Convergence Performance of the Cascaded RLS-LMS Prediction

Dong-Yan Huang and Susanto Rahardja
Signal Processing Department
Institute for Infocomm Research
21 Heng Mui Keng Terrace
Telephone: (65) 6874 2581
Fax: (65) 6874 7588

Abstract—In this paper, we use a stochastic fixed-point theorem to analyze the stochastic convergence properties of the cascaded RLS-LMS prediction filter in terms of conditions of convergence and the misadjustment. It is shown that the cascaded RLS-LMS prediction filter converges to almost the same optimal solution of the conventional RLS filter. The misadjustment is shown to be exponentially dependent on the number of stages in the cascade structure and is higher than the misadjustment of the conventional RLS filter. However, the cascaded RLS-LMS prediction filter allows us to build up a low complexity RLS-like predictor with time-varying learning rate, which may be useful in uncertain and non-stationary environments.

I. INTRODUCTION

Recent research results have shown that the structure of cascaded adaptive filters displays an interesting ability to converge to a good approximation of the optimal predictor [1], even better than more expensive computational and sophisticated approaches [2]. The eminent practical cascade structure results in the interest of applying this structure into various applications such as speech and audio compression [1], [3], [4], system identification [5], echo cancellation [6]. Little work has been devoted to the fundamental problems in the convergence of the structure of cascaded adaptive filters, such as a special case of cascaded predictors using one-tap LMS filters based on Wiener filter theory [1]. Computer simulations have, however, shown that the structure of the cascade features faster adaptation speed and higher misadjustment and better numerical stability. Especially, the most successful experiments have happened to the cascade structures of minimized sized stages using LMS and RLS-like adaptation. Therefore, the theoretical analysis of this structure is major concern of this paper.

A stochastic version of a fixed-point theorem, which was developed by Oza [7], has been taken of the adaptive gradient lattice based on the notion of a contraction mapping on a "stochastic" Hilbert space [8], [9]. Under the same framework, this paper applies this fixed-point theorem as the basis in establishing the convergence conditions, and misadjustment of the cascaded RLS-LMS prediction filter by considering a "distance" measure in the appropriate space. In this work, the convergence of the filter is taken as an operator, rather than convergence of some filter parameters [1].

The paper is organized as follows. In Sec. II, we introduce the concepts such like operator and inner product in Hilbert space $l_2(\Omega)$, the stochastic fixed-point theorem for the study of the convergence properties of the cascade structure. In Section III, we briefly present the RLS, the LMS algorithms, as well as assumptions for the cascaded RLS-LMS prediction filter. In Section IV-A, we study the convergence conditions of the RLS and the LMS prediction filters under the fixed-point theorem respectively. In section V, we derive the misadjustment of the cascaded RLS-LMS algorithm. The simulation results will be shown to confirm our theoretical analysis in Section VI. The paper concludes with Section VII.

II. BACKGROUND

A. The Cascaded Linear Prediction

The general structure of the cascade for the linear prediction is depicted in Fig. 1. To address the study the convergence properties of this structure, we introduce the following concepts. First of all, each stage will formally be considered as the elementary operator $T_k$ (Fig. 2): The global predictor transfer function can be expressed as the cascade of elementary operators

$$T = \prod_{k=1}^{M} T_k$$

where $M$ is the number of stages. Secondly, the $T_k$ is assumed to be an operator on $l_2(\Omega)$, the Hilbert space of zero-mean, wide sense stationary stochastic ($\Omega \times T$) sequences at $k$th stage with inner product

$$\langle e_k(\cdot), e_k(\cdot) \rangle = \Delta \left\{ e_k(\cdot)e_k(\cdot) \right\}$$

Fig. 2. $k$th element in the cascaded predictors
It is also assumed that \( e_k(\cdot) \) has finite average power. In the cascade structure, each stage is an FIR predictor of order \( l_k \). Assume that \( x_k(n) \) be the input to stage \( k \), and let \( e_k(n) \) be the corresponding prediction error, we obtain

\[
e_k(n) = x_k(n) - \sum_{m=1}^{l_k} h_k(m)x_k(n - m)
\]

\[
\Delta = h_{1k} U e_{k-1}(n)
\]

where \( z^{-1} \) is the unit shift operator, and \( U \) is the matrix defined by

\[
U = \begin{bmatrix}
1 & 0 & \cdots & 0 \\
0 & z^{-1} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \cdots & z^{-l_k}
\end{bmatrix}
\]

and

\[
h_{1k} = \begin{bmatrix}
1 & -h_k(1) & \cdots & -h_k(l_k)
\end{bmatrix}
\]

The cascade structure is such that \( x_{k+1}(n) = e_k(n); x_1(n) = x(n) \), where \( x(n) \) is the signal to predict. The global prediction error of the structure is the error of the last stage \( e_M(n) \). The norm of \( T_k \) is given by the following formulate

\[
\|T_k\| = \sup_{\|e_{k-1}(n)\| = 1} \{\|h_{1k} U e_{k-1}(n)\|\}
\]

where \( \|\cdot\| \) denotes the norm induced by the inner product (2). Since \( U \) is an unitary operator, (6) reduces to the simple filter coefficients vector \( h_{1k} \), which is given by its spectral radius [10]

\[
\|T_k\| = \|h_{1k}\|_{\text{max}}
\]

where \( \|h_{1k}\|_{\text{max}} \) is the largest value of \( |h_{1k}| \). Likewise, a lower bound on the output energy can be obtained by replacing \( \sup \) by \( \inf \) and using the same procedure. We can obtain

\[
\|h_{1k}\|_{\text{min}} \|e_{k-1}(n)\| \leq \|e_k(n)\| \leq \|h_{1k}\|_{\text{max}} \|e_{k-1}(n)\|
\]

Note that the lower bounds is actually attained when \( h_{1k} \) is optimal linear predictor.

### B. The Cascaded Adaptive Prediction Filter

So far, \( T_k \) is considered as a fixed operator on the space \( l_2(\Omega) \). If each stage in the cascade uses the adaptive algorithm, however, two issues should be considered. First of all, if a time-recursive method is used, the filter becomes time-varying, thus the assumption of wide-sense stationarity of the inputs to the successive stage is no longer valid. Secondly, the parameters of filter are statistical measured, consequently, the assumption of Gaussian statistics of the input to each stage will be violated. In the cascade structure, the adaptive predictor is independently adapting at each stage. In order to evaluate the transient behavior as well as the steady-state response of the cascaded adaptive predictor, we make the following assumption for each stage of the cascade structure:

**Assumption 1:** The statistics of the sequences \( x_k(n-1) \) and the desired response \( x_k(n) \) is jointly Gaussian at each stage.

To calculate the mean-square norm of the stochastic cascade element, we take again (3) where now the filter coefficients \( h_{1k}(n) \) are random variables:

\[
\|T_k(n)\|^2 = \sup_{\|e_{k-1}(n)\| = 1} \{\|e_{k-1}(n)h_{1k}(n)\|^2\}
\]

\[
\leq \sup_{\|e_{k-1}(n)\| = 1} E\{\|e_{k-1}(n)h_{1k}(n)h_k(n)\|^2\}
\]

\[
\leq E\{\|h_{1k}(n)\|^2\}
\]

Under the widely used assumption that the operator \( h_k(n) \) and the sequence \( e_{k-1}(n) \) are uncorrelated, (9) becomes

\[
\|T_k(n)\|^2 \leq E\{\|h_{1k}(n)h_k(n)\|^2\}
\]

(10)

(using the Schwarz inequality and the definition of vector norm). As the filter coefficient-vector norm is attained by an eigenvector, in which case the equality in the Schwarz inequality holds, we have

\[
\|T_k(n)\|^2 = E\{\|h_{1k}(n)\|^2\}
\]

(11)

Assume \( \sigma_k^2 \) and \( \bar{h}_{1k} \) are the variance and mean, respectively, of \( h_{1k}(n) \), we can write bounds for \( \|e_k(n)\| \) similar to (8)

\[
(\sigma_k^2 + \bar{h}_{1k}^2)\|e_{k-1}(n)\|^2 \leq \|e_k(n)\|^2 \leq (\sigma_k^2 + \bar{h}_{1k,\text{max}}^2)\|e_{k-1}(n)\|^2
\]

(12)

### C. The Stochastic Fixed-Point Theorem

Here we briefly describe a stochastic fixed-point theorem used by Oza [7] in a system identification problem. We shall
employ it for the discussion on the conditions on the step-size \( \mu_k \), the misadjustment, as well as the convergence rate in the following Sections.

**Theorem 1:** Let \( \{T_k(n)\}_{n=0}^{\infty} \) be a sequence of (random) operators on a Hilbert space \( \mathcal{H} \) and let \( T_k(n) \rightarrow T_k \) where \( T_k \) is a contraction mapping i.e.,

\[
\lim_{n \rightarrow \infty} \| T_k(n)y - T_ky \| = 0 \quad \forall y \in \mathcal{H}
\]  

(13)

and

\[
\| T_ky_1 - T_ky_2 \| \leq \| y_1 - y_2 \| \quad \forall y_1, y_2 \in \mathcal{H}
\]  

(14)

If it is assumed that \( T_k \) has a fixed point, then the sequence generated by

\[
y(n+1) = T_k(n)y(n) \quad (15)
\]

\( y_0 \) fixed, but arbitrary in \( \mathcal{H} \), converges strongly to the fixed-point of \( T_k \).

### III. THE CASCADED RLS-LMS PREDICTION FILTER

#### A. The Recursive Least Squares (RLS) Algorithm

The RLS algorithm consists to estimate the filter coefficient-vector \( \hat{h}_k(n) \) at the \( k \)-th stage, that satisfies the following least-square (or \( H^2 \)) criterion:

\[
\min_{\hat{h}_k} \| \hat{h}_k - h_k(-1) \|^2 + \sum_{n=0}^{i} |y_k(n) - x_k(n-1)h_k(n-1)|^2
\]  

(16)

where \( \hat{h}_k(-1) \) is the initial estimate of \( h_k \), and \( \delta > 0 \) represents the relative weight that we give to our initial estimate compared to the "sum of squared-error" term \( \sum_{n=0}^{i} |y_k(n) - x_k(n-1)h_k| \).

The RLS (Recursive Least Square) algorithm can get the exact solution to the above criterion:

\[
\hat{h}_k(n) = h_k(n-1) + K_k(n)[x_k(n) - x_k(n-1)\hat{h}_k(n-1)]
\]  

(17)

and \( \hat{h}_k(-1) \) with

\[
K_k(n) = \frac{P_k(n-1)x_k(n-1)}{1 + x_k(n-1)P_k(n-1)x_k^*(n-1)}
\]  

(18)

and \( k(n) \) satisfying the Riccati recursion

\[
P_k(n) = P_k(n-1) - K_k(n)x_k(n-1)P_k(n-1)
\]  

(19)

and

\[
P(0) = \delta I
\]

#### B. The Least-Mean Squares (LMS) Algorithm

It is well known that the LMS filter is a natural implementation of the Wiener solution to solve the linear filter problem under small-step-size theory [11]. The coefficients of the predictor at the \( k \)-th stage are updated according to the LMS (least-Mean-Squares) algorithm by their instantaneous values:

\[
\hat{h}_k(n) = \hat{h}_k(n-1) + \mu_k e_k(n)x_k^*(n-1)
\]

\[
= \hat{h}_k(n-1) + \mu_k x_k^*(n-1)[x_k(n) - x_k(n-1)\hat{h}_k(n-1)]
\]  

(20)

where \( e_k(n) = x_k(n) - x_k(n-1)\hat{h}_k(n-1) \). \( \mu_k \) is the so-called stepsize of the algorithm and \( \hat{h}_k(n) \) denotes the adaptive estimates.

### C. Assumptions

We discuss the convergence behavior of the cascaded RLS-LMS prediction filter in a stationary environment, we make the reasonable assumptions in the following,

**Assumption 2:** The desired predicted signal \( x_k(n) \) and the tap-input signal vector \( x_k(n-1) \) are related function by the multiple linear regression model

\[
x_k(n) = x_k(n-1)h_k(n) + e_{k0}(n)
\]  

(21)

where \( h_k(n) \) is the regression parameter vector and \( e_{k0}(n) \) is the measurement noise.

**Assumption 3:** The input signal vector \( x_k(n-1) \) drawn from a stochastic process, which is ergodic in the autocorrelation function.

The implication of Assumption 3 is that we may substitute time average for ensemble averages. In particular, we may express the ensemble average correlation matrix of the input vector \( x_k(n-1) \) as

\[
R_k \approx \frac{1}{n} \Phi_k(n) \quad \text{for} \quad n > M
\]  

(22)

where \( \Phi_k(n) \) is the time-average correlation of \( x_k(n-1) \) and the requirement \( n > M \) guarantee that the input spreads across all the taps of the filter.

**Assumption 4:** The fluctuations in the weight-error vector \( \| h_k(n) - h_k \| \) are slow compared with those of the input signal vector \( x_k(n) \).

**Assumption 5:** The step-size \( \mu_k, k = 1, \cdots, M \) is sufficiently small such that the inputs to successive stages are at least locally stationary.

### IV. CONDITIONS OF THE CONVERGENCE

Strictly speaking, an adaptive element \( T_k(n) \) converges to the optimal element \( T_k \) (the ‘optimal’ operator) if

\[
\lim_{n \rightarrow \infty} \| T_k(n)e_{k-1}(n) - T_ke_{k-1}(n) \| = 0
\]  

(23)

and \( \mu_k \| x_k(n) \|^2 \leq 1 \). In practice, however, most adaptive algorithms converge only to the values which are within a certain “distance” of the optimal value i.e.,

\[
\lim_{n \rightarrow \infty} \| T_k(n)e_{k-1}(n) - T_ke_{k-1}(n) \| = M_k \leq \infty
\]  

(24)

where \( M_k \) is the (unnormalized) misadjustment. Using (3), the norm in (24) can be written as

\[
\| T_k(n)e_{k-1}(n) - T_ke_{k-1}(n) \|^2 = E\{\| h_k(n) - h_k \|^2\}E\{\| \hat{e}_{k-1}(n) \|^2\}
\]  

(25)

Thus, convergence of the adaptive element is determined by convergence of the filter coefficients.
A. Condition of the Convergence of the RLS Prediction Filter

In order to study the convergence of the RLS algorithm, (17) can be rewritten in terms of a stochastic fixed-point theorem as follows:

\[ \mathbf{h}_k(n) - \mathbf{h}_k = \begin{pmatrix} 1 - K_k(n) \mathbf{x}^T_k(n-1)(\mathbf{h}_k(n-1) - \mathbf{h}_k) \\ + K_k(n)(x_k(n) - \mathbf{x}_k(n-1)\mathbf{h}_k) \end{pmatrix} \]  

(26)

The stochastic operator \( \mathbf{T}_k(n) \) is defined on the space of Gaussian random variables by

\[ \mathbf{T}_k(n)\mathbf{h} \overset{\Delta}{=} [1-K_k(n)\mathbf{x}^T_k(n-1)]\mathbf{h} + K_k(n)[x_k(n) - \mathbf{x}_k(n-1)\mathbf{h}_k] \]

(27)

for all \( \mathbf{h} \). Then (26) can be reformulated as a fixed-point problem as follows:

\[ \mathbf{h}_k(n) - \mathbf{h}_k = \mathbf{T}_k(n)(\mathbf{h}_k(n-1) - \mathbf{h}_k) \]  

(28)

and using the contraction mapping principle of Section (II-C), the recursion in (28) will converge if

\[ \| [1 - K_k(n)\mathbf{x}^T_k(n-1)](y_1 - y_2) \| < \| y_1 - y_2 \| \]

(29)

for \( y_1 \) and \( y_2 \) fixed but arbitrary. Therefore, the condition of convergence in the mean value becomes

\[ E[1 - K_k(n)\mathbf{x}^T_k(n-1)] = 1 - E[\mathbf{\Phi}^{-1}(n)\mathbf{x}_k(n-1)\mathbf{x}^T_k(n-1)] \]

(30)

According to assumption 3, the \( \mathbf{\Phi}^{-1}(n) \) is invertible matrix and is in the form (22), (31) becomes

\[ E[1 - K_k(n)\mathbf{x}^T_k(n-1)] = 1 - \frac{1}{n} \]  

(31)

It means that if \( \mathbf{\Phi}^{-1}(n) \) is invertible matrix, the RLS algorithm converges to the exact solution.

B. Conditions on the Step Size

In order to obtain the conditions on the step-size \( \mu_k \), we employ the stochastic fixed-point theorem to (20), which can be rewritten as follows

\[ \mathbf{h}_k(n) - \mathbf{h}_k = \begin{pmatrix} 1 - \mu_k \mathbf{x}_k(n-1)\mathbf{x}^T_k(n-1) \\ (\mathbf{h}_k(n-1) - \mathbf{h}_k) + \mu_k[x_k(n)\mathbf{x}_k(n-1) - \mathbf{x}_k^T(n-1)\mathbf{x}_k(n-1)\mathbf{h}_k] \end{pmatrix} \]

(32)

The stochastic operator \( \mathbf{T}_k(n) \) is defined on the space of Gaussian random variables by

\[ \mathbf{T}_k(n)\mathbf{h} \overset{\Delta}{=} [1 - \mu_k \mathbf{x}_k(n-1)\mathbf{x}^T_k(n-1)]\mathbf{h} + \mu_k[x_k(n)\mathbf{x}_k(n-1) - \mathbf{x}_k^T(n-1)\mathbf{x}_k(n-1)\mathbf{h}_k] \]

(33)

for all \( \mathbf{h} \). Then (32) can be reformulated as a fixed-point problem as follows:

\[ \mathbf{h}_k(n) - \mathbf{h}_k = \mathbf{T}_k(n)(\mathbf{h}_k(n-1) - \mathbf{h}_k) \]

(34)

and using the contraction mapping principle of Section (II-C), the recursion in (34) will converge if

\[ \| [1 - \mu_k \mathbf{x}_k(n-1)\mathbf{x}^T_k(n-1)](y_1 - y_2) \| < \| y_1 - y_2 \| \]

(35)

for \( y_1 \) and \( y_2 \) fixed but arbitrary. Therefore, the condition of convergence becomes

\[ 0 < \mu_k < \frac{2}{\text{tr}(\mathbf{R}_k)} \]

(36)

where \( \mathbf{R}_k = E[\mathbf{x}_k^T(n-1)\mathbf{x}_k(n-1)] \). The derived result of convergence condition from the fixed-point theorem is the same as the result in the book [11].

V. MISADJUSTMENT

The misadjustment due to the \( k \)th stage in the cascade is defined by (24). Since (25) shows that

\[ M_k^2 = \lim_{n \to \infty} E\{\|\mathbf{h}_k(n) - \mathbf{h}_k\|^2\} \]  

(37)

It is clear that the the normalized misadjustment is nothing but the norm in the fixed-point problem, (25)

\[ m_k^2 = \frac{E\{\|\mathbf{e}_{k-1}(n)\|^2\}}{\lim_{n \to \infty} E\{\|\mathbf{h}_k(n) - \mathbf{h}_k\|^2\}} \]

(38)

The norm of \( E\{\|\mathbf{h}_k(n) - \mathbf{h}_k\|^2\} \) is need for the RLS algorithm and the LMS algorithm. Under the assumptions 2, 3, 4, we can get the mean-squared error in the RLS algorithm [11] if the \( k \)th stage using the RLS:

\[ m_k^2 = \frac{\lim_{n \to \infty} E\{\|\mathbf{h}_k(n) - \mathbf{h}_k\|^2\}}{\frac{M}{n}, n > M} \]

(39)

Similarly, under the assumptions 2, 5, we can obtain the mean-square error in the LMS algorithm [11]:

\[ m_k^2 = \frac{\mu_k}{2} E\{\|\mathbf{x}_k(n)\|^2\} \]

(40)

In order to compute the total misadjustment at the output of the cascade, (12) can be used. Essentially, (12) corresponds to the model shown in Fig.3. The total output power in Fig.3 is bounded by (12), where \( \sigma^2_k \) and \( \mathbf{R}_{k\text{ll}} \) equal to \( m_k^2 \) and \( \mathbf{h}_{k\text{ll}} \), respectively, and we obtain Eq.(41), where \( \|\mathbf{e}_0(n)\|^2 = \|\mathbf{x}(n)\|^2 \). Since the optimal output power is given in (12), the output misadjustment of the last stage normalized with respect to the total power is given by

\[ \prod_{k=1}^{M} \left( m_k^2 + \frac{\|\mathbf{h}_{k\text{ll}}\|^2}{\|\mathbf{h}_{k\text{ll}}\|^2_{\text{min}}} \right)^{-1} \leq m_M^2 \leq \prod_{k=1}^{M} \left( m_k^2 + \frac{\|\mathbf{h}_{k\text{ll}}\|^2}{\|\mathbf{h}_{k\text{ll}}\|^2_{\text{max}}} \right)^{-1} \]

(42)
If the input sequence is very random, \( \| \mathbf{h}_{1k} \|^2_{\text{min}} \approx \| \mathbf{h}_{1k} \|^2_{\text{max}} \approx 1 \) and both bounds approach to the same value in this case. We have approximately

\[
m^2_M \approx \prod_{k=1}^{M} (1 + m^2_k) - 1 \quad (43)
\]

For the LMS algorithm, if \( \mu_k \) is normalized with respect to \( E[\|x_k(n)\|^2] \), which is the case in most application, we have from (40)

\[
m^2_k = \frac{\mu_k}{2} \quad (44)
\]

As the RLS is used in the first stage and the LMS algorithm are used for the rest stages in the cascaded RLS-LMS prediction filter respectively, finally we obtain the final misadjustment of the cascaded RLS-LMS prediction filter:

\[
m^2_M = (1 + \frac{M}{n}) \prod_{k=1}^{M-1} \left[ 1 + \frac{\mu_k}{2} \right] - 1 \quad (45)
\]

We observe that this misadjustment of the cascaded RLS-LMS is slightly higher than the misadjustment of the RLS algorithm [11], but the former posses less complexity than the latter. The exponential dependence on \( M \) for the cascade structure suggests that, in general, a higher misadjustment can be expected for the cascade structure, which will be confirmed by the simulation showed in the Section VI.

VI. SIMULATION RESULTS

The experimental results can be used to verify these theoretical results about the performance of the cascade structure. Considering a stationary process \( x(n) \) generated by filtering unit variance Gaussian white noise through a two-pole filter,

\[
H(z) = \frac{1}{1 - 1.8706z^{-1} + 0.9025z^{-2}} \quad (46)
\]

the filter coefficients are updated using the standard RLS algorithm (20). Fig. 4 illustrates the mean-square errors for the conventional RLS algorithm, the cascaded RLS prediction filter and the cascaded RLS-LMS prediction filter. We observe that the misadjustment of the RLS-LMS is greatest among these three algorithms, but it is very close to the solution of the RLS prediction filter. These results confirm our theoretical analysis.

VII. CONCLUSION

In this paper, we carried out the analysis of the conditions of the convergence and the misadjustment of the cascaded RLS-LMS prediction filter based on a stochastic fixed-point theorem. The theoretical analysis showed that the misadjustment of the cascaded RLS-LMS algorithm is higher than that of the conventional RLS algorithm. However, the complexity of the cascaded RLS-LMS algorithm is lower than the conventional RLS algorithm. It allows us to build up an RLS-like algorithm with time-varying learning rate, which may be useful in uncertain and non-stationary environments.

REFERENCES