On the Convergence of a Direct Adaptive Controller for ATM ABR Congestion Control

Kenneth P. Laberteaux and Charles E. Rohrs
Tellabs Research Center
3740 Edison Lakes Parkway
Mishawaka, IN 46545
{Ken.Laberteaux, Charlie.Rohrs}@tellabs.com

***Abstract* – One of the more challenging and yet unresolved issues which is paramount to the success of ATM networks is that of congestion control for Available Bit Rate (ABR) traffic. Unlike other ATM service categories, ABR provides a feedback mechanism, allowing interior nodes to dictate source rates. Previous work has demonstrated how linear adaptive control theory can be utilized to create a stable and efficient control system for the purposes of ATM ABR congestion control. This paper investigates our recently proposed adaptive controller that uses a finite impulse response (FIR) filter to approximately invert the FIR plant. Specifically, convergence issues are addressed in depth. In doing so, a proof is provided for the convergence of the Normalized Least Mean Square (NLMS) adaptive algorithm employing a DC tap (drift tap), allowing the proof to extend beyond the (often-assumed) zero-mean case. Using a set of reasonable assumptions, parameter convergence in the mean and mean square is proven. Other issues pertaining to the stability of this controller are presented.**

I. Introduction

In 1984, the Consultative Committee on International Telecommunications and Telegraph (CCITT), a United Nations organization responsible for telecommunications standards, selected Asynchronous Transfer Mode (ATM) as the paradigm for broadband integrated service digital networks (B-ISDN) [3]. ATM networks provide 6 service categories. Each category of service is customized for a particular type of traffic. Of these 5 categories, only one, Available Bit Rate (ABR), uses a feedback mechanism to create a closed-loop congestion control. The creation of a control mechanism for a switch that can work with the closed-loop congestion control mechanism specified by the ATM Forum [2] is the focus of the present study.

The complete ABR congestion control mechanism is described in [1] and [3]. This paper limits its consideration to explicit rate congestion control. The plant description of Section II is an approximation to the mechanisms specified in [1].

The present challenge is to devise a controller that resides at the output queue of an ATM switch and produces a single Explicit Rate $u(n)$ to be sent to all ABR sources passing through the queue. The Explicit Rate $u(n)$ must be chosen such that the incoming ABR bandwidth y matches the available ABR bandwidth y^* in some appropriate sense. Specifying a single Explicit Rate at time n for all sources ensures fairness. Matching y to y^* attains efficiency.

Previous contributions to the problem of ATM ABR congestion control include [3]-[6]. In addition, there has been significant contributions made in the ATM Forum [2]. Recently, we proposed a new multi-parameter controller [6]. Specifically, this controller employs an adaptive Finite Impulse Response (FIR) filter to approximately invert the FIR plant model. In this paper, we analyze the convergence and stability of the controller proposed in [6]. It will be shown that the controller parameters converge to their optimal values in both the mean and the mean square sense. Further, the form of the controller ensures stability.

The convergence analysis in this paper is based on a proof of convergence for the NLMS algorithm by Tarrab and Feuer [8]. However we make different assumptions (Section II.B); most notably, we do not require zero-mean signals, which were assumed in [8]. Further, our filter includes a DC tap (drift tap) that ensures the mean of the estimated signal equals the mean of the signal being estimated.

Due to the page limitations, many technical details have been omitted here. A full presentation can be found in [7].

The remainder of this paper is organized as follows: Section II defines the plant and the controller under consideration and derives the optimal parameter values for the controller. Section III consists of a proof that the controller parameters converge to their optimal value in the mean. Mean square convergence is shown in Section IV. Section V discusses the global stability of the system. Conclusions are made in Section VI.

II. Problem Statement

In this section, the plant and controller are defined.

A. Plant and Controller

The plant and controller under consideration were presented in [6] and are briefly summarized here. Since each switch implements its own, independent controller, one may consider the plant from the perspective of a single switch *SW*. Port *j* of switch *SW* carries *N* simultaneous Available Bit Rate (ABR) sessions. To be fair to all connections, port *j* generates a single desired rate $u(n)$ for all connections. Rates $u(n)$, $y(n)$, and $y^*(n)$ are in units of cells/second.

Output port *j* will observe changes to its input rate $y(n)$ as various sources (S_i) react to previously specified Explicit Rates $u(n\text{-}m)$. These reaction delays will vary for different sources. In addition, one or more of the N flows may be unresponsive to $u(n)$. The part of $y(n)$ comprised of non-responsive flows is C cells/second. It is assumed that C, b_0, $b_1, \ldots, b_{dB}$ remain constant for periods of time long enough for adaptive identification to occur. Faster convergence speed of the adaptive algorithm results in better tracking of these time-varying parameters. More details are in [6].

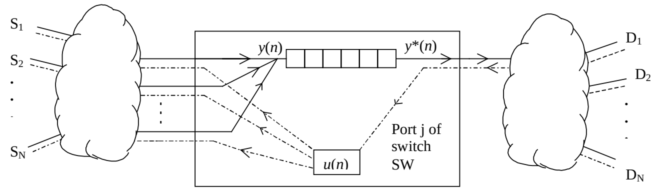

Figure 1 – Plant from perspective of Switch

The plant is therefore given by

$$y(n) = b_0 u(n-d) + \cdots + b_{dB} u(n-d-dB) + C \quad (1)$$

$$y(n) = B(z^{-1})u(n-d) + C \quad (2)$$

Note that for convenience, filters in z^{-1} and time sequences in n will be mixed in expressions such as (2); (2) is equivalent to (1).

The controller $\hat{\mathbf{Q}}(n)$ is comprised of an adaptive FIR filter with a DC tap, with the desired future input rate **y***(n+d+V|n) (see [6]) as its input and the Explicit Rate u(n) as its output.

$$u(n) = \hat{\mathbf{Q}}(n)^T \mathbf{y}^*(n+d+V) \quad (3)$$

$$\hat{\mathbf{Q}}(n) = [\hat{q}_0(n), \hat{q}_1(n), ..., \hat{q}_{dQ}(n), \hat{q}_{DC}(n)]^T$$

$$\mathbf{y}^*(n+d+V \mid n) \equiv [y^*(n+d+V \mid n), ...,$$
$$y^*(n+d+V-dQ \mid n-dQ), y_{DC}]^T$$

Identification of the controller employs Normalized Least Mean Square (NLMS) [9]:

$$\hat{u}(n-d-V) = \hat{\mathbf{Q}}(n)^T \mathbf{y}(n) \quad (4)$$

$$\mathbf{y}(n) = [y(n), y(n-1), ..., y(n-dQ), y_{DC}]^T \quad (5)$$

$$e(n) \equiv e_u(n-V-d) \equiv u(n-d-V) - \hat{u}(n-d-V) \quad (6)$$

$$\hat{\mathbf{Q}}(n+1) = \hat{\mathbf{Q}}(n) + \frac{\mu \mathbf{y}(n)}{\mathbf{y}(n)^T \mathbf{y}(n)} e(n) \quad (7)$$

d is the minimum plant delay, V is an operator chosen (non-negative) inversion polynomial delay (discussed at length in [6]), and μ is the adaptive gain chosen such that 0 < μ < 2. The constant y_{DC} is operator-chosen, appended to the delay-chain values of {y} in (5) so that the final tap of $\hat{\mathbf{Q}}(n)$ becomes a DC tap $\hat{q}_{DC}(n)$ (discussed further in Section II.C),

$$\hat{\mathbf{Q}}(n) \equiv [\hat{\mathbf{Q}}_{lin}(n)^T, \hat{q}_{DC}(n)]^T . \quad (8)$$

For notational convenience, define another vector, identical to (5) except for the DC term:

$$\mathbb{Y}(n) \equiv [y(n), y(n-1), ..., y(n-dQ)]^T \quad (9)$$

The complete system under consideration, as presented in [6], is shown in Figure 2. Plant (2) is controlled by Controller (3). The Controller is identified with (4)-(7).

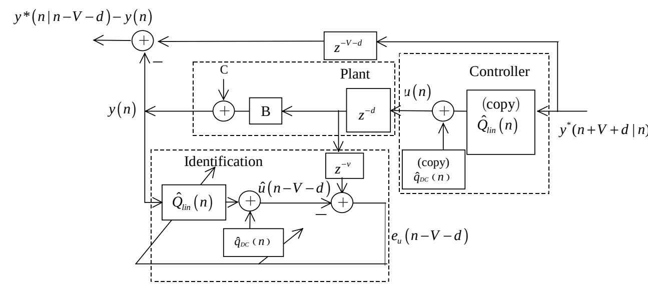

Figure 2 – Complete System ($y_{DC} = 1$)

The system will operate optimally if $\hat{\mathbf{Q}}(n)$ produces u(n) that minimizes $E[(y(n+d+V) - y^*(n+d+V \mid n))^2]$, which occurs if $\hat{\mathbf{Q}}(n)$ converges to its stationary, minimum mean square error optimal value, $\mathbf{Q}_0$ (see (10) and Lemma 1). In this paper, we prove that $\hat{\mathbf{Q}}(n)$ converges to $\mathbf{Q}_0$. To quantify the amount of convergence at time n, define the parameter error vector $\tilde{\mathbf{Q}}(n) = \hat{\mathbf{Q}}(n) - \mathbf{Q}_0$. Given certain assumptions (Section II.B, plus 0<μ<2), it is shown that $\lim_{n\to\infty} E[\tilde{\mathbf{Q}}(n)] = \mathbf{0}$ and $\lim_{n\to\infty} \sum_{i=1}^{dQ+1} E[(\tilde{Q}_i(n))^2] < \alpha$, $\alpha < \infty$.

B. Assumptions

Here are the assumptions made:

Assumption 1: u(n) is gaussian.

Assumption 2: $\mathbf{y}(n)$ *and* $\tilde{\mathbf{Q}}(n)$ *are independent. Also* $u(n-V-d)$ *and* $\tilde{\mathbf{Q}}(n)$ *are independent.*

Assumption 3: The auto-covariance matrix, $\boldsymbol{\sigma}^2 \equiv E[(\mathbb{Y}(n) - E[\mathbb{Y}(n)])(\mathbb{Y}(n) - E[\mathbb{Y}(n)])^T]$*, is full rank.*

Assumption 4: $\alpha_0 \le \|\mathbf{y}(n)\|^2 < \alpha_1, \ \alpha_0 > 0, \ \alpha_1 < \infty$

Assumption 4 ensures that finite adaptation adjustments in (7) will occur. In implementation, it is common to impose Assumption 4 by simply skipping the adaptation of (7) unless Assumption 4 is satisfied.

Assumption 3 is a sufficient excitation condition. From Assumption 3, it follows that $E[\mathbf{y}(n)\mathbf{y}(n)^T]$ is full rank [7], which ensures that the plant will be fully identified, allowing the discovery of a unique $\mathbf{Q}_0$; see (10).

Assumption 2 is an often-made assumption in convergence proofs. If {y} were white (an assumption we generally do not make, but do here only for illustration), both $u(n-V-d)$ and **y**(n) would be independent of $\tilde{\mathbf{Q}}(n-dQ)$, and if μ <<1, then $\tilde{\mathbf{Q}}(n) \approx \tilde{\mathbf{Q}}(n-dQ)$. Signals $u(n-V-d)$ and **y**(n) make their most significant contribution to $\tilde{\mathbf{Q}}(n+1)$. For ease of computations, the much smaller impact on $\tilde{\mathbf{Q}}(n)$ is ignored.

Assumption 2 replaces an assumption made in [8]. The proof of [8] assumes that the excitation signal is Gaussian and further, if $\mathbf{y}(n)$ is a vector of the excitation signal at time n, $E[\mathbf{y}(n)\mathbf{y}(m)^T]=\mathbf{0}$ for $n \neq m$, even if $m = n+1$. This assumption tends not to be even approximately true if $\mathbf{y}(n)$ is the input of an FIR filter, thus it is replaced with Assumption 2.

Assumption 1 assures that $u(n\text{-}d)$ and $\mathbf{y}(n)$ are jointly Gaussian. It is significant to note that [8] further requires $\mathbf{y}(n)$ to be zero-mean. No such assumption is made here. Broadening [8] beyond the zero-mean case is the primary contribution of this paper.

C. DC Identification

Consider the Identification section of Figure 2 ((4) – (7)), redrawn as Figure 3.

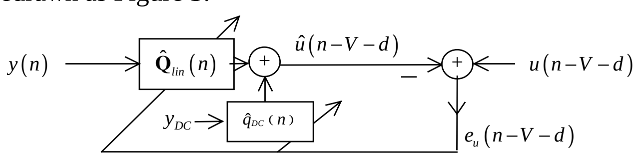

Figure 3– Adaptive System for Calculating $\hat{\mathbf{Q}}(n)$

Recalling (8), from Figure 3, the optimal solution $\mathbf{Q}_0$ for the adaptive coefficients $\hat{\mathbf{Q}}(n)$ is defined as $\mathbf{Q}_0 = \arg\min_{\hat{\mathbf{Q}}}\{e(n)^2\}$. Defining $\mathbf{R} \equiv E[\mathbf{y}(n)\mathbf{y}(n)^T]$, $\boldsymbol{\rho} \equiv E[\mathbf{y}(n)u(n-V-d)]$, $\mathbf{Q}_0$ is known to be [9]

$$\mathbf{Q}_0 = \mathbf{R}^{-1}\boldsymbol{\rho}\,. \tag{10}$$

This solution exists and is unique since $\mathbf{R}$ is full rank [7].

Now consider a different yet similar scheme where the DC tap is not employed but the means of y and u are removed, as in Figure 4.

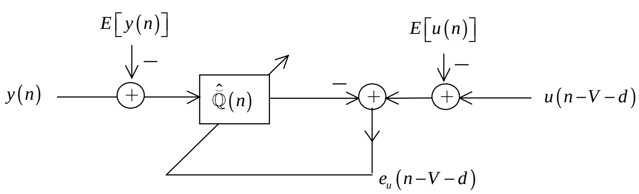

Figure 4 – Adaptive System with Means Explicitly Removed

Define $\breve{\mathbb{Y}}(n) \equiv \mathbb{Y}(n) - E[\mathbb{Y}(n)]$ and $\breve{u}(n) \equiv u(n) - E[u(n)]$. The optimal solution for the adaptive coefficients $\hat{\breve{\mathbb{Q}}}(n)$ is $\breve{\mathbb{Q}}_0$, which solves $E[\breve{\mathbb{Y}}(n)\breve{\mathbb{Y}}(n)^T]\breve{\mathbb{Q}}_0 = E[\breve{\mathbb{Y}}(n)\breve{u}(n-V-d)]$.

The solutions $\breve{\mathbb{Q}}_0$ and $\mathbf{Q}_0$ are closely related, as shown by Lemma 1.

Lemma 1 – The unique solution of (10) is

$$\mathbf{Q}_0 = \left[\breve{\mathbb{Q}}_0^{\,T}, \frac{-E[y(n)]\sum_i \breve{\mathbb{Q}}_{0,i} + E[u(n-V-d)]}{y_{DC}}\right]^T.$$

Proof: As noted before, $\mathbf{Q}_0$ is unique due to $\mathbf{R}$ being full-rank [7]. Describe

$$\left[\breve{\mathbb{Q}}_0^{\,T}, \frac{-E[y(n)]\sum_i \breve{\mathbb{Q}}_{0,i} + E[u(n-V-d)]}{y_{DC}}\right]^T \tag{11}$$

as the *proposed solution* to (10). Directly substituting the proposed solution into (10) verifies (after some algebraic manipulation) that it is indeed a solution, and thus the unique solution, completing the proof.

Lemma 1 demonstrates that by using a DC tap in the adaptive estimator as in Figure 3, the optimal solution for $\mathbf{Q}_{lin,0}$ is equivalent to $\breve{\mathbb{Q}}_0$. To gain intuition, consider that for a general linear estimator, i.e. not using a DC tap, the optimal solution would create the best possible match between the frequency spectrum of the desired signal (u) and the spectrum of the estimated signal ($\hat{u}$), given the regressor (y), across all frequencies, including DC. A DC tap, if included, can only affect the spectrum of the estimated signal at DC, but by doing so, allows the linear taps to ignore DC in their spectrum matching, as if there was no DC content in either regressor signal (y) or the desired signal (u).

The DC tap creates an additional similarity between Figure 3 and Figure 4 – a zero-mean optimal error. By defining the optimal error $e^*(n) \equiv u(n-d-V) - \mathbf{Q}_0(n)^T\mathbf{y}(n)$, then with (11), it is easy to show that

$$E[e^*(n)] = 0\,. \tag{12}$$

D. Other Notation

Now that the control and estimation methods have been described, what remains is to show convergence. Before proceeding with the proofs, some notation needs to be introduced. Let the matrices of orthonormalized eigenvectors and eigenvalues of $\mathbf{R}$ be $\mathbf{W}^T$ and $\boldsymbol{\Lambda}$ respectively, where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_i),\ i = 1, \cdots, (dQ+1)$,

$$\mathbf{W}^T\mathbf{W} = \mathbf{I}\,,\ \mathbf{W}\mathbf{R}\mathbf{W}^T = \boldsymbol{\Lambda} \tag{13}$$

Because $\mathbf{R}$ is full-rank, $\mathbf{W}$ is full-rank. Now define a linear transformation of the random vector $\mathbf{y}(n)$ as follows.

$$\boldsymbol{\psi}(n) \equiv \mathbf{W}\mathbf{y}(n),\ \mathbf{W}^T\boldsymbol{\psi}(n) = \mathbf{W}^T\mathbf{W}\mathbf{y}(n) = \mathbf{y}(n) \tag{14}$$

$$\mathbf{L}(n) \equiv \mathbf{W}\tilde{\mathbf{Q}}(n),\ \mathbf{W}^T\mathbf{L}(n) = \tilde{\mathbf{Q}}(n) \tag{15}$$

$$E[\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T] = \boldsymbol{\Lambda} \tag{16}$$

Substituting (4) and (6) into (7), pre-multiplying by $\mathbf{W}$ and adding and subtracting a term, then subtracting $\mathbf{W}\mathbf{Q}_0$ from both sides produces

$$\mathbf{L}(n+1) = \left(I - \mu\frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n)}\right)\mathbf{L}(n) + \frac{\mu\ \boldsymbol{\psi}(n)e^*(n)}{\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n)} \tag{17}$$

and

$$\mathbf{L}(n+1)\mathbf{L}(n+1)^T = \mathbf{L}(n)\mathbf{L}(n)^T - \mu\frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n)}\mathbf{L}(n)\mathbf{L}(n)^T$$
$$-\mu\mathbf{L}(n)\mathbf{L}(n)^T\frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n)} + \mu^2\frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n)}\mathbf{L}(n)\mathbf{L}(n)^T\frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n)}$$
$$+\mu\frac{e^*(n)}{\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n)}[\mathbf{L}(n)\boldsymbol{\psi}(n)^T + \boldsymbol{\psi}(n)\mathbf{L}(n)^T]$$
$$-\mu\frac{e^*(n)}{\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n)}\left[\frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n)}\mathbf{L}(n)\boldsymbol{\psi}(n)^T + \boldsymbol{\psi}(n)\mathbf{L}(n)^T\frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n)}\right]$$
$$+\frac{\mu^2\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T(e^*(n))^2}{(\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n))^2} \quad (18)$$

The following notations will be used extensively:

$$\mathbf{A} \equiv E\left[\frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n)}\right], \mathbf{C}(n) \equiv E[\mathbf{L}(n)\mathbf{L}(n)^T]$$
$$\mathbf{D}(n) \equiv E\left[\frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n)}\ \mathbf{C}(n)\ \frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n)}\right]$$
$$\mathbf{H} \equiv E\left[\frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{(\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n))^2}\right]$$

III. Parameter Convergence in the Mean

In this section it will be shown that $\lim_{n\to\infty} E[\mathbf{L}(n)] = \mathbf{0}$, that in turn, by (15), shows $\lim_{n\to\infty} E[\tilde{\mathbf{Q}}(n)] = \mathbf{0}$.

We begin by noting a few key independencies. From Assumption 2, and the fact that **W** provides a one-to-one mapping, Section 5.4 of [10] shows that $\boldsymbol{\psi}(n)$ and $\mathbf{L}(n)$ are independent. Similarly $u(n-V-d)$ and $\mathbf{L}(n)$ are independent.

Note that $e^*(n)$ and $\mathbf{y}(n)$ are jointly Gaussian, and uncorrelated ($E[\mathbf{y}(n)e^*(n)] = \boldsymbol{\rho} - \mathbf{R}\,\mathbf{R}^{-1}\,\boldsymbol{\rho} = \mathbf{0}$). The auto-covariance matrix of two jointly-Gaussian, uncorrelated random variables where at least one is zero-mean ($e^*(n)$) is diagonal. Therefore, $\mathbf{y}(n)$ and $e^*(n)$ are independent. By a similar argument, $\boldsymbol{\psi}(n)$ and $e^*(n)$ are also independent.

The auto-covariance $\boldsymbol{\psi}(n)$ is diagonal ((13) gives $\mathbf{W}_i^T\mathbf{W}_j = 0$ where $\mathbf{W}_j$ is the j'th row of **W**), thus the elements of $\boldsymbol{\psi}(n)$ are independent.

Lemma 2 – $\mathbf{A}$ and $\mathbf{H}$ are diagonal matrices.

Proof: The proofs for $\mathbf{A}$ and $\mathbf{H}$ are nearly identical; only the former will be shown. Let $\mathbf{A}_{ij}$ indicate the element of $\mathbf{A}$ in the i'th row, j'th column. Then

$$|\mathbf{A}_{ij}| = \left|E\left[\frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n)}\right]_{ij}\right| = \left|\int_{\mathbf{X}}\frac{x_i x_j}{\|\mathbf{X}\|^2}\ f_{\psi}(\mathbf{X})\,d\mathbf{X}\right|$$

From Assumption 4,

$$|\mathbf{A}_{ij}| \le \frac{1}{\alpha_0}\left|\int_{\mathbf{X}} x_i x_j\ f_{\psi}(\mathbf{X})\,d\mathbf{X}\right| = \frac{1}{\alpha_0}|E[\psi_i(n)\psi_j(n)]|$$

From (16),

$$|A_{ij}| \le \begin{cases} 0 \text{ if } i \ne j \\ \dfrac{\lambda_i}{\alpha_0} \text{ if } i = j \end{cases} \quad (19)$$

Thus all non-diagonal elements of **A** equal zero, completing the proof (note that **A** is positive definite and thus all of its eigenvalues λ_i are positive).

Lemma 3 – $0 < \lambda_i/\alpha_1 < A_{ii} \le 1$.

Proof: $A_{ii} = E\left[\psi_i(n)^2\left(\sum_{j=1}^{dQ+1}\psi_j(n)^2\right)^{-1}\right]$. Since $\psi_i(n)^2$, $i = 1,\ldots,dQ+1$, are real and non-negative, $0 \le A_{ii} \le 1$. However, from Assumption 4

$$A_{ii} > \frac{1}{\alpha_1}\int_{\mathbf{X}} x_i^2\ f_{\psi}(\mathbf{X})\,d\mathbf{X} = \frac{1}{\alpha_1}E[\psi_i(n)^2] = \frac{\lambda_i}{\alpha_1}$$

Also, since **R** is full-rank, the eigenvalues λ_i of **R** are non-zero. Thus $0 < \lambda_i/\alpha_1 < A_{ii} \le 1$, completing the proof.

Now we reach the main result of this Section.

Theorem 1 *- Given Assumption 1– Assumption 4 and* $0 < \mu < 2$, $\lim_{n\to\infty} E[\tilde{\mathbf{Q}}(n)] = 0$.

<u>Proof</u>: Consider (17). From (12),

$$E\left[\frac{\mu\boldsymbol{\psi}(n)e^*(n)}{\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n)}\right] = \mu\ E\left[\frac{\boldsymbol{\psi}(n)}{\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n)}\right]E[e^*(n)] = 0$$

Since $\boldsymbol{\psi}(n)$ and $\mathbf{L}(n)$ are independent:

$$E[\mathbf{L}(n+1)] = (\mathbf{I} - \mu\ \mathbf{A})\ E[\mathbf{L}(n)]$$

From Lemma 2, the linear system completely decouples:

$$E[L_i(n+1)] = (1 - \mu A_{ii})E[L_i(n)],\ i = 1, \ldots, dQ+1$$

From Lemma 3 and if $0 < \mu < 2$, then $|(1-\mu A_{ii})| < 1$, $i = 1, \ldots, dQ+1$. Thus, $\lim_{n\to\infty} E[\mathbf{L}(n)] = \mathbf{0}$. Eq. (15) gives $\lim_{n\to\infty} E[\tilde{\mathbf{Q}}(n)] = \mathbf{0}$, thus completing the proof.

Theorem 1 states that given Assumption 1–Assumption 4, if μ is bounded as $0 < \mu < 2$, then our NLMS adaptive system of estimating $\hat{\mathbf{Q}}$, given by (4) – (7), converges in the mean to the ideal $\mathbf{Q}_0$, given by Lemma 1. It is distinctive from the proof of [8] in that it uses (12) – instead of a zero-mean assumption for $\mathbf{y}(n)$ – to eliminate the expectation of the second term of (17).

IV. Parameter Convergence in the Mean Square

Now that Theorem 1 has been presented, a statement bounding the variance of $\tilde{\mathbf{Q}}(n)$ would give much more credibility to the proposed controller, and is the goal of this section.

The proof in [8] relies heavily on a zero-mean assumption on $\mathbf{y}(n)$, an assumption not made here. However, Lemma 4 shows that almost all of the terms of $\boldsymbol{\psi}(n)$ are zero-mean. Because of this, we can adopt a strategy similar to that of [8].

Lemma 4 – Given the independence of the terms of $\boldsymbol{\psi}(n)$*, (16) is the necessary and sufficient condition that no less than dQ of the dQ+1 elements of* $\boldsymbol{\psi}(n)$ *are zero mean.*

Proof: Sufficiency: By contradiction. For $i \neq j$, $E[\psi_i(n)\psi_j(n)] = E[\psi_i(n)]E[\psi_j(n)]$. If less than dQ elements of $\boldsymbol{\psi}(n)$ are zero mean, $E[\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T]$ cannot be diagonal, contradicting (16). Necessity: If no less than dQ of the $dQ+1$ elements of $\boldsymbol{\psi}(n)$ are zero mean, then the independence of the terms of $\boldsymbol{\psi}(n)$ gives (16), concluding the proof.

For convenience, we shall name the element of $\boldsymbol{\psi}(n)$ that generally has non-zero mean $\psi_\zeta(n)$, i.e.

$$E[\psi_i(n)] = 0, i = 1,\ldots,dQ+1,\ i \neq \zeta, \tag{20}$$

noting that if $E[y(n)] = 0$, $E[\psi_\zeta(n)] = 0$.

Now we begin the mean-square convergence proof.

Lemma 5 – The expectation of the fifth through eighth term of (18) is zero.

Proof: Examine the expectation of the seventh term on the right side of (18), specifically the term of row *i*, column *j*:

$$\mu E\left[\frac{\sum_{k=1}^{dQ+1} L_k(n)e^*(n)\psi_i(n)\psi_j(n)\psi_k(n)}{(\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n))^2}\right] \tag{21}$$

For the *k*'th term of the summation, since $L_k(n)$ is independent from $\boldsymbol{\psi}(n)$, and $e^*(n)$ is independent from $\boldsymbol{\psi}(n)$, $E[e^*(n)L_k(n)]$ can be separated from the remaining terms inside the expectation of (21). Now, from (15), Assumption 2, and (12):

$E[e^*(n)L_k(n)] = \mathbf{W}_k E[e^*(n)]E[\tilde{\mathbf{Q}}(n)] = 0$, for all *k*,

which shows that the expectation of row *i*, column *j* of the seventh term equals zero. Since this is true for every (i,j), the expectation of the seventh term results in a matrix of zeros. By a similar argument, one can show the expectations of the fifth, sixth, and eighth terms of (18) produce matrices of zeros, thus completing the proof.

With Lemma 5, (18) is equivalent to

$$\begin{aligned}\mathbf{C}(n+1) = \mathbf{C}(n) - \mu(\mathbf{AC}(n) + \mathbf{C}(n)\mathbf{A}) \\ +\mu^2\mathbf{D}(n) + \mu^2\varepsilon^*\mathbf{H}\end{aligned} \tag{22}$$

where ε^* is the minimal mean-square error $\varepsilon^* \equiv E[(e^*(n))^2]$.

Since (22) is a discrete, linear, time invariant difference equation, its convergence is guaranteed if its homogeneous part is asymptotically stable, as this implies BIBO stability, and if its forcing term $\mu^2\varepsilon^*\mathbf{H}$ is bounded. We now show these in turn.

Lemma 6 – Since the elements of $\boldsymbol{\psi}(n)$ *are Gaussian and independent, and given (20), the expectation of an Odd Function of* $\boldsymbol{\psi}(n)$ *around* $\psi_i(n)$[1] *for* $i \neq \zeta$ *is zero.*

Proof: Mentioned in [8] (using (20)). Shown in [7].

We now show that $D_{ij}(n)$, the row *i*, column *j* element of $\mathbf{D}(n)$, can be computed as

$$D_{ij}(n) = 2G_{ij}C_{ij},\ i \neq j \tag{23}$$

$$D_{ii}(n) = \sum_{J=1}^{dQ+1} G_{iJ}C_{JJ} \tag{24}$$

where $\mathbf{G}$ is defined as

$$G_{ij} \equiv E\left[\frac{(\psi_i(n))^2(\psi_j(n))^2}{(\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n))^2}\right] \tag{25}$$

By direct substitution, we have

$$D_{ij}(n) = \sum_{K=1}^{dQ+1}\sum_{J=1}^{dQ+1} E\left[\frac{\psi_i(n)\psi_j(n)\psi_J(n)\psi_K(n)}{(\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n))^2}\right]C_{JK} \tag{26}$$

Since the elements of $\boldsymbol{\psi}(n)$ are independent and all but one are zero-mean, the numerator of the expectation in (26) will equal zero in many cases. To show which terms of the double summation of (26) equal zero, note that for any (i,j,J,K) the numerator can always be expressed as

$$\begin{aligned}&E[\psi_i(n)\psi_j(n)\psi_J(n)\psi_K(n)] = \\ &E[(\psi_{\pi_1}(n))^{\rho_1}]E[(\psi_{\pi_2}(n))^{\rho_2}]E[(\psi_{\pi_3}(n))^{\rho_3}]E[(\psi_{\pi_4}(n))^{\rho_4}]\end{aligned} \tag{27}$$

such that $\rho_1,\rho_2,\rho_3,\rho_4 \in \{0,1,2,3,4\}$, $\rho_1+\rho_2+\rho_3+\rho_4 = 4$, $\rho_1 \geq \rho_2 \geq \rho_3 \geq \rho_4$, $\pi_1,\pi_2,\pi_3,\pi_4 \in \{1,\ldots,dQ+1\}$, and $\pi_\beta \neq \pi_\delta$ for $\beta \neq \delta$. Then each (i,j,J,K) maps to exactly one of the following five cases:

Case 1: $\rho_1 = 1, \rho_2 = 1, \rho_3 = 1, \rho_4 = 1$

Case 2: $\rho_1 = 2, \rho_2 = 1, \rho_3 = 1, \rho_4 = 0$

Case 3: $\rho_1 = 2, \rho_2 = 2, \rho_3 = 0, \rho_4 = 0$

Case 4: $\rho_1 = 3, \rho_2 = 1, \rho_3 = 0, \rho_4 = 0$

Case 5: $\rho_1 = 4, \rho_2 = 0, \rho_3 = 0, \rho_4 = 0$

Remembering that the element not assured to be zero-mean is denoted as $\psi_\zeta(n)$, (see (20)), then for Case 1, at most one member of the set $\{\pi_1,\pi_2,\pi_3,\pi_4\}$ equals ζ, thus (27) would equal zero. Similarly for Case 2, at most one member of the set $\{\pi_1,\pi_2,\pi_3\}$ equals ζ, thus (27) would equal zero.

For Case 4, if $\pi_1 = \zeta$, then from (20), (27) would equal zero. If $\pi_2 = \zeta$, then by Lemma 6, (27) would equal zero. If

[1] A function $\Gamma_i(\boldsymbol{\psi}(n))$, is defined as an *Odd Function around* $\psi_i(n)$ if $\Gamma_i(\psi_1(n),\ldots,\psi_i(n),\ldots,\psi_{dQ+1}(n)) = -\Gamma_i(\psi_1(n),\ldots,-\psi_i(n),\ldots,\psi_{dQ+1}(n))$

$\pi_1, \pi_2 \neq \zeta$, then by either (20) or Lemma 6, (27) would equal zero, thus (27) would equal zero for the entirety of Case 4.

For Cases 3 and 5, (27) could be non-zero.

Using an argument similar to that found in the proof of Lemma 2, the instances of (i, j, J, K) that fall into Cases 1, 2, and 4 can be shown to have a zero contribution to the double summation of (26). For the remaining cases, consider first the instances where $i \neq j$, which eliminates Case 5 in addition to Cases 1, 2, and 4, thus (26) equals (23). For $i = j$, after eliminating Cases 1, 2 and 4, Cases 3 and 5 make (26) equal to (24).

Having shown (22), (23), and (24), the techniques used for remainder of the mean-square proof are nearly identical to those presented in [8], and thus will only be outlined here (see [7] for details). Off-diagonal elements of $\mathbf{C}(n)$ are treated separately from the diagonal elements. The off-diagonal term of (22) is

$$C_{ij}(n+1) = \gamma_{ij} C_{ij}(n), i \neq j \tag{28}$$

$\gamma_{ij} = 1 - \mu(A_{ii} + A_{jj}) + 2\mu^2 G_{ij}, i \neq j$, with $|\gamma_{ij}| < 1$ ([8],[7]), and thus (28) goes to zero as n approaches infinity.

Focussing now on the diagonal entries of $\mathbf{C}(n)$, define a vector of the diagonal entries of $\mathbf{C}(n)$, $\mathbf{\Omega}(n) \equiv [C_{11}(n), C_{22}(n), \ldots, C_{(dQ+1)(dQ+1)}(n)]^T$. From (22),

$$\mathbf{\Omega}(n+1) = \mathbf{F}\mathbf{\Omega}(n) + \mu^2 \varepsilon * \mathbf{h}, \tag{29}$$

$\mathbf{F} = \text{diag}\{(1 - 2\mu \mathbf{A}_{ii})\} + \mu^2 \mathbf{G}$, $\mathbf{h} \equiv [\mathbf{H}_{11}, \mathbf{H}_{22}, \ldots, \mathbf{H}_{(dQ+1)}]^T$. It can be shown ([8],[7]) that (29) is BIBO stable. Assuming $0 < \mu < 2$, the forcing term of (29) is bounded, i.e. $|\mu^2 \varepsilon * \mathbf{H}_{ii}| < \alpha_2, \alpha_2 < \infty$, $|\mu^2 \varepsilon * \mathbf{H}_{ij}| = 0, i \neq j$, since $|H_{ii}| \leq 1/(dQ-2)\lambda_{\min}$ ([8],[7]). Thus we can state the main result of this section:

Theorem 2*: Given Assumption 1– Assumption 4 and*

$$0 < \mu < 2, \lim_{n\to\infty} \sum_{i=1}^{dQ+1} E[(\tilde{Q}_i(n))^2] < \alpha, \alpha < \infty.$$

Proof: Linear time-invariant system (29) is BIBO stable ([8],[7]). Its input signal is bounded, thus $\lim_{n\to\infty} \mathbf{\Omega}(n) < \alpha_3 [1,1,\ldots,1]^T, \alpha_3 < \infty$. Then with (28) and $|\gamma_{ij}| < 1$, $\lim_{n\to\infty} \mathbf{C}(n) < \alpha_3 \mathbf{I}$, $\lim_{n\to\infty} tr\, \mathbf{C}(n) < \alpha, \alpha < \infty$. Since $tr\, \mathbf{C}(n) = tr\, E[\tilde{\mathbf{Q}}(n)\tilde{\mathbf{Q}}(n)^T]$, the proof is completed.

The key contribution of Section IV comes from our showing (22)-(24) without requiring $E[\mathbf{y}(n)]=\mathbf{0}$. Expressions (22)-(24) exist in [8], but required $E[\mathbf{y}(n)]=\mathbf{0}$. By demonstrating (22)-(24) without requiring $E[\mathbf{y}(n)]=\mathbf{0}$, the results of [8] are significantly extended.

Equations (28) and (29) actually show more than stated in Theorem 2. They show that each element of $\mathbf{C}(n)$ is bounded at each n, that the off-diagonal elements of $\mathbf{C}(n)$ converge to zero, and that the diagonal elements also converge, $\lim_{n\to\infty} \mathbf{\Omega}(n) = (\mathbf{I} - \mathbf{F})^{-1} \mu^2 \varepsilon * \mathbf{h}$.

V. Global Stability

Global Stability has been built into the control structure. Both plant (1) and controller (3) are FIR filters. The controller simply conditions the set point $\{y^*\}$; all other control is open loop. The parameters of the plant are obviously bounded. The parameters of the controller are random variables that have been shown to have mean square values that are finite for all n and converge, implying a bounded mean-square gain for the controller. From an implementation view, the controller parameters can be kept bounded at each time n by a simple limiter after the adaptation of (7). The FIR structure of the controller then guarantees BIBO stability for the modified system.

VI. Conclusions

After the congestion control problem was presented, a controller was examined for its convergence and stability properties. Theorem 1 and Theorem 2 proved that the controller parameters converge to their optimal values in the mean and mean square sense, respectively. It was observed that the controller runs essentially open loop, only conditioning the set point, thus global convergence is assured.

References

[1] J. Kenney, Editor, *Traffic Management Specification Version 4.1*, available from [2].

[2] ATM Forum web site, http://www.atmforum.org.

[3] R. Jain, "Congestion Control and Traffic Management in ATM Networks: Recent Advances and A Survey," *Comp Net ISDN Sys*, Vol. 28, No. 13, Oct. 1996, pp. 1723-1738.

[4] C. Rohrs, R. Berry, and S. O'Halek, "Control engineer's look at ATM congestion avoidance," *Comp Comm*. v 19 n 3 Mar 1996. pp. 226-234.

[5] K. Laberteaux and C. Rohrs, "Application Of Adaptive Control To ATM ABR Congestion Control," *Globecom'98*.

[6] K. Laberteaux and C. Rohrs, "A Direct Adaptive Controller for ATM ABR Congestion Control," Amer Control Conf 2000, Chicago, IL, June 2000.

[7] K. Laberteaux and C. Rohrs, "A Proof of Convergence for a Direct Adaptive Controller for ATM ABR Congestion Control," ND-ISIS-2000-02, available from http://www.nd.edu/~isis.

[8] M. Tarrab and A. Feuer, "Convergence and Performance Analysis of the Normalized LMS Algorithm with Uncorrelated Gaussian Data," *Trans Info Theory*, Vol. 34, No. 4, July 1988, p. 680-691.

[9] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 1991.

[10] P. Peebles, *Probability, Random Variables, and Random Signal Principles, 2nd Ed.* McGraw-Hill, New York, NY, 1987.