

Relative Convergence of the Cascade RLS with Subsection Adaptation Algorithm

G. Zakaria^{1,2} and A. A. (Louis) Beex²

¹Hughes Network Systems, Germantown, MD 20876, USA

²Systems Group - DSP Research Laboratory

The Bradley Department of Electrical and Computer Engineering
VIRGINIA TECH, Blacksburg, VA 24061-0111, USA

Abstract

We analyze the convergence behavior of the CRLS-SA algorithm for inverse filtering. CRLS-SA is a cascade adaptive filter based on the RLS algorithm, with each section adapted independently based on global minimization. The subsection adaptation results in reduced computational complexity. The rate of convergence is evaluated based on the convergence time constant, defined as the ratio of condition number and sensitivity. The smaller the convergence time constant, the faster the structure converges. Analysis and simulation explain and show that CRLS-SA exhibits faster convergence than the direct form RLS adaptive filter for speech type signals.

1. Introduction

For short data records the *Cascade Recursive Least Squares with Sub-section Adaptation* (CRLS-SA) algorithm estimates an AR process more accurately than the autocorrelation method (ACM) [12]. CRLS-SA presents a computational load comparable to ACM. Furthermore, CRLS-SA can be modified to compute line spectral frequencies directly [2], without additional computations and without the approximations common in a widely used method [9].

The CRLS-SA achieves a low computational load because for linear prediction (LP) applications, the gradient autocorrelation matrix can be decoupled. Hence the adaptation of each section involves only the autocorrelation matrix of that section, while the minimization is still done globally.

Here we look at the convergence properties of the CRLS-SA algorithm. The rate of convergence is based on the conjecture [1] that a structure that has a smaller convergence time constant τ will converge faster than a structure with larger τ . The time constant τ is the ratio of the condition number and the sensitivity.

We will show that CRLS-SA converges faster than the direct form RLS (DFRLS) for speech-like signals. The rate of convergence is defined by how fast the estimated model approaches the known system in terms of Itakura distance.

This paper is organized as follows. Section 2 looks at the convergence properties of CRLS-SA. In Section 3 we report on simulations using parameters derived from real speech signals. We conclude the paper in Section 4.

2. Relative Convergence

The CRLS-SA block diagram is shown in Figure 1.

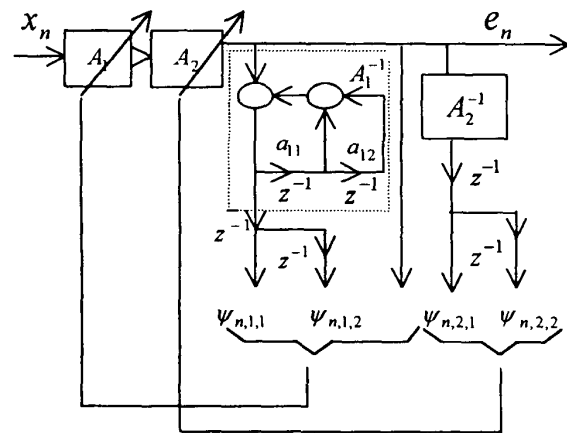


Fig. 1: CRLS-SA Gradient Evaluation.

We analyze the convergence rate of CRLS-SA using the convergence time constant [1].

2.1 Condition Number

The condition number is defined as:

$$\chi(R) = \frac{\lambda_{\max}}{\lambda_{\min}} \quad (1a)$$

where λ_{\max} and λ_{\min} are the maximum and minimum eigenvalues of the $N \times N$ autocorrelation matrix R , of a wss vector signal \underline{x}_n of size N , defined on the basis of

$$r(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} x(n)x(n-k); k=0,1,\dots,N-1 \quad (1b)$$

$$R = \begin{bmatrix} r(0) & \dots & r(N-1) \\ \vdots & \ddots & \vdots \\ r(N-1) & \dots & r(0) \end{bmatrix} \quad (1c)$$

Given that $S(\omega)$ is the power spectral density of \underline{x}_n , the condition number is bounded by [7]:

$$\chi(R) = \frac{\lambda_{\max}}{\lambda_{\min}} \leq \frac{\max_{\omega} S(\omega)}{\min_{\omega} S(\omega)} \quad (1d)$$

For an AR process of order N , the components of \underline{x}_n are generated by:

$$x_n = \frac{1}{A(z)} u_n \quad (2)$$

where $A(z) = 1 - \sum_{k=1}^N a_k z^{-k}$, and u_n is white Gaussian noise.

The power spectral density (PSD) of the AR process can be written in terms of $S_i(\omega)$, the PSD for a pole pair.

$$S(\omega) = \prod_{i=1}^N S_i(\omega) \quad (3)$$

Because of interactions among the poles, the peaks of the PSD due to each pole pair can be different from the peaks of the PSD for that pole pair itself without interaction, i.e. for a second order AR process. Let

$$\omega_{i,\max} = \arg \max_{\omega} S_i(\omega) \quad (4a)$$

$$\omega_{i,\min} = \arg \min_{\omega} S_i(\omega) \quad (4b)$$

so that we have for any ω_i ($\omega_{i,\max}$ or $\omega_{i,\min}$)

$$\prod_{\substack{k=1 \\ k \neq i}}^N S_k(\omega_{i,\max}) \underset{b}{\overset{a}{>}} 1 \Leftrightarrow S(\omega_{i,\max}) \underset{b}{\overset{a}{>}} S_i(\omega_{i,\max}) \quad (4c)$$

Under condition a , $\forall i$, it follows that

$$\begin{aligned} \max_{\omega} S(\omega) &\geq \max_i S(\omega_{i,\max}) \\ &\geq \max_k S(\omega_{k,\max}) \\ &\geq S_k(\omega_{k,\max}); k=1, \dots, N \end{aligned} \quad (4d)$$

Equations (4d) tend to be satisfied when the poles of the AR process are closely spaced, and say that the biggest peak of $S(\omega)$ exceeds the largest PSD peak for any individual pole pair. Using analogous reasoning, condition b leads to the minimum of $S(\omega)$ being smaller than the minimum of any individual pole pair PSD, that is

$$\min_{\omega} S(\omega) \leq S_k(\omega_{k,\min}); k=1, \dots, N \quad (4e)$$

For a direct form AR (DFAR) process with tightly clustered poles, condition a in equation (4c) is likely to hold for $\max S(\omega)$ while condition b is likely to hold for $\min S(\omega)$. Hence we surmise that for an order N DFAR process with closely clustered poles, the condition number is always greater than the condition number for any of its constituent individual second order AR processes. We offer two cases as examples: Case 1, a fourth order AR process with closely clustered poles, and Case 2, a fourth order AR process with well separated poles. The theoretical results are shown in Table 1.

Table 1. Condition Number

	Case		DF	Section	
	Pole 1	Pole 2		1	2
I	$.8e^{\pm j\pi/8}$	$.9e^{\pm j\pi/5}$	1950	12.3	5.7
II	$.8e^{\pm j\pi/8}$	$-.9e^{\pm j\pi/5}$	3.9	1.5	5.7

For a higher order AR prediction error filter (PEF) in cascade form, the input signal to the i -th section of a cascade structure can be expressed as

$$x_{n,i} = \frac{\prod_{k=1}^{i-1} \tilde{A}_k(z)}{A(z)} u_n \quad (5)$$

where $\tilde{A}_i(z)$ is the estimate for the i -th section. The input to the first section, $x_{n,1}$, is equal to x_n . This shows that the input signal to the first section reflects all of the poles, just as the input signal to the DFAR parameter estimator, while the input signal to the last section reflects only poles that have not been removed in the previous sections. Hence, the autocorrelation matrix of the input signal to the first section is the same as the leading 2×2 component of the $N \times N$ autocorrelation matrix R of the DF. As a result, the maximum eigenvalue of the first section of the cascade is always less than the maximum eigenvalue of the DF and the minimum eigenvalue of the first section of the cascade is always bigger than the minimum eigenvalue of the DF. In other words, the condition number of the first section of the cascade is always less than the condition number of the DF. Recall, for NND matrices, that if we take an $m \times m$

matrix out of an $N \times N$ matrix, where $m < N$, then the maximum eigenvalue of the $m \times m$ matrix is smaller than the maximum eigenvalue of the $N \times N$ matrix, while the minimum eigenvalue of the $m \times m$ matrix is larger than the minimum eigenvalue of the $N \times N$ matrix.

For the final section, if (4d) and (4e) hold, the condition number is smaller than that of DF. For any other sections, (4d) and (4e) hold if the poles are closely spaced. In other words, the condition number of each section of the cascade structure is always smaller than the condition number of the DF for signals with closely spaced poles, such as speech signals for example.

2.2 Sensitivity

Sensitivity is defined as [3]:

$$S = \sum_{i=1}^N \left| \frac{\partial h(n, \theta)}{\partial \theta_i} \right|^2 \quad (6)$$

where $h(n, \theta)$ is the impulse response, and θ are the parameters in the structure. Define the impulse response of DF PEF for an AR process with N coefficients as

$$h_{n,DF} = A(z) \delta_n \quad (7)$$

where

$$A(z) = 1 - \sum_{i=1}^N a_i z^{-i} \quad (8)$$

Now the derivative of the impulse response with respect to each coefficient is

$$\frac{\partial h_n}{\partial a_i} = \frac{\partial A(z)}{\partial a_i} \delta_n = -\delta_{n-i} \quad (9)$$

where δ_n is the unit sample sequence. By substituting (9) into (6), the sensitivity of a DF PEF is:

$$S = \sum_{i=1}^N |\delta_{n-i}|^2 = N \quad (10)$$

The system response $A(z)$ of a cascade structure PEF, with each section a second order filter is

$$A(z) = \prod_{i=1}^{N/2} A_i(z) \quad (11)$$

$$A_k(z) = 1 - a_{k,1} z^{-1} - a_{k,2} z^{-2} \quad (12)$$

and its impulse response is

$$h_{n,c} = \prod_{i=1}^{N/2} A_i(z) \delta_n \quad (13a)$$

We can show that the derivative of the impulse response with respect to each coefficient is:

$$\frac{\partial h_{n,c}}{\partial a_{k,l}} = z^{-l} \prod_{\substack{i=1 \\ i \neq k}}^{N/2} A_i(z) \delta_n = \frac{z^{-l}}{A_k(z)} h_{n,c} \quad (13b)$$

where $l=1,2$.

From (13b) we see that the derivative of the impulse response of the cascade structure with respect to the coefficients of the k -th section are obtained by taking the impulse response of all sections but the k -th section. Here, we see that the DF, from (10), has potentially low sensitivity compared to the cascade structure. Note that for the DF PEF, a change in a coefficient will affect all its roots, the zeros in this case. On the other hand, changing a coefficient in a section of the cascade structure will affect only the roots of that section.

Using the poles in the previous sections, the sensitivities of the DF and cascade structure PEFs are computed. The theoretical results are shown in Table 2. Here we see that the DF PEF has lower sensitivity than the cascade structure PEF. Note that for section 1, Case I and II yield the same sensitivity because the poles are still the same. For section 2, the radius of the poles for Case I and II are the same, but the frequencies are different. As a result, the sensitivity for Case I and II are also the same.

Table 2. Sensitivity

	Case		DF	Section	
	Pole 1	Pole 2		1	2
I	$.8e^{\pm j\pi/8}$	$.9e^{\pm j\pi/5}$	4.0	6.5	7.1
II	$.8e^{\pm j\pi/8}$	$-.9e^{\pm j\pi/5}$	4.0	6.5	7.1

Even though the DF has a lower sensitivity than the cascade structure, this does not mean that the overall effect of a change in the DF to the system is less than the overall effect in the cascade structure. To verify this, a small perturbation is imposed on the parameters of the DF and cascade structures. Then the distances of the perturbed DF and cascade structures to the true system are compared. The Itakura distance is used, as it is considered more appropriate than the Euclidean distance for measuring the distance between two sets of highly correlated linear prediction (LP) parameters [5]. The Itakura distance is

$$d[A(z), B(z)] = \log \frac{B'(z)RB(z)}{A'(z)RA(z)} \quad (14a)$$

where R is as defined in (1), but of size $(N+1) \times (N+1)$, and the LP prediction parameter polynomials are:

$$A(z) = 1 - \sum_{k=1}^N a_k z^{-k} \quad (14b)$$

$$B(z) = 1 - \sum_{k=1}^N b_k z^{-k} \quad (14c)$$

Averaging over 100 realizations we found that for closely spaced poles (Case I), the Itakura distance of the DF PEF is 17.59 after perturbing its parameters, while for the cascade structure PEF it is 1.81. For poles that are spaced far apart (Case II), the Itakura distance of the DF PEF is 1.18 after parameter perturbation, while for the cascade structure PEF it is 1.81. For these simulations, the coefficients were perturbed by adding a small random quantity (zero mean, equal variance). We also see here that sensitivity alone does not characterize the structure. In the two DF systems above, which have the same sensitivity, the effect of parameter perturbation is different. On the other hand, for the two cascade systems, the parameter perturbations yield the same Itakura distance.

2.3 Convergence Time Constant

Define the convergence time constant τ [3] for the direct form as

$$\tau = \frac{\text{condition number}}{\text{sensitivity}} \quad (15)$$

and for the cascade structure use

$$\tau_c = \prod_{k=1}^{N/2} \tau_k \quad (16)$$

where τ_k is the convergence time constant for the k^{th} section. Using (1d) for the condition number and (10) and (13b) for the DF and cascade sensitivity respectively, we see that for an AR process with tightly clustered poles, such as a speech signal, the cascade structure will have a smaller convergence time constant. Hence the cascade structure is expected to converge faster than the DF. For an AR process with poles that are well separated, there is a possibility that the DF will converge as fast as, or possibly faster than the cascade structure. To verify this, some simulations are conducted. Using the result from the condition number in Section 2.1 and the sensitivity in Section 2.2, the τ 's are as shown in Table 3.

Table 3. Convergence Time Constant

Case	Pole		DF	Section		
	1	2		1	2	1+2
I	$.8e^{\pm j\pi/8}$	$.9e^{\pm j\pi/5}$	487	1.87	.8	1.4
II	$.8e^{\pm j\pi/8}$	$-.9e^{\pm j\pi/5}$.975	.22	.8	.16

We see that for an AR process with tightly-clustered poles (Case I), the τ for the cascade structure is smaller than for DF. Hence the cascade structure is expected to

converge faster. For a system with poles that are far apart, some sections of the cascade structure might converge faster than the DF, while other sections converge slower.

It is difficult to determine the convergence rate from the trajectory of the coefficients. Instead, we measure the Itakura distance between the estimated and true denominator polynomial at any time instant n . Then we look at how fast these distances, for DF and cascade, reach steady state and how small the distance has become.

3. Simulation Experiments

Our convergence analysis is supported with results based on averaging 100 realizations. The first system has poles that are closely spaced (Case I). Figure 2 shows that CRLS-SA converges much faster than DF. This corresponds to τ for CRLS-SA, which at 1.4 is much smaller than that for DF, which is 487.

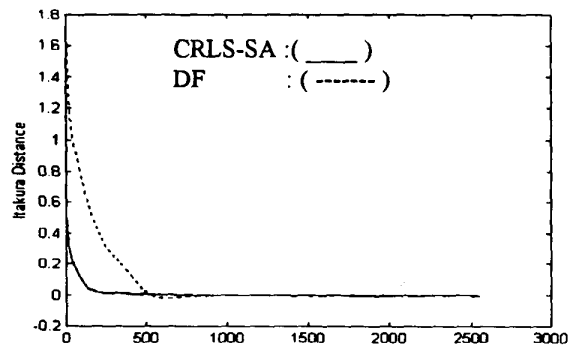


Fig. 2: Convergence for Case I.

The second system has poles that are far apart (Case II). In Figure 3 we see that, because the convergence time constant for CRLS-SA (of 0.16) is just a little bit smaller than that for DF (which is .975), CRLS-SA converges just a little bit faster than DF.

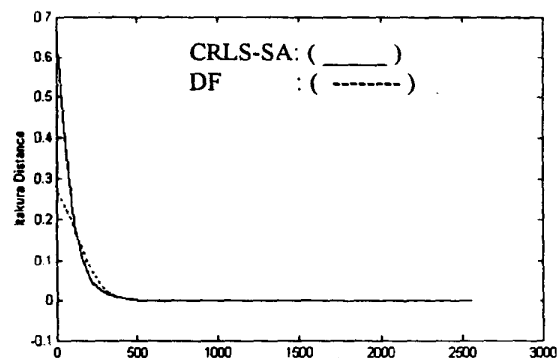


Fig. 3: Convergence for Case II.

In Figures 2 and 3 we also see that CRLS-SA and DF start at a different initial distance. This is due to the different initializations that were used. CRLS-SA is initialized by distributing the poles evenly around the unit circle and assigning each pole pair to a section, while DF is initialized with zero values.

The next simulation example uses 10th order AR processes. The coefficients of the AR processes are obtained from speech signals. That is, from a frame of speech, we estimated the AR coefficients. Then we use this estimate to generate outputs for the corresponding AR process. The speech frames were chosen to correspond to voiced and unvoiced speech respectively. The results for CRLS-SA, along with those for DF, are shown in Figures 4 and 5 for voiced and unvoiced speech frames respectively.

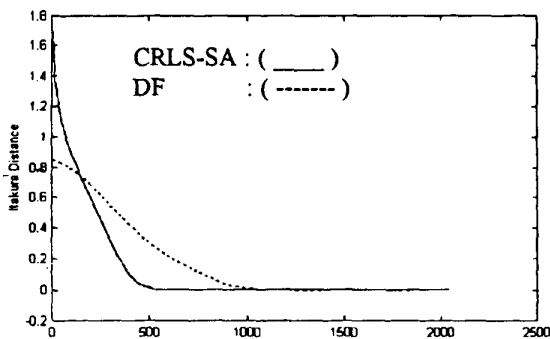


Fig. 4: Convergence for Voiced Speech.

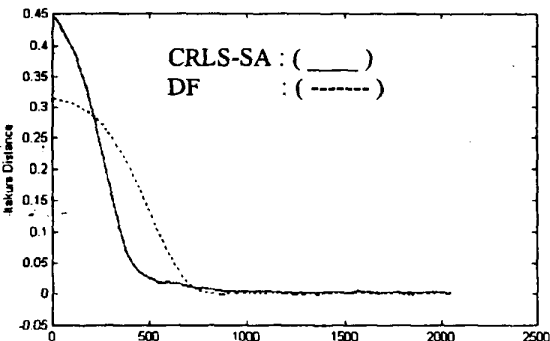


Fig. 5: Convergence for Unvoiced Speech.

We see from Figure 4 that because the convergence time constant of about 0.3 for CRLS-SA, is much smaller than that for DF, of about 124, the Itakura distance for CRLS-SA reaches steady state faster than for DF. This means that CRLS-SA approaches the true system faster than DF.

For the unvoiced speech in Figure 5, we see that because the convergence time constant of about 1.5, for

DF, is just a little bit larger than that of about 0.04 for CRLS-SA, DF converges just a little bit slower than CRLS-SA.

4. Conclusion

The CRLS-SA algorithm converges faster than the direct form for linear prediction applications, especially for systems where the poles are tightly clustered. This is because the convergence time constant for CRLS-SA is smaller than for the direct form. The condition number of each section of CRLS-SA is pretty much determined by its own poles, and there is little influence from the poles of the other sections, as is the case with the direct form. For a real system in which the order is relatively high, the pole locations cannot be that far apart, so that interaction among the poles cannot be avoided. The latter makes CRLS-SA a more favorable choice than DF for real systems such as used in speech processing.

5. References

1. A. A. (Louis) Beex and V. E. DeBrunner, *The Effect of Identifier Structure on Parameter Convergence*, *IEEE Transactions on Signal Processing*, vol. 40, no. 11, pp.2819-2822, Nov. 1992.
2. A. A. (Louis) Beex and G. Zakaria, *Direct Computation of Line Spectral Frequency Using Cascade Recursive Least Squares with Subsection Adaptation*, 33rd Asilomar, 1999.
3. V. E. DeBrunner and A. A. (Louis) Beex, *Sensitivity of Structures for the Identification of Linear Systems from Impulse Response Data*, ICASSP'89, Glasgow, Scotland, pp. 2214-2217, May 1989.
4. _____, *An Informational Approach to the Convergence of Output Error Adaptive IIR Filter Structures*, ICASSP'90, pp. 1261-1264.
5. J. R Deller Jr., J. G. Proakis, J. H. L. Hansen, *Discrete-time Processing of Speech Signals*, New York, Macmillan, 1993.
6. H. Eves, *Elementary Matrix Theory*, Dover Publications, Inc., New York, 1966.
7. S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, Prentice Hall, 1991.
8. L. B. Jackson and S. L. Wood, *Linear Prediction in cascade form*, *IEEE Transactions on ASSP*, vol. ASSP-26, pp. 519-528, Dec. 1978.
9. P. Kabal and R. P. Ramachandran, *The Computation of Line Spectral Frequencies Using Chebyshev Polynomials*, *IEEE Transactions on ASSP*, vol. ASSP-34, no.6, December 1986.
10. S. M. Kay, *Modern Spectral Estimation*, Prentice Hall, 1988.
11. B. D. Rao, *Adaptive IIR Filtering Using Cascade Structures*, International Symposium on Circuits and Systems, pp. 194-197, 1993.
12. G. Zakaria and A. A. (Louis) Beex, *Cascade Recursive Least Squares with Subsection Adaptation for AR parameter Estimation*, ICASSP 98, Seattle, pp. 953-956, May 1998.