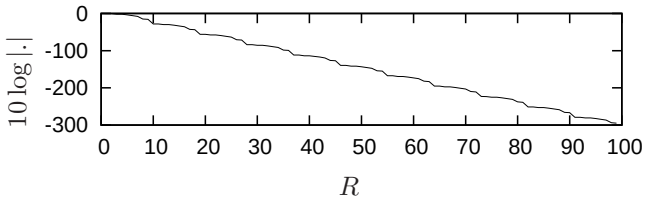

Fig. 3. Truncation error $\sum_{n=-\infty}^{-R} h_b^2(n)$.

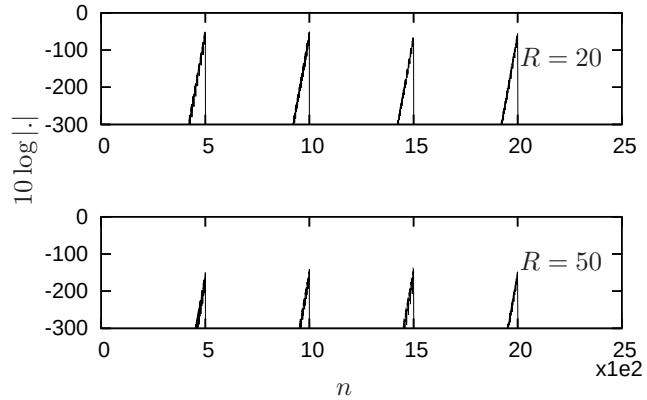

Fig. 4. Square error in overlap-save approximation of the output of anti-causal system $H_b(z)$, $L - R = 500$.

B. *Temporary adaptation freeze*

An additional aspect is raised by considering in more detail the adaptation of the parameters, for which we assume that the recursive gradient algorithm (for a detailed description see, for instance, [2, p. 271]) is employed. In this case and fixing $a_M = 1$, the following signals are necessary:

$$u_f(n) = u(n) - \sum_{k=1}^{M} a_{M-k}(n) \cdot u_f(n+k)$$

$$\widehat{y}_{bf}(n) = -\widehat{y}_b(n) - \sum_{k=1}^{M} a_{M-k}(n) \cdot \widehat{y}_{bf}(n+k).$$

The signal $u_f(n)$ already exists within the filtering structure, whereas $\widehat{y}_{bf}(n)$ has to be generated by post-filtering the output $\widehat{y}_b(n)$. After the initialization transient of these filtering operations, the mean values of the products $u_f(n+k)e(n)$ and $\widehat{y}_{bf}(n+k)e(n)$ would ideally (for slow adaptation) correspond to scaled versions of the derivatives of the mean square error with respect to parameters b_{M-k} and a_{M-k}, respectively. The problem is that, unlike a conventional forward adaptive algorithm, these filtering operations must be reset to a zero initial state – and therefore undergo a new initialization transient – after each block of samples, leading the mean value of the aforementioned products away from the desired derivatives. In order to reduce this disturbance, one procedure is to freeze the adaptation of the parameters during the first $S \approx R$ backward calculations in each block.

As a numerical example, we consider that such an anti-causal (or more precisely, backward) adaptive filter is used to identify the same anti-causal all-pass system used in Section 2. The results for a white gaussian input $u(n)$, adaptation steps $\mu_a = \mu_b = 0.04$, $L = 500$, $R = 70$ are shown in Figure 5, for

$S = 100$ and $S = 0$. The negative effect of adopting $S = 0$ is evident. The backward decrease of the squared error within each block should also be noted.

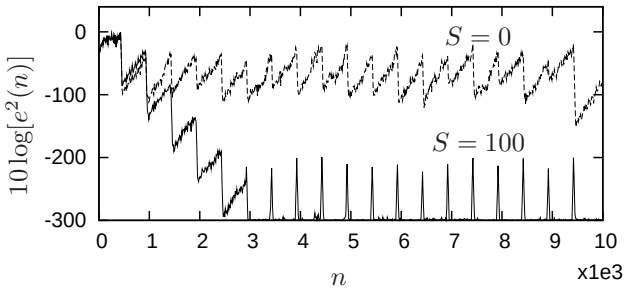

Fig. 5. Identification of an anti-causal system.

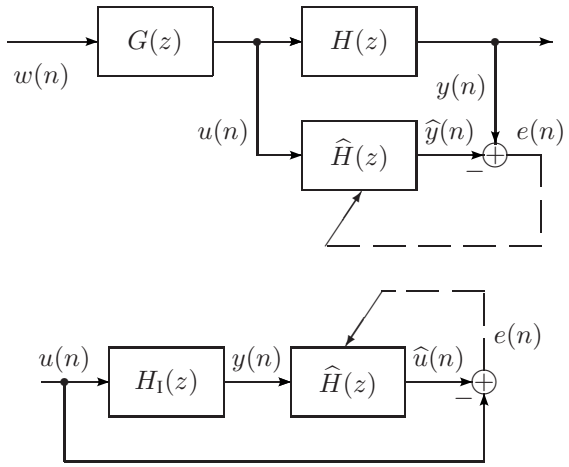

Fig. 6. Identification (top) and inverse identification (bottom) adaptive filter configurations.

IV. Convergence speed analysis

A. Previous results

In [3]-[5] we have analysed the convergence speed of causal adaptive IIR filters in the identification and inverse identification configurations. Adaptive filtering in the identification configuration and with a rational spectrum input signal is represented in Figure 6, where $w(n)$ is a unit variance white signal and $G(z)$ is a rational minimum-phase transfer function. The particular case in which, for some stable and causal minimum-phase $H_I(z)$, we have $H(z) = 1/H_I(z)$ and $G(z) = H_I(z)$ is equivalent to the inverse identification configuration with white input, also represented in Figure 6. Moreover, inverse identification with a non-white input can be obtained with $G(z) = G'(z)H_I(z)$. Thus, both configurations can be analyzed in an unified manner.

As shown in [5], a certain decomposition of the mean square error brings up the "projected function"

$$H_p(z) \triangleq \left[G\left(z^{-1}\right) H(z)\right]_{\oplus},$$

where $[.]_{\oplus}$ is the causal projection operator. We note that in the identification configuration with white input ($G(z) \equiv 1$) we obtain $H_p(z) = H(z)$ and in the inverse identification configuration with white input, $H_p(z) = [H_{\rm I}(z^{-1})/H_{\rm I}(z)]_{\oplus}$. A smaller spread of the Hankel singular values of $H_p(z)$ (eigenvalue spread which achieves its minimum when $H_p(z)$ is all-pass) tends to make convergence faster. Regarding the choice of parameterization, the proximity of the poles of $H_p(z)$ one to the other has a more negative impact on convergence speed for a direct form parameterization than for a lattice parameterization.

B. Application to non-causal adaptive filtering

Let us see now how these results can be applied to particular situations involving non-causal adaptive filtering. The identification of an anticausal $H_b(z)$ (as in the example in Section 3) is completely equivalent to the case of a causal $H(z)$ and, therefore, has equivalent convergence speed properties. However, this case seems to be of small practical interest.

On the other hand, the inverse identification of a causal non-minimum phase system $H_I(z)$ is of greater practical interest, as it relates to the equalization of communication channels. Assuming sufficient modelling, the optimal solution $\widehat{H}(z) = 1/H_I(z)$ is non-causal and, in particular, it is anti-causal if $H_I(z)$ is maximum-phase. Assuming additionally that the input to $H_I(z)$ is white, writing $H_I(z) = C(z)/D(z)$ and imposing $\widehat{H}(z) = \widehat{H}_b(z)$, where $\widehat{H}_b(z)$ is anti-causal, the mean square error can be written in terms of L_2 norms as

$$\begin{aligned} E[e^2(n)] &= \left\| 1 - \frac{C(z)}{D(z)} \widehat{H}_b(z) \right\|^2 \\ &= \left\| \frac{\overline{D}(z)}{D(z)} \frac{C(z)}{\overline{D}(z)} \left[\frac{D(z)}{C(z)} - \widehat{H}_b(z) \right] \right\|^2 \\ &= \left\| \frac{C(z)}{\overline{D}(z)} \left[\frac{D(z)}{C(z)} - \widehat{H}_b(z) \right] \right\|^2, \end{aligned}$$

where $\overline{D}(z) = z^N D(z^{-1})$, N is the order of $H_I(z)$ and the last equality follows from the fact that $\overline{D}(z)/D(z)$ is all-pass. This corresponds to the identification of an anti-causal system $H_b(z) = D(z)/C(z)$ with a non-white input signal generated with the anti-causal maximum-phase transfer function $G_b(z) = C(z)/\overline{D}(z)$, and is completely equivalent to the causal identification case in Figure 6. Therefore, the anti-causal projected function $H_{bp}(z) \triangleq \left[G_b\left(z^{-1}\right) H_b(z)\right]_{\ominus}$ is of interest, and is given by

$$H_{bp}(z) = \left[\frac{\overline{C}(z)}{D(z)} \frac{D(z)}{C(z)} \right]_{\ominus} = \frac{\overline{C}(z)}{C(z)},$$

where $\overline{C}(z) = z^N C(z^{-1})$. Quite surprisingly then, regardless of the particular maximum-phase $H_I(z)$, $H_{bp}(z)$ is always all-pass, contributing to faster convergence.

As a numerical example, we consider a unit L_2 norm $H_I(z)$ with poles at $0.7^{-1}\angle\pm 130$ and $0.7^{-1}\angle\pm 100$ and with zeros at $0.95\angle\pm 45$ and $0.95\angle\pm 135$. The frequency response $H_I(e^{j\omega})$ is shown in Figure 7 and the truncation error $\sum_{n=-\infty}^{-R} h_b^2(n)$ is shown in Figure 8. The result of the adaptation for a white gaussian signal $u(n)$ with unit variance, adaptation steps $\mu_a = 0.01$ and $\mu_b = 0.1$, $L = 1200$, and $R = S = 400$ is shown in Figure 9. In this case the poles of $H_{bp}(z)$ (which are the zeros of $H_I(z)$) are uniformly distributed, which together with the fact that $H_{bp}(z)$ is all-pass contributes to a relatively fast convergence.

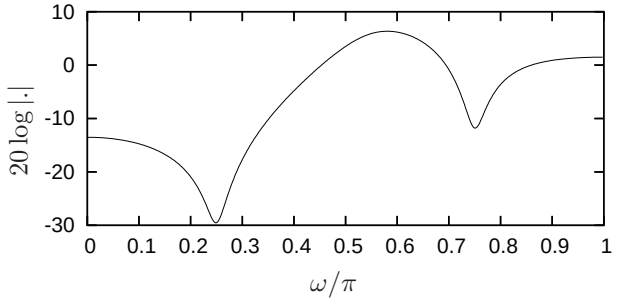

Fig. 7. Frequency response $H_I(e^{j\omega})$.

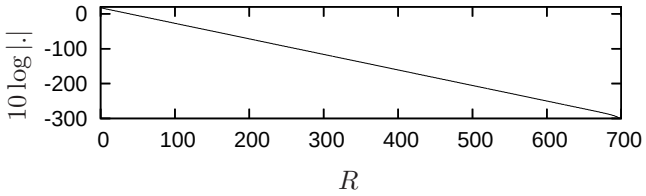

Fig. 8. Truncation error $\sum_{n=-\infty}^{-R} h_b^2(n)$.

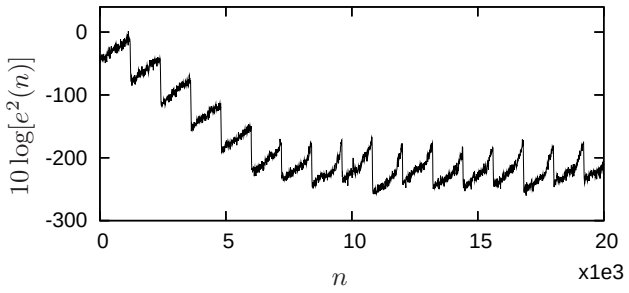

Fig. 9. Inverse identification of causal maximum-phase $H_I(z)$.

V. Comparison with FIR adaptive filtering

In the example above $H_I(z)$ has zeros close to the unit circle, leading to deep spectral notches and to a long (anti-causal) impulse response for $1/H_I(z)$. Therefore, in the inverse identification of $H_I(z)$ an FIR adaptive filter would require a large number of coefficients. Using the modified error-feedback lattice least-square (EF-LSL) algorithm of [6] with, for instance, $M = 200$ coefficients, leads to a steady-state mean square error of approximately -75 dB, as showed in Figure 10 (a forgetting factor $\lambda = 0.9$ was used and the prediction error energies were initialized with 1). A similar steady-state performance can be obtained using an IIR adaptive filter as described above, now with $L = 600$, $R = S = 200$, $\mu_a = 0.01$ and $\mu_b = 0.09$, as can also be seen in Figure 10. The EF-LSL algorithm converges in approximately $2M = 400$ samples and is faster than the IIR algorithm (which requires 3500 samples). However, its computational complexity is much larger, as presented in Table I. Considering now the LMS adaptive algorithm, it has smaller computational complexity than the EF-LSL algorithm (though still much larger than the IIR algorithm) but converges much slower, due to the large number of coefficients and the correlation of the input samples (an adaptation step of $\mu = 0.005$ was used, which is close to the maximum value allowed in this case). In cases such as this, therefore, it is likely that the IIR algorithm would represent the best trade-off between convergence speed and computational complexity.

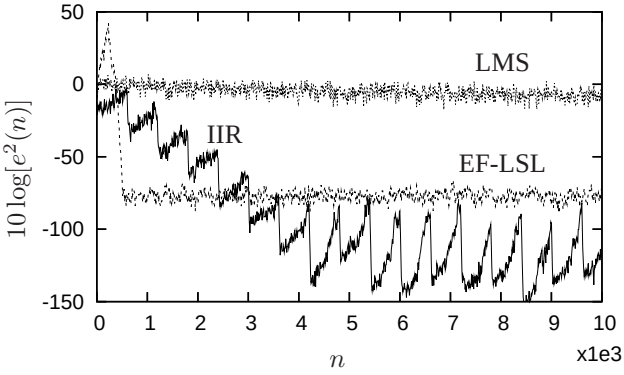

Fig. 10. Inverse identification of maximum-phase $H_I(z)$: typical realizations of the IIR, EF-LSL and LMS algorithms (smoothed for better visualization).

TABLE I

Number of computations per signal sample

	IIR	EF-LSL	LMS
N, M	4	200	200
$+$	$(5N+2)(1+\frac{R}{L})$ $=$ 29.3	$9M$ $=$ 1800	$2M$ $=$ 400
$\times$	$(5N+4)(1+\frac{R}{L})$ $=$ 32	$13M$ $=$ 2600	$2M+1$ $=$ 401
$\div$	0	$2M$ $=$ 400	0

VI. Conclusion

A procedure to approximate non-causal adaptive IIR filtering was described and some important implementation aspects were discussed. Regarding convergence speed, we showed that in the inverse identification of a causal maximum-phase system with an anti-causal adaptive filter, a previously obtained condition for faster convergence is, quite surprisingly, always satisfied, regardless of the particular response of the system. The presented IIR algorithm can achieve a better trade-off between convergence speed and computational complexity than FIR adaptive filters, in certain situations. An important continuation of this work would be to take into acount more general non-causal adaptive filters, comprising causal and anti-causal stages.

References

[1] C. M. Rader and L. B. Jackson, "Approximating noncausal IIR digital filters having arbitrary poles, including new Hilbert transformer designs, via forward/backward block recursion," *IEEE Transactions on Circuits and Systems I*, vol. 53, no. 12, pp. 2779–2787, Dec. 2006.

[2] P. A. Regalia, *Adaptive IIR filtering in signal processing and control*, Marcel Dekker, New York, 1995.

[3] P. M. S. Burt and P. A. Regalia, "A new framework for convergence analysis and algorithm development of adaptive IIR filters," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3129–3140, Aug. 2005.

[4] P. M. S. Burt, "Inverse identification adaptive IIR filtering: convergence speed analysis and successive approximations algorithm," in *Proc. IEEE Icassp*, Honolulu, 2007, vol. 3, pp. 1309–1312.

[5] T. E. Filgueiras Filho and P. M. S. Burt, "On the convergence speed of adaptive IIR filters with rational spectrum input signals," in *Proc. European Signal Processing Conference (Eusipco)*, Lausanne, 2008.

[6] M. Miranda, M. Gerken, and M.T.M. da Silva, "Efficient implementation of error-feedback LSL algorithm," *Electronics Letters*, pp. 1308–1309, 1999.