A Proof of Convergence for a Direct Adaptive Controller for ATM ABR Congestion Control

Technical Report of the ISIS Group
At the University of Notre Dame
ISIS-2000-002
September, 2000

Kenneth P. Laberteaux and Charles E. Rohrs
Tellabs Research Center
3740 Edison Lakes Parkway, Mishawaka, IN 46545
{Ken.Laberteaux, Charlie.Rohrs}@tellabs.com
219.258.6400 (voice), 219.258.6446 (fax)

Interdisciplinary Studies of Intelligent Control

A Proof of Convergence for 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
219.258.6400 (voice), 219.258.6446 (fax)

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 most ATM traffic classes, ABR provides a feedback mechanism, allowing interior nodes to dictate source rates. Previous work has demonstrated that how linear control theory can be utilized to create a stable and efficient control system for the purposes of ATM ABR congestion control. This paper investigates one recently proposed direct adaptive controller. Specifically, convergence issues are addressed in depth. 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.

1 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 congestion control mechanism is described in [1] and [3]. This paper limits its consideration to explicit rate congestion control. The plant description of Section 2.1 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]-[11]. In addition, there has been significant contributions made in the ATM Forum [2]. Recently, we proposed a new multi-parameter controller [11]. 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 [11]. 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 [12]. However we make different assumptions (Section 2.2); most notably, we do not require zero-mean signals, which were assumed in

[12]. 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.

The remainder of this paper is organized as follows: Section 2 defines the plant and the controller under consideration and derives the optimal parameter values for the controller. Section 3 consists of a proof to show that the controller parameters converge to their optimal value in the mean. Section 4 consists of a proof to show that the controller parameters converge to their optimal value in the mean square. Section 5 discusses the global stability of the system. Conclusions are made in Section 6.

This technical report contains many detailed proofs. Readers may prefer the treatments of this material in either [17] or [18].

2 Problem Statement

In this section, the plant and controller are defined.

2.1 Plant Definition

Since each switch implements its own independent controller, one may consider the plant from the perspective of a single switch *SW*, as in Figure 1. A discrete-time model is used, where sample intervals correspond to control intervals, i.e. a new control action $u(n)$ is calculated for each n. Port j of switch *SW* carries N simultaneous Available Bit Rate (ABR) sessions, and serves as output port for data cells and input port for Backward Resource Management Cells.

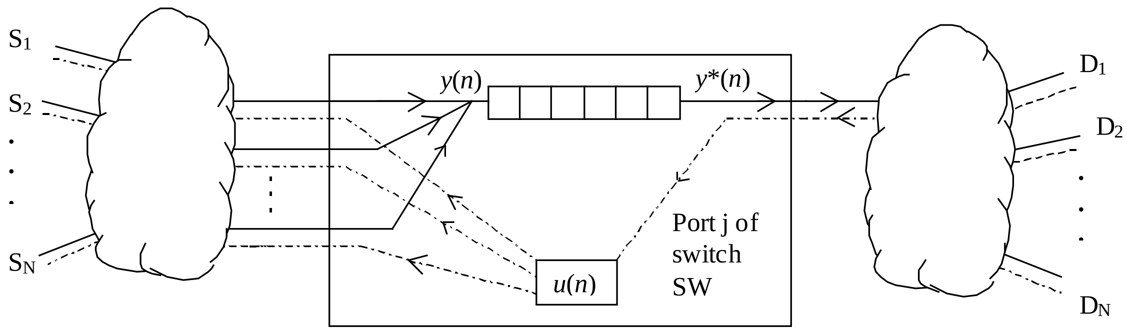

Figure 1 – Plant from perspective of Switch

The present challenge is to devise a controller that resides at output Port *j* of switch *SW* and produces a single Explicit Rate u to be sent to all ABR sources passing through the port. The Explicit Rate u 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.

Port *j* generates a single desired rate $u(n)$ for all connections. As Resource Management (RM) cells for the *N* ABR Virtual Connections (VC's) pass through *j* on their return from destination to source, Port *j* examines each, specifically the contents of each Explicit Rate (ER) Field. If Port *j* finds the ER Field contains a rate above its current $u(n)$, Port *j* overwrites the ER Field with $u(n)$. The RM cell transports this Explicit Rate $u(n)$ to each ABR Source. It is assumed that for each of the *N* ABR Virtual Connections, at least one RM cell passes *j* during each sample interval. Rates $u(n)$, $y(n)$, and $y*(n)$ are in units of cells/second.

Although N sources share $y*(n)$ of available bandwidth, it is assumed that a subset N_c of the N sources are constrained to a rate different than $u(n)$. There are at least two reasons for this possibility. First, a source may be controlled or bottlenecked by another switch along its path. Second, a source may have been guaranteed a Minimum Cell Rate (MCR) greater than the rate assigned by Port *j*, or have insufficient data to take advantage of the offered bandwidth. Thus only the $N_u \equiv N - N_c$ unconstrained sources will react to $u(n)$. These N_u sources are assumed greedy, i.e. will send cells continuously at the maximum Allowed Cell Rate (ACR) dictated by the switch output ports through which they pass. The aggregate bandwidth of the $N_c(n)$ constrained sources

Output Port *j* will observe changes to its input rate $y(n)$ as various sources (S_i) react to previously specified Explicit Rates $u(n-m)$. The *reaction delay*, *m*, as viewed by *j* for source S_i, is the time between *j*'s adjustment of its explicit rate at time $n-m$ to the time *j* measures this explicit rate as its input rate from S_i. These reaction delays will vary for different sources. Assume that there are b_0 sources that respond with reaction delay *d*, b_1 sources that respond with delay $d+1$, and b_{dB} with delay $d+dB$, where dB is a known upper bound on *j*'s reaction delay. 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. The plant is therefore given by

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

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

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).

Since the minimum delay in the plant is d, adjustments in $u(n)$ will not be observed until $n+d$. Therefore to generate $u(n)$, it must be decided at time n what the desired value of $y(n+d)$ should be. This desired bandwidth, which is notated as $y^*(n+d\,|\,n)$, may reflect both bandwidth and buffer measurements[1] made up to time n (this may be generated by a prediction filter as in [10]). By extension, in many cases, the input of the algorithm will be $y^*(n+d+V\,|\,n)$ (for some non-negative V), i.e. the desired value of $y(n+d+V)$ known at time n. Rates $u(n)$, $y(n)$, and $y^*(n)$ are in units of cells per second.

The goal of the congestion control mechanism of *SW* is to choose at time n the control signal $u(n)$ so as to minimize $E\left[\left(y(n+d+V)-y^*(n+d+V\,|\,n)\right)^2\right]$.

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

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

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

$$\mathbf{y}^*(n+d+V\,|\,n)\equiv\left[y^*(n+d+V\,|\,n),\ldots,y^*(n+d+V-dQ\,|\,n-dQ),y_{DC}\right]^T$$

Identification of the controller employs Normalized Least Mean Square (NLMS) [13]:

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

$$\mathbf{y}(n)=\left[y(n),y(n-1),\ldots,y(n-dQ),y_{DC}\right]^T \tag{5}$$

---

[1] Requested bandwidth can be reduced to shrink the buffer if it is too large.

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

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

d is the minimum plant delay, V is an operator chosen (non-negative) inversion polynomial delay (discussed at length in [11]), and μ is the adaptive gain chosen such that $0 < \mu < 2$. 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 2.3),

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

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

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

The complete system under consideration, as presented in [11], is shown in Figure 2. Plant (2) is controlled by Controller (3). The Controller is identified with (4)-(7).

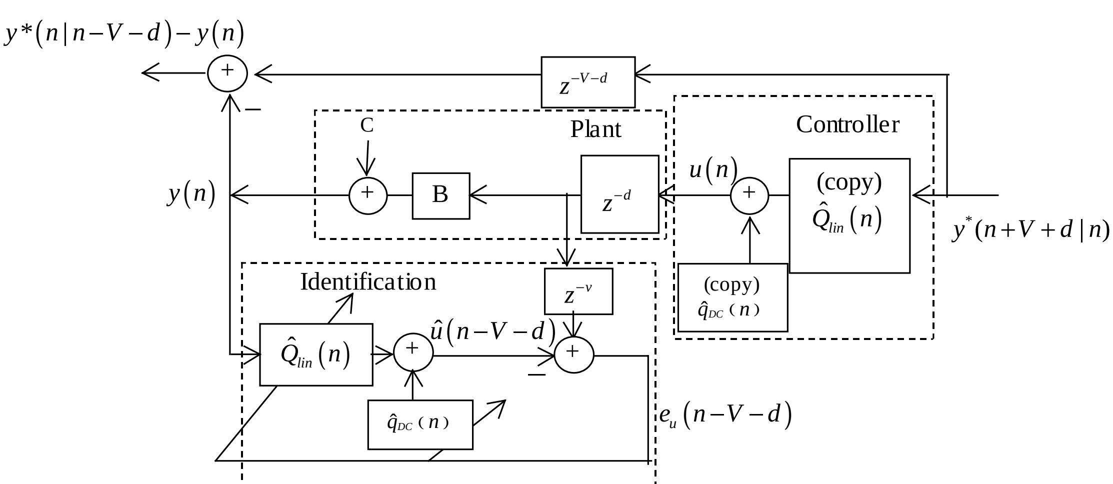

Figure 2 – Complete System ($y_{DC} = 1$)

The system will operate optimally if $\hat{\mathbf{Q}}(n)$ produces, at time n, the control signal $u(n)$ that minimizes $E\left[\left(y(n+d+V)-y^*(n+d+V\,|\,n)\right)^2\right]$, which occurs if $\hat{\mathbf{Q}}(n)$ converges to its stationary, minimum mean square error optimal value, $\mathbf{Q}_0 \equiv \left\{E\left[\mathbf{y}(n)\mathbf{y}(n)^T\right]\right\}^{-1} E\left[\mathbf{y}(n)u(n-d-V)\right]$. This report shows that $\hat{\mathbf{Q}}(n)$ converges to $\mathbf{Q}_0$, and that $y(n)$ appropriately emulates $y^*(n\,|\,n-d-V)$. To quantify the amount of convergence at time *n*, define the parameter error vector $\tilde{\mathbf{Q}}(n) \equiv \hat{\mathbf{Q}}(n) - \mathbf{Q}_0$. Given certain assumptions (Section 2.2, plus 0<μ<2), it is shown that

$$\lim_{n\to\infty} E\left[\tilde{\mathbf{Q}}(n)\right] = \mathbf{0}$$

and

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

2.2 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 $u(n-d-V)$ and $\tilde{\mathbf{Q}}(n)$ are independent.

Assumption 3: $\boldsymbol{\sigma}^2 = E\left[\left(\mathbb{Y}(n) - E\left[\mathbb{Y}(n)\right]\right)\left(\mathbb{Y}(n) - E\left[\mathbb{Y}(n)\right]\right)^T\right]$ is full rank.

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

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

Assumption 3 is a sufficient excitation condition. From Assumption 3, it follows that $E\left[\mathbf{y}(n)\mathbf{y}(n)^T\right]$ is full rank (Lemma 2), which ensures that the plant will be fully identified, allowing the discovery of a unique $\mathbf{Q}_0$; see (13).

Assumption 2 is an often-made assumption in convergence proofs. If $\{y\}$ was white (an assumption we generally do not make, but do here only for illustration), both $u(n-V-d)$ and $\mathbf{y}(n)$ would be independent of $\tilde{\mathbf{Q}}(n-dQ)$, and if $\mu << 1$, then $\tilde{\mathbf{Q}}(n) \approx \tilde{\mathbf{Q}}(n-dQ)$. More generally, signals $u(n-V-d)$ and $\mathbf{y}(n)$ 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 the assumption made in [12]. The proof of [12] 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 makes little sense if $\mathbf{y}(n)$ is the input of an FIR filter, thus it is replaced with Assumption 2.

Assumption 1 assures that $u(n-d)$ and $\mathbf{y}(n)$ are jointly Gaussian. It is significant to note that [12] further requires $\mathbf{y}(n)$ to be zero-mean. No such assumption is made here. Broadening [12] beyond the zero-mean case is the primary contribution of this report.

Lemma 1: $E\left[\mathbb{Y}(n)\mathbb{Y}(n)^T\right]$ is positive definite, thus full rank.

Proof: An auto-correlation matrix, e.g. $\boldsymbol{\sigma}^2$, is positive semi-definite [13], thus for any real, non-zero vector $\mathbf{x}$, $\mathbf{x}^T\boldsymbol{\sigma}^2\mathbf{x} \geq 0$. A necessary and sufficient condition for the positive semi-definiteness of a matrix is that all eignevalues of are ≥ 0. Similarly a necessary and sufficient condition for the positive definiteness of a matrix is that all eignevalues of the matrix are > 0 [16]. From Assumption 3, all eigenvalues of $\boldsymbol{\sigma}^2$ are non-zero, thus $\boldsymbol{\sigma}^2$ is positive definite. Since $E\left[\mathbb{Y}(n)\mathbb{Y}(n)^T\right] = \boldsymbol{\sigma}^2 + \left(E\left[y(n)\right]\right)^2 \mathbf{1}\mathbf{1}^T$, where $\mathbf{1}$ is a column vector of ones, $\mathbf{x}^T E\left[\mathbb{Y}(n)\mathbb{Y}(n)^T\right]\mathbf{x} > 0$, thus $E\left[\mathbb{Y}(n)\mathbb{Y}(n)^T\right]$ is positive definite, thus full rank ([16], p. 98), completing the proof.

Lemma 2: $\mathbf{R} = E\left[\mathbf{y}(n)\mathbf{y}(n)^T\right]$ is positive definite, thus full rank.

Proof: Consider an arbitrary, real, non-zero vector $\mathbf{x}_+ \equiv [\mathbf{x}^T, x_{DC}]^T$ that is length $dQ+1$, i.e. $\mathbf{x}$ is length dQ and x_{DC} is a scalar.

$$\mathbf{x}_+^T\mathbf{R}\mathbf{x}_+ = E\left[\left(\mathbf{x}^T\mathbf{y}(n) + x_{DC}y_{DC}\right)^2\right] = E\left[\left(\mathbf{x}^T\mathbf{y}(n)\right)^2 + 2\mathbf{x}^T\mathbf{y}(n)x_{DC}y_{DC} + \left(x_{DC}y_{DC}\right)^2\right] \geq 0$$

Since $\mathbf{x}_+^T\mathbf{R}\mathbf{x}_+$ cannot be negative, determine if it can equal zero. Define $\eta \equiv x_{DC}y_{DC}$ and find the roots η_{root} of $\mathbf{x}_+^T\mathbf{R}\mathbf{x}_+ = 0$.

$$\begin{aligned}
\eta_{root} &= \frac{-2\mathbf{x}^T E\left[\mathbf{y}(n)\right] \pm \sqrt{4\mathbf{x}^T E\left[\mathbf{y}(n)\right]E\left[\mathbf{y}(n)^T\right]\mathbf{x} - 4\mathbf{x}^T E\left[\mathbf{y}(n)\mathbf{y}(n)^T\right]\mathbf{x}}}{2} \\
&= \mathbf{x}^T E\left[\mathbf{y}(n)\right] \pm \sqrt{-\mathbf{x}^T E\left[\left(\mathbf{y}(n) - E\left[\mathbf{y}(n)\right]\right)\left(\mathbf{y}(n)^T - E\left[\mathbf{y}(n)^T\right]\right)\right]\mathbf{x}} \\
&= \mathbf{x}^T E\left[\mathbf{y}(n)\right] \pm \sqrt{-\mathbf{x}^T\boldsymbol{\sigma}^2\mathbf{x}}
\end{aligned}$$

However, from Lemma 1, $\mathbf{x}^T\boldsymbol{\sigma}^2\mathbf{x} > 0$, thus η_{root} must be complex, contradicting the definition of x_{DC}. Therefore, for any real $\mathbf{x}_+$, $\mathbf{x}_+^T\mathbf{R}\mathbf{x}_+ > 0$ and $\mathbf{R}$ is full rank, completing the proof.

2.3 Optimal Controller

With Assumption 1–Assumption 4, the adaptive process (4)–(7) with the plant (2) can be represented as in Figure 3.

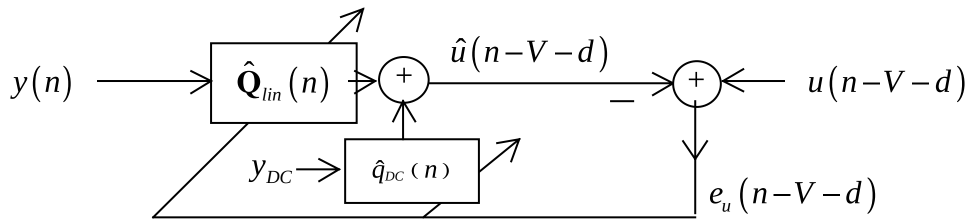

Figure 3– Adaptive System for Calculating $\hat{\mathbf{Q}}(n)$

The optimal solution $\mathbf{Q}_0$ for the adaptive coefficients $\hat{\mathbf{Q}}(n)$ is defined as

$$\mathbf{Q}_0 = \arg\min_{\hat{\mathbf{Q}}} \left\{ e(n)^2 \right\} \tag{10}$$

Defining

$$\mathbf{R} = E\left[\mathbf{y}(n)\mathbf{y}(n)^T \right], \quad \boldsymbol{\rho} = E\left[\mathbf{y}(n)u(n-V-d) \right] \tag{11}$$

$\mathbf{Q}_0$ is known to be [13] the solution of

$$\mathbf{R}\mathbf{Q}_0 = \boldsymbol{\rho} \tag{12}$$

$$\mathbf{Q}_0 = \mathbf{R}^{-1}\boldsymbol{\rho} \tag{13}$$

Note that this solution exists and is unique due to Lemma 2.

Now consider a different yet similar scheme than that shown in Figure 3 where the DC tap is not employed but the means of *y* and *u* are removed.

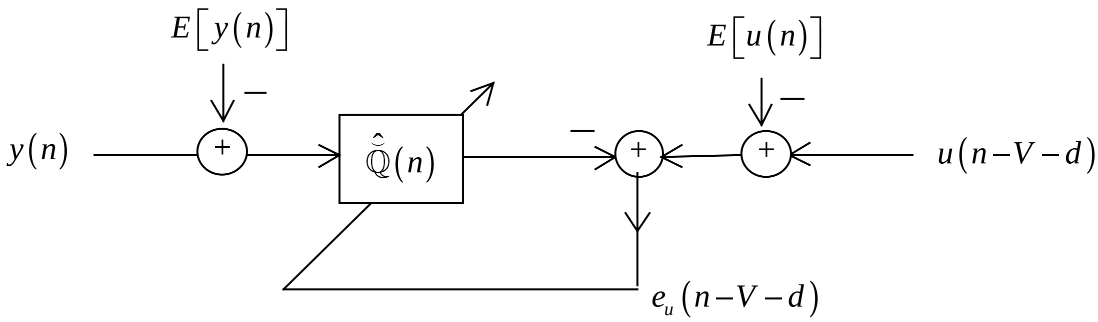

Figure 4 – Adaptive System with Means Explicitly Removed

Define

$$\breve{\mathbb{Y}}(n) \equiv \mathbb{Y}(n) - E\left[\mathbb{Y}(n)\right] \text{ and } \breve{u}(n) \equiv u(n) - E\left[u(n)\right], \tag{14}$$

i.e., $\breve{\mathbb{Y}}(n)$ and $\breve{u}(n)$ are the mean-removed versions of $\mathbb{Y}(n)$ and $u(n)$. The optimal solution for the adaptive coefficients $\hat{\breve{\mathbb{Q}}}$ is $\breve{\mathbb{Q}}_0$, which solves

$$E\left[\breve{\mathbb{Y}}(n)\breve{\mathbb{Y}}(n)^T\right]\breve{\mathbb{Q}}_0 = E\left[\breve{\mathbb{Y}}(n)\breve{u}(n-V-d)\right], \text{ or} \tag{15}$$

$$\breve{\mathbb{Q}}_0 = \left\{E\left[\breve{\mathbb{Y}}(n)\breve{\mathbb{Y}}(n)^T\right]\right\}^{-1} E\left[\breve{\mathbb{Y}}(n)\breve{u}(n-V-d)\right] \tag{16}$$

The solutions $\breve{\mathbb{Q}}_0$ and $\mathbf{Q}_0$ are closely related, as shown by Lemma 3.

Lemma 3 – The unique solution of (12) is

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

<u>Proof</u>: As noted before, $\mathbf{Q}_0$ is unique due to Lemma 2. Describe

$$\left[\breve{\mathbb{Q}}_0^{\,T}, \frac{-E\left[y(n)\right]\sum_i \breve{\mathbb{Q}}_{0,i} + E\left[u(n-V-d)\right]}{y_{DC}}\right]^T \tag{17}$$

as the *proposed solution* of (12). If the proposed solution is in fact a solution to (12), then it is the unique solution. Equation (12) defines $dQ+1$ linear equations. The proposed solution is in fact the solution of the equation defined by the first row of (12), after substituting the definitions of (11), as shown by:

$$\begin{aligned}
&\sum_{i=1}^{dQ} E\left[y(n)\,y(n-i-1)\right]\breve{\mathbb{Q}}_{0,i} + y_{DC}E\left[y(n)\right]\frac{-E\left[y(n)\right]\sum_i \breve{\mathbb{Q}}_{0,i} + E\left[u(n-V-d)\right]}{y_{DC}} \\
&= \sum_{i=1}^{dQ} E\left[y(n)\,y(n-i-1) - \left(E\left[y(n)\right]\right)^2\right]\breve{\mathbb{Q}}_{0,i} + E\left[y(n)\right]E\left[u(n-V-d)\right] \\
&= \sum_{i=1}^{dQ} E\left[\breve{y}(n)\,\breve{y}(n-i-1)\right]\breve{\mathbb{Q}}_{0,i} + E\left[y(n)\right]E\left[u(n-V-d)\right]
\end{aligned} \tag{18}$$

where the last line uses the fact that

$$\begin{aligned}
E\left[y(n)\,y(n-i)\right] - \left(E\left[y(n)\right]\right)^2 &= E\left[\left(y(n)-E\left[y(n)\right]\right)\left(y(n-i)-E\left[y(n-i)\right]\right)\right] \\
&= E\left[\breve{y}(n)\,\breve{y}(n-i)\right]
\end{aligned}$$

From the equation defined by the first row of (15)

$$\sum_{i=1}^{dQ} E\left[\breve{y}(n)\,\breve{y}(n-i-1)\right]\breve{\mathbb{Q}}_{0,i} = E\left[\breve{y}(n)\,\breve{u}(n-V-d)\right] \tag{19}$$

Substituting (19) into the last line of (18) gives

$$\begin{aligned}
&\sum_{i=1}^{dQ} E\left[y(n)\,y(n-i-1)\right]\breve{\mathbb{Q}}_{0,i} + y_{DC}E\left[y(n)\right]\frac{-E\left[y(n)\right]\sum_i \breve{\mathbb{Q}}_{0,i} + E\left[u(n-V-d)\right]}{y_{DC}} \\
&= E\left[\breve{y}(n)\,\breve{u}(n-V-d)\right] + E\left[y(n)\right]E\left[u(n-V-d)\right] \\
&= E\left[y(n)\,u(n-V-d)\right] - E\left[y(n)\right]E\left[u(n-V-d)\right] + E\left[y(n)\right]E\left[u(n-V-d)\right] \\
&= E\left[y(n)\,u(n-V-d)\right]
\end{aligned} \tag{20}$$

since $\breve{y}(n)$ and $\breve{u}(n-V-d)$ are both zero mean. Thus from (20), the proposed solution does in fact solve the first row of (12). In a similar manner, the proposed solution can be shown to solve rows 2 through dQ of (12). It is trivial to show that the proposed solution solves the last row of (12):

$$E\left[y(n)\right]y_{DC}\sum_{i=1}^{dQ}\breve{\mathbb{Q}}_{0,i}+\left(y_{DC}\right)^2\frac{-E\left[y(n)\right]\sum_i\breve{\mathbb{Q}}_{0,i}+E\left[u(n-V-d)\right]}{y_{DC}} = y_{DC}E\left[u(n-V-d)\right] \tag{21}$$

Thus the proposed solution is indeed a solution to (12) and in fact the unique solution, concluding the proof.

Lemma 3 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 $e^*(n)\equiv u(n-d-V)-\mathbf{Q}_0(n)^T\,\mathbf{y}(n)$,

$$e^*(n)\equiv u(n-d-V)-\mathbf{Q}_0(n)^T\,\mathbf{y}(n) \tag{22}$$

then $E\left[e^*(n)\right]=0$, as shown in Lemma 4.

Lemma 4 - $E\left[e^*(n)\right]=0$.

Proof: From (22)

$$E\left[e*(n)\right]=E\left[u(n-V-d)\right]-E\left[\mathbf{Q}_0(n)^T\,\mathbf{y}(n)\right]$$

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

$$=0$$

completing the proof.

2.4 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. For matrix **R**, let the matrices of normalized eigenvectors and eigenvalues be $\mathbf{W}^T$ and $\mathbf{\Lambda}$ respectively, where $\mathbf{\Lambda}=\mathrm{diag}(\lambda_i),\ i=1,\ \cdots,\ (dQ+1)$,

$$\mathbf{W}^T\mathbf{W}=\mathbf{I} \tag{23}$$

$$\mathbf{W}\,\mathbf{R}\,\mathbf{W}^T=\mathbf{\Lambda} \tag{24}$$

Note that **W** is full rank by Lemma 2. Now define a linear transformation of the random vector $\mathbf{y}(n)$ as follows.

$$\boldsymbol{\psi}(n)=\mathbf{W}\,\mathbf{y}(n),\ \mathbf{W}^T\boldsymbol{\psi}(n)=\mathbf{W}^T\mathbf{W}\,\mathbf{y}(n)=\mathbf{y}(n) \tag{25}$$

$$\mathbf{L}(n)=\mathbf{W}\tilde{\mathbf{Q}}(n)=\mathbf{W}\left(\hat{\mathbf{Q}}(n)-\mathbf{Q}_0(n)\right),\ \mathbf{W}^{\mathbf{T}}\mathbf{L}(n)=\tilde{\mathbf{Q}}(n) \tag{26}$$

$$E\left[\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T\right]=\mathbf{\Lambda} \tag{27}$$

Lemma 5 – The elements of $\boldsymbol{\psi}(n)$ are independent.

Proof: For $i\neq j$

$$
\begin{aligned}
&E\left[\left(\psi_i(n)-E\left[\psi_i(n)\right]\right)\left(\psi_j(n)-E\left[\psi_j(n)\right]\right)\right] \\
&=E\left[\left(y_i(n)-E\left[y_i(n)\right]\right)^T \mathbf{W}_i^T \mathbf{W}_j \left(y_j(n)-E\left[y_j(n)\right]\right)\right] \\
&=0
\end{aligned}
\tag{28}
$$

since (23) gives $\mathbf{W}_i^T \mathbf{W}_j = 0$ where $\mathbf{W}_j$ is the *j*'th row of $\mathbf{W}$. From the definition of the probability density function of two guassian random variables, (28) gives that $\psi_i(n)$ and $\psi_j(n)$ are independent (p. 126 of [15]), thus completing the proof.

The update equation (7) can be written as

$$
\hat{\mathbf{Q}}(n+1) = \left(I - \mu \frac{\mathbf{y}(n)\mathbf{y}(n)^T}{\mathbf{y}(n)^T \mathbf{y}(n)} \right) \hat{\mathbf{Q}}(n) + \frac{\mu\, \mathbf{y}(n)}{\mathbf{y}(n)^T \mathbf{y}(n)} u(n-V-d)
$$

Pre-multiplying by **W** and adding and subtracting a term,

$$
\begin{aligned}
\mathbf{W}\hat{\mathbf{Q}}(n+1) = \mathbf{W}\hat{\mathbf{Q}}(n) - \mu\,\mathbf{W}\frac{\mathbf{y}(n)\mathbf{y}(n)^T}{\mathbf{y}(n)^T \mathbf{y}(n)}\hat{\mathbf{Q}}(n) + \frac{\mu\,\mathbf{W}\,\mathbf{y}(n)\mathbf{y}(n)^T}{\mathbf{y}(n)^T \mathbf{y}(n)}\mathbf{Q}_o \\
-\mu\,\frac{\mathbf{W}\,\mathbf{y}(n)\mathbf{y}(n)^T}{\mathbf{y}(n)^T \mathbf{y}(n)}\mathbf{Q}_0 + \frac{\mu\,\mathbf{W}\,\mathbf{y}(n)u(n-d-V)}{\mathbf{y}(n)^T \mathbf{y}(n)}
\end{aligned}
\tag{29}
$$

Using (22) and subtracting $\mathbf{W}\mathbf{Q}_0$ from both sides of (29)

$$
\mathbf{L}(n+1) = \left(I - \mu \frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)} \right) \mathbf{L}(n) + \frac{\mu\ \boldsymbol{\psi}(n)e^*(n)}{\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)}
\tag{30}
$$

and

$$
\begin{aligned}
\mathbf{L}(n+1)\mathbf{L}(n+1)^T = {} & \mathbf{L}(n)\mathbf{L}(n)^T \\
& -\mu \frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)} \mathbf{L}(n)\mathbf{L}(n)^T \\
& -\mu \mathbf{L}(n)\mathbf{L}(n)^T \frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)} \\
& +\mu^2 \frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)} \mathbf{L}(n)\mathbf{L}(n)^T \frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)} \\
& +\mu \frac{e^*(n)}{\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)} \left[\mathbf{L}(n)\boldsymbol{\psi}(n)^T \right] \\
& +\mu \frac{e^*(n)}{\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)} \left[\boldsymbol{\psi}(n)\mathbf{L}(n)^T \right] \\
& -\mu \frac{e^*(n)}{\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)} \left[\frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)} \mathbf{L}(n)\boldsymbol{\psi}(n)^T \right] \\
& -\mu \frac{e^*(n)}{\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)} \left[\boldsymbol{\psi}(n)\mathbf{L}(n)^T \frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)} \right] \\
& + \frac{\mu^2 \boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T \left(e^*(n)\right)^2}{\left(\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)\right)^2}
\end{aligned} \tag{31}
$$

The following notations will be used extensively:

$$
\mathbf{A} \equiv E\left[\frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)} \right] \tag{32}
$$

$$
\mathbf{C}(n) \equiv E\left[\mathbf{L}(n)\mathbf{L}(n)^T \right] \tag{33}
$$

$$
\mathbf{D}(n) \equiv E\left[\frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)} \; \mathbf{C}(n) \; \frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)} \right] \tag{34}
$$

$$
\mathbf{H} \equiv E\left[\frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{\left(\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)\right)^2} \right] \tag{35}
$$

3 Parameter Convergence in the Mean

In this section it will be shown that

$$\lim_{n\to\infty} E\left[\mathbf{L}(n)\right] = \mathbf{0}, \tag{36}$$

which in turn, by (26), shows $\lim_{n\to\infty} E\left[\tilde{\mathbf{Q}}(n)\right] = \mathbf{0}$. With the contributions of Section 2.4, the proof is already half-completed. Before proceeding, we show a few key independencies.

Lemma 6: $\boldsymbol{\psi}(n)$ and $\mathbf{L}(n)$ are independent.

Proof: This follows from Assumption 2, which gives

$$f_{\mathbf{Y},\tilde{\mathbf{Q}}}(\mathbf{y},\mathbf{q}) = f_{\mathbf{Y}}(\mathbf{y})\, f_{\tilde{\mathbf{Q}}}(\mathbf{q}) \tag{37}$$

Define

$$\mathbf{R} = \begin{bmatrix} \mathbf{y}(n) \\ \tilde{\mathbf{Q}}(n) \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} \boldsymbol{\psi}(n) \\ \mathbf{L}(n) \end{bmatrix} \tag{38}$$

noting that

$$f_{\mathbf{R}}(\mathbf{r}) = f_{\mathbf{Y},\tilde{\mathbf{Q}}}(\mathbf{y},\mathbf{q}) \tag{39}$$

$$f_{\mathbf{S}}(\mathbf{s}) = f_{\boldsymbol{\psi},\mathbf{L}}(\boldsymbol{\Psi},\mathbf{l}) \tag{40}$$

Also define

$$\mathbf{U} = \begin{bmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{W} \end{bmatrix} \tag{41}$$

noting that from Lemma 2, $\mathbf{U}$ is full rank. Also, like $\mathbf{W}$, $\mathbf{U}$ is orthonormal (see (23)), $\mathbf{U}\mathbf{U}^T = \mathbf{I}$, thus $\mathbf{U}^{-1} = \mathbf{U}^T$. Then

$$\mathbf{U}\mathbf{R} = \mathbf{S} \tag{42}$$

Since **U** is full rank and thus has no null space, each unique instance of **R** maps to a unique instance of **S**. Likewise, since

$$\mathbf{U}^{\mathbf{T}}\mathbf{S} = \mathbf{R} \tag{43}$$

and $\mathbf{U}^T$ has no null space, each unique instance of **S** maps to a unique instance of **R**.

From [15], given $f_{\mathbf{R}}(\mathbf{r})$ and that the transformations (42) and (43) are unique, (40) is given by

$$f_{\mathbf{S}}(s) = f_{\mathbf{R}}\left(\mathbf{r} = \mathbf{U}^T\mathbf{s}\right)|\Xi| \tag{44}$$

where Ξ is the Jacobian of the transformation of (43) ([15], p. 129), which can be shown to equal $\mathbf{U}^T$. Since the determinant of an orthonormal matrix, e.g. $\mathbf{U}^T$, is either 1 or –1, $|\Xi| = 1$, thus from (39), (40), and (37)

$$f_{\boldsymbol{\psi},\mathbf{L}}(\boldsymbol{\Psi},\mathbf{l}) = f_{\mathbf{Y}}\left(\mathbf{y} = \mathbf{W}^T\boldsymbol{\psi}\right) f_{\tilde{\mathbf{Q}}}\left(\mathbf{q} = \mathbf{W}^T\mathbf{l}\right) \tag{45}$$

By inspection,

$$f_{\boldsymbol{\psi}}(\boldsymbol{\Psi}) = f_{\mathbf{Y}}\left(\mathbf{y} = \mathbf{W}^T\boldsymbol{\psi}\right), f_{\mathbf{L}}(\mathbf{l}) = f_{\tilde{\mathbf{Q}}}\left(\mathbf{q} = \mathbf{W}^T\mathbf{l}\right) \tag{46}$$

thus $\boldsymbol{\psi}(n)$ and $\mathbf{L}(n)$ are independent, completing the proof.

Lemma 7: $u(n-V-d)$ and $\mathbf{L}(n)$ are independent. Also $\mathbf{y}(n)$ and $\mathbf{L}(n)$ are independent.

Proof: From Assumption 2 and (26).

Lemma 8: $\mathbf{y}(n)$ and $e^*(n)$ are independent. $\boldsymbol{\psi}(n)$ is also independent of $e^*(n)$.

Proof: First note that

$$E\left[\mathbf{y}(n)e^*(n)\right] = E\left[\mathbf{y}(n)u(n-V-d)\right] - E\left[\mathbf{y}(n)\mathbf{y}(n)^T\right]\mathbf{Q}_0 \\ = \boldsymbol{\rho} - \mathbf{R}\ \mathbf{R}^{-1}\ \boldsymbol{\rho} = \mathbf{0} \tag{47}$$

Since e*(n) consists of linear combinations of jointly Gaussian random variables $u(n-V-d)$ and **y**(n), $e^*(n)$ is jointly Gaussian with **y**(n) [15]. Let

$$\varsigma(n) \equiv \left[e^*(n), \mathbf{y}(n)^T\right]^T \tag{48}$$

Therefore, using (47),

$$E\left[\varsigma(n)\varsigma(n)^T\right] = \left[\begin{array}{c|ccc} E\left[e^*(n)^2\right] & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \mathbf{R} & \\ 0 & & & \end{array}\right] \tag{49}$$

From the definition of a determinant as the sum of weighted cofactors (see p. 226 of [16]),

$$\det E\left[\varsigma(n)\varsigma(n)^T\right] = E\left[\left(e^*(n)\right)^2\right]\det \mathbf{R} \tag{50}$$

and since

$$E\left[\varsigma(n)\varsigma(n)^T\right]^{-1} = \left[\begin{array}{c|ccc} 1/E\left[e^*(n)^2\right] & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \mathbf{R}^{-1} & \\ 0 & & & \end{array}\right] \tag{51}$$

by direct substitution

$$\left(\varsigma(n) - E\left[\varsigma(n)\right]\right)^T \left(E\left[\varsigma(n)\varsigma(n)^T\right]\right)^{-1} \left(\varsigma(n) - E\left[\varsigma(n)\right]\right) \\ = \left(\mathbf{y}(n) - E\left[\mathbf{y}(n)\right]\right)^T \mathbf{R}^{-1}\left(\mathbf{y}(n) - E\left[\mathbf{y}(n)\right]\right)\left(E\left[e^*(n)^2\right]\right)^{-1}\left(e^*(n) - E\left[e^*(n)\right]\right)^2 \tag{52}$$

Then by inspection of the multi-variable Gaussian pdf of ς, from (50) and (52), $e^*(n)$ and $\mathbf{y}(n)$ are independent, completing the proof for the first part. Stated more succinctly, The auto-covariance matrix of two jointly-Gaussian, uncorrelated random variables where at least one is zero-mean ($e^*(n)$) is diagonal, therefore independent.

Since $E\left[\boldsymbol{\psi}(n)e^*(n)\right] = \mathbf{W}E\left[\mathbf{y}(n)e^*(n)\right] = 0$, $\boldsymbol{\psi}(n)$ and $e^*(n)$ can be shown to be independent using a similar argument as above, thus completing the proof.

Lemma 8 states that the error $e^*(n)$ produced when the optimal mean-square estimate filter $\mathbf{Q}_0$ is used is orthogonal to the excitation signal $\mathbf{y}(n)$, an intuitive result.

Lemma 9: $\mathbf{A}$ 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

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

From Assumption 4,

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

From (27),

$$\left|A_{ij}\right| \le \begin{cases} 0 \text{ if } i \ne j \\ \dfrac{\lambda_i}{\alpha_0} \text{ if } i = j \end{cases} \tag{53}$$

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

Lemma 10: $0 < \alpha_1 < A_{ii} \le 1$.

Proof: $A_{ii} = \int_{\mathbf{X}} \frac{x_i(n)^2}{\sum_{j=1}^{dQ+1} x_j(n)^2} f_{\boldsymbol{\psi}(n)}(\mathbf{X})\, d\mathbf{X}$. $0 \le \frac{x_i(n)^2}{\sum_{j=1}^{dQ+1} x_j(n)^2} \le 1$ and $0 \le f_{\boldsymbol{\psi}(n)}(\mathbf{X}) \le 1$ for any $\mathbf{X}$. Then it is possible to choose a closed, bounded set $\boldsymbol{\Omega}$ that simultaneously satisfies four constraints: 1. $x_i \ne 0$, 2. $\sum_j x_j{}^2 < \infty$, 3. $0 < f_{\boldsymbol{\psi}(n)}(\mathbf{X})$, 4. $\int_{\boldsymbol{\Omega}} \frac{x_i(n)^2}{\sum_{j=1}^{dQ+1} x_j(n)^2} f_{\boldsymbol{\psi}(n)}(\mathbf{X})\, d\mathbf{X} > \alpha_1$.

Since $\frac{x_i(n)^2}{\sum_{j=1}^{dQ+1} x_j(n)^2} f_{\boldsymbol{\psi}(n)}(\mathbf{X})$ is non-negative for every $\mathbf{X}$, the proof is completed.

Now we reach the main result of this Section.

Theorem 1 - Given Assumption 1– Assumption 4 and $0 < \mu < 2$,

$$\lim_{n\to\infty} E\left[\tilde{\mathbf{Q}}(n)\right] = \mathbf{0}$$

Proof: Consider (30). From Lemma 8 and Lemma 4

$$E\left[\frac{\mu\boldsymbol{\psi}(n)e^*(n)}{\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n)}\right] = \mu E\left[\frac{\boldsymbol{\psi}(n)}{\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n)}\right] E\left[e^*(n)\right] = 0 \tag{54}$$

From Lemma 6, (54), (32), and (30)

$$E\left[\mathbf{L}(n+1)\right] = (\mathbf{I} - \mu\ \mathbf{A})\ E\left[\mathbf{L}(n)\right] \tag{55}$$

From Lemma 9, we see that the linear system of (55) completely decouples:

$$E\left[L_i(n+1)\right] = (1 - \mu A_{ii}) E\left[L_i(n)\right],\ i = 1,\ \ldots,\ dQ+1 \tag{56}$$

From Lemma 10 and if $0 < \mu < 2$

$$\left|(1 - \mu A_{ii})\right| < 1,\ i = 1,\ \ldots,\ dQ+1 \tag{57}$$

Taking (56) and (57) together

$$\lim_{n\to\infty} E\left[\mathbf{L}(n)\right] = \mathbf{0} \tag{58}$$

and since (26) gives that $\tilde{\mathbf{Q}}(n) = \mathbf{W}^T \mathbf{L}(n)$, from (23) and (58)

$$\lim_{n\to\infty} E\left[\tilde{\mathbf{Q}}(n)\right] = \mathbf{0} \tag{59}$$

thus completing the proof.

Theorem 1 states that given Assumption 1–Assumption 4, if we bound μ as

$$0 < \mu < 2 \tag{60}$$

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 3. It is distinctive from the proof of [12] in that it uses Lemma 4 – instead of a zero-mean assumption for $\mathbf{y}(n)$ – to eliminate the expectation of the second term of (30).

4 Parameter Convergence in the Mean Square

Theorem 1 tells us that $\hat{\mathbf{Q}}(n)$ will converge in the mean, i.e. the mean value of $\hat{\mathbf{Q}}(n)$ will equal the desired $\mathbf{Q}_0$. However, Theorem 1 only speaks to how the parameter estimation error vector $\tilde{\mathbf{Q}}(n)$ behaves in the mean. It does not bound the variance of the terms of $\tilde{\mathbf{Q}}(n)$. Thus if $\mathbf{Q}_{0,i}$ was a small positive number, $\hat{\mathbf{Q}}_i(n)$ could alternate between a huge positive number and a huge negative number for successive *n* and still not violate Theorem 1. Clearly this would be a poor controller for our congestion control system. 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 [12] relies heavily on a zero-mean assumption on $\mathbf{y}(n)$, an assumption not made here. However, Lemma 11 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 [12].

Lemma 11 – Given Lemma 5, (27) 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. From Lemma 5, $E\left[\psi_i(n)\psi_j(n)\right]=E\left[\psi_i(n)\right]E\left[\psi_j(n)\right]$. If less than dQ elements of $\boldsymbol{\psi}(n)$ are zero mean, then $E\left[\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T\right]$ cannot be diagonal, contradicting (27). Necessity: If no less than dQ of the $dQ+1$ elements of $\boldsymbol{\psi}(n)$ are zero mean, then Lemma 5 gives (27) concluding the proof.

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

$$E\left[\psi_i(n)\right]=0, i=1,\ldots,dQ+1, i\neq\zeta \quad (61)$$

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

Now we massage (31) into something more manageable. Begin by taking the expectation of (31). The expectation of the fifth through eighth term on the right side of (31) are zero, as shown in

Lemma 12 – The expectation of the fifth through eighth term on the right side of (31) is zero.

Proof: Examine the expectation of the seventh term on the right side of (31), specifically at row i, column j:

$$E\left[\mu \frac{e^*(n)}{\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)} \frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}(n)^T}{\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)} \mathbf{L}(n)\boldsymbol{\psi}(n)\right]_{i,j} = \mu E\left[\frac{e^*(n)\psi_i(n)\psi_j(n)\sum_{k=1}^{dQ+1} L_k(n)\psi_k(n)}{\left(\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)\right)^2}\right] \quad (62)$$

From Lemma 6, $L_k(n)$ is independent from $\boldsymbol{\psi}(n)$, and from Lemma 8, $e^*(n)$ is independent from $\boldsymbol{\psi}(n)$. Thus

$$\mu E\left[\frac{e^*(n)\psi_i(n)\psi_j(n)L_k(n)\psi_k(n)}{\left(\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)\right)^2}\right] = \mu E\left[e^*(n)L_k(n)\right] E\left[\frac{\psi_i(n)\psi_j(n)\psi_k(n)}{\left(\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)\right)^2}\right] \quad (63)$$

Now

$$E\left[e^*(n)L_k(n)\right] = \mathbf{W}_k E\left[e^*(n)\tilde{\mathbf{Q}}(n)\right] = \mathbf{W}_k E\left[e^*(n)\right] E\left[\tilde{\mathbf{Q}}(n)\right] = 0 \quad (64)$$

where the first equality is due to (26), the second equality is due to (22) and Assumption 2, and the third equality is due to Lemma 4. Substituting (64) into (63) and (63) into (62) 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 (31) produce matrices of zeros, thus completing the proof.

With (32) – (35), (31) becomes the much-less cumbersome

$$\mathbf{C}(n+1) = \mathbf{C}(n) - \mu\left(\mathbf{A}\,\mathbf{C}(n) + \mathbf{C}(n)\mathbf{A}\right) + \mu^2\mathbf{D}(n) + \mu^2\,\varepsilon^*\mathbf{H} \quad (65)$$

where ε^* is the minimal mean-square error

$$\varepsilon^* \equiv E\left[\left(e^*(n)\right)^2\right] \quad (66)$$

In this section we will show that the linear, time-invariant difference equation (65) converges to a finite value, i.e.

$$\lim_{n\to\infty}\mathbf{C}(n) < \alpha_2\mathbf{I}, \alpha_2 < \infty \tag{67}$$

Since (65) is a discrete, linear, time invariant difference equation, its convergence (67) is guaranteed if its autonomous part is asymptotically stable, as this implies BIBO stability, and if its forcing term $\mu^2\varepsilon * \mathbf{H}$ is bounded. We begin with the latter.

Lemma 13: $|H_{ii}| \le 1/(dQ-2)\lambda_{\min}$

Proof:

$$|\mathbf{H}_{ii}| \le tr[\mathbf{H}] = E\left[\frac{1}{\boldsymbol{\psi}(n)^T\,\boldsymbol{\psi}(n)}\right] = E\left[\frac{1}{\boldsymbol{\psi}(n)^T\,\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{-1}\,\boldsymbol{\Lambda}^{1/2}\boldsymbol{\psi}(n)}\right]$$

$$|\mathbf{H}_{ii}| \le \frac{1}{\lambda_{\min}}E\left[\frac{1}{\boldsymbol{\psi}(n)^T\;\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda}^{-1/2}\;\boldsymbol{\psi}(n)}\right]$$

$\boldsymbol{\upsilon}(n) \equiv \boldsymbol{\Lambda}^{-1/2}\;\boldsymbol{\psi}(n)$ is a vector of independent Gaussian random variables with and unit variance with the further property that all but the ζ'th term is zero mean, i.e. $E[\upsilon_i(n)] = 0, i = 1,\ldots,dQ+1, i \ne \zeta$ (see (61)). [12] shows that

$$E\left[\left(\sum_{\substack{i=1,\\ i\ne\zeta}}^{dQ+1}\upsilon_i(n)^2\right)^{-1}\right] = \frac{1}{dQ-2} \tag{68}$$

and since

$$E\left[\left(\sum_{i=1}^{dQ+1}\upsilon_i(n)^2\right)^{-1}\right] \le E\left[\left(\sum_{\substack{i=1,\\ i\ne\zeta}}^{dQ+1}\upsilon_i(n)^2\right)^{-1}\right]$$

then $|H_{ii}| \le 1/(dQ-2)\lambda_{\min}$, thus completing the proof.

Lemma 14: Assuming $0 < \mu < 2$, the forcing term of (65) is bounded,

i.e. $\left|\mu^2 \varepsilon^* \mathbf{H}_{ii}\right| < \alpha_3,\ \alpha_3 < \infty$, $\left|\mu^2 \varepsilon^* \mathbf{H}_{ij}\right| = 0, i \neq j$

<u>Proof</u>: From Lemma 9, $\mathbf{H}$ is diagonal, and from Lemma 13, $\left|H_{ii}\right| < 1/(dQ-2)\lambda_{\min}$. μ^2 is positive and bounded by assumption. ε^* is also, by convention, bounded and non-negative, thus completing the proof.

Lemma 15: Since the elements of $\boldsymbol{\psi}(n)$ are gaussian and independent, and given Lemma 11, the expectation of an Odd Function of $\boldsymbol{\psi}(n)$ around $\psi_i(n)$ for $i \neq \zeta$ is zero.

<u>Proof</u>: A function of $\boldsymbol{\psi}(n)$, $\Gamma_i(\boldsymbol{\psi}(n))$, is defined as an *Odd Function around* $\psi_i(n)$ if

$$\begin{aligned}&\Gamma_i\left(\psi_1(n), \psi_2(n), \ldots, \psi_i(n), \ldots, \psi_{dQ+1}(n)\right) \\ &\qquad = -\Gamma_i\left(\psi_1(n), \psi_2(n), \ldots, -\psi_i(n), \ldots, \psi_{dQ+1}(n)\right)\end{aligned} \tag{69}$$

Also define $\hat{\boldsymbol{\psi}}(n)$ as $\boldsymbol{\psi}(n)$ without the i'th element, i.e.

$$\hat{\boldsymbol{\psi}}(n) \equiv \left[\psi_1(n), \psi_2(n), \ldots, \psi_{i-1}(n), \psi_{i+1}(n), \ldots, \psi_{dQ+1}(n)\right] \tag{70}$$

Because the elements of $\boldsymbol{\psi}(n)$ are independent

$$f_{\boldsymbol{\psi}(n)}(\boldsymbol{\psi}) = f_{\psi_i(n)}(\psi_i)\, f_{\hat{\boldsymbol{\psi}}(n)}(\hat{\boldsymbol{\psi}}) \tag{71}$$

and since $E\left[\psi_i(n)\right] = 0$ and $\psi_i(n)$ is Gaussian,

$$f_{\psi_i(n)}(\psi_i) = f_{\psi_i(n)}(-\psi_i) \tag{72}$$

Then

$$E\left[\Gamma_i\left(\boldsymbol{\psi}(n)\right)\right]=\int_{\hat{\boldsymbol{\psi}}=-\infty}^{\infty}\int_{\psi_i=-\infty}^{\infty}\Gamma_i\left(\boldsymbol{\psi}\right)f_{\psi_i(n)}\left(\psi_i\right)f_{\hat{\boldsymbol{\psi}}(n)}\left(\hat{\boldsymbol{\psi}}\right)d\psi_i d\hat{\boldsymbol{\psi}}$$

$$=\int_{\hat{\boldsymbol{\psi}}=-\infty}^{\infty}\left[\int_{\psi_i=-\infty}^{\infty}\Gamma_i\left(\boldsymbol{\psi}\right)f_{\psi_i(n)}\left(\psi_i\right)d\psi_i\right]f_{\hat{\boldsymbol{\psi}}(n)}\left(\hat{\boldsymbol{\psi}}\right)d\hat{\boldsymbol{\psi}}$$

$$=\int_{\hat{\boldsymbol{\psi}}=-\infty}^{\infty}\left[\int_{\psi_i=-\infty}^{0}\Gamma_i\left(\boldsymbol{\psi}\right)f_{\psi_i(n)}\left(\psi_i\right)d\psi_i+\int_{\psi_i=0}^{\infty}\Gamma_i\left(\boldsymbol{\psi}\right)f_{\psi_i(n)}\left(\psi_i\right)d\psi_i\right]f_{\hat{\boldsymbol{\psi}}(n)}\left(\hat{\boldsymbol{\psi}}\right)d\hat{\boldsymbol{\psi}}$$

Performing a change a variable on the first integral inside the brackets,

$$\begin{aligned}E\left[\Gamma_i\left(\boldsymbol{\psi}(n)\right)\right]&=\int_{\hat{\boldsymbol{\psi}}=-\infty}^{\infty}\left[\int_{t=\infty}^{0}\Gamma_i\left(\psi_1,\psi_2,\ldots,-t,\ldots,\psi_{dQ+1}\right)f_{\psi_i(n)}\left(-t\right)\left(-1\right)dt\right]f_{\hat{\boldsymbol{\psi}}(n)}\left(\hat{\boldsymbol{\psi}}\right)d\hat{\boldsymbol{\psi}}\\&\quad+\int_{\hat{\boldsymbol{\psi}}=-\infty}^{\infty}\left[\int_{\psi_i=0}^{\infty}\Gamma_i\left(\psi_1,\psi_2,\ldots,\psi_i,\ldots,\psi_{dQ+1}\right)f_{\psi_i(n)}\left(\psi_i\right)d\psi_i\right]f_{\hat{\boldsymbol{\psi}}(n)}\left(\hat{\boldsymbol{\psi}}\right)d\hat{\boldsymbol{\psi}}\\&=\int_{\hat{\boldsymbol{\psi}}=-\infty}^{\infty}\left[\int_{t=0}^{\infty}-\Gamma_i\left(\psi_1,\psi_2,\ldots,t,\ldots,\psi_{dQ+1}\right)f_{\psi_i(n)}\left(t\right)dt\right]f_{\hat{\boldsymbol{\psi}}(n)}\left(\hat{\boldsymbol{\psi}}\right)d\hat{\boldsymbol{\psi}}\\&\quad+\int_{\hat{\boldsymbol{\psi}}=-\infty}^{\infty}\left[\int_{\psi_i=0}^{\infty}\Gamma_i\left(\psi_1,\psi_2,\ldots,\psi_i,\ldots,\psi_{dQ+1}\right)f_{\psi_i(n)}\left(\psi_i\right)d\psi_i\right]f_{\hat{\boldsymbol{\psi}}(n)}\left(\hat{\boldsymbol{\psi}}\right)d\hat{\boldsymbol{\psi}}\\&=0\end{aligned}\qquad(73)$$

where the third line of (73) uses (69) and (72), thus completing the proof.

Lemma 16: $D_{ij}\left(n\right)$, the row *i*, column *j* element of $\mathbf{D}\left(n\right)$, can be computed as:

$$D_{ij}\left(n\right)=2G_{ij}C_{ij}\,,\ i\neq j\qquad(74)$$

$$D_{ii}\left(n\right)=\sum_{J=1}^{dQ+1}G_{iJ}C_{JJ}\qquad(75)$$

where **G** is defined as

$$G_{ij}\equiv E\left[\frac{\left(\boldsymbol{\psi}_i\left(n\right)\right)^2\left(\boldsymbol{\psi}_j\left(n\right)\right)^2}{\left(\boldsymbol{\psi}\left(n\right)^T\boldsymbol{\psi}\left(n\right)\right)^2}\right]\qquad(76)$$

<u>Proof</u>: By direct substitution, we have

$$D_{ij}(n)=\sum_{K=1}^{dQ+1}\sum_{J=1}^{dQ+1}E\left[\frac{\psi_i(n)\psi_j(n)\psi_J(n)\psi_K(n)}{\left(\boldsymbol{\psi}(n)^T\boldsymbol{\psi}(n)\right)^2}\right]C_{JK} \tag{77}$$

Since the elements of $\boldsymbol{\psi}(n)$ are independent and all but one are zero-mean, the numerator of the expectation in (77) will equal zero in many cases. To show which terms of the double summation of (77) equal zero, note that for any (i,j,J,K) the numerator can always be expressed as

$$\begin{aligned}&E\left[\psi_i(n)\psi_j(n)\psi_J(n)\psi_K(n)\right]\\&=E\left[\left(\psi_{\pi_1}(n)\right)^{\rho_1}\right]E\left[\left(\psi_{\pi_2}(n)\right)^{\rho_2}\right]E\left[\left(\psi_{\pi_3}(n)\right)^{\rho_3}\right]E\left[\left(\psi_{\pi_4}(n)\right)^{\rho_4}\right]\end{aligned} \tag{78}$$

such that $\rho_1,\rho_2,\rho_3,\rho_4\in\{0,1,2,3,4\}$, $\rho_1+\rho_2+\rho_3+\rho_4=4$, $\rho_1\geq\rho_2\geq\rho_3\geq\rho_4$, $\pi_1,\pi_2,\pi_3,\pi_4\in\{1,...,dQ+1\}$, 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:

<u>Case 1</u> $\rho_1=1,\rho_2=1,\rho_3=1,\rho_4=1$

<u>Case 2</u> $\rho_1=2,\rho_2=1,\rho_3=1,\rho_4=0$

<u>Case 3</u> $\rho_1=2,\rho_2=2,\rho_3=0,\rho_4=0$

<u>Case 4</u> $\rho_1=3,\rho_2=1,\rho_3=0,\rho_4=0$

<u>Case 5</u> $\rho_1=4,\rho_2=0,\rho_3=0,\rho_4=0$

Remembering that the element not assured to be zero-mean is denoted as $\psi_\zeta(n)$, (see (61)), then for Case 1, at most one member of the set $\{\pi_1,\pi_2,\pi_3,\pi_4\}$ equals ζ, thus (78) would equal zero. Similarly for Case 2, at most one member of the set $\{\pi_1,\pi_2,\pi_3\}$ equals ζ, thus (78) would equal zero.

For Case 4, if $\pi_1 = \zeta$, then from (61), (78) would equal zero. If $\pi_2 = \zeta$, then by Lemma 15, (78) would equal zero. If $\pi_1, \pi_2 \neq \zeta$, then by either (61) or Lemma 15, (78) would equal zero, thus (78) would equal zero for the entirety of Case 4.

For Cases 3 and 5, (78) could be non-zero.

Using an argument similar to that found in the proof of Lemma 9, 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 (77). 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 (77) equals (74). For $i = j$, after eliminating Cases 1, 2 and 4, Cases 3 and 5 make (77) equal to (75), thus completing the proof.

With the previous lemmas now presented, we can now directly address the task of showing the boundedness of convergence of $\mathbf{C}(n)$. We will consider the off-diagonal elements of $\mathbf{C}(n)$ separately from the diagonal elements. Note that $\mathbf{C}(n)$ is symmetric, i.e. $C_{ij}(n) = C_{ji}(n)$. Since $\mu^2 \varepsilon * \mathbf{H}$ is diagonal (see Lemma 9), the off-diagonal term of (65) is

$$C_{ij}(n+1) = \gamma_{ij} C_{ij}(n), i \neq j \tag{79}$$

where, using (74), (76), and Lemma 9,

$$\gamma_{ij} = 1 - \mu\left(A_{ii} + A_{jj}\right) + 2\mu^2 G_{ij}, i \neq j \tag{80}$$

Lemma 17: $\left|\gamma_{ij}\right| < 1$ and thus (79) is asymptotically stable.

<u>Proof</u>: Since $\left(\psi_i(n) - \psi_j(n)\right)^2 \geq 0$

$$\left(\psi_i(n)\right)^2 + \left(\psi_j(n)\right)^2 \geq 2\left(\psi_i(n)\right)\left(\psi_j(n)\right) \tag{81}$$

With $0<\mu<2$, (80) becomes

$$1-\gamma_{ij}=\mu\left[\left(A_{ii}+A_{jj}\right)-2\mu G_{ij}\right]$$
$$=\mu\ E\left[\frac{\left(\psi_i(n)\right)^2+\left(\psi_j(n)\right)^2}{\boldsymbol{\psi}(n)^T\,\boldsymbol{\psi}(n)}-\frac{2\mu\left(\psi_i(n)\right)^2\left(\psi_j(n)\right)^2}{\left(\boldsymbol{\psi}(n)^T\,\boldsymbol{\psi}(n)\right)^2}\right]$$

Squaring both sides of (81), we have

$$1-\gamma_{ij}\geq\mu\ E\left[\frac{\left(\psi_i(n)\right)^2+\left(\psi_j(n)\right)^2}{\boldsymbol{\psi}(n)^T\,\boldsymbol{\psi}(n)}\ \left\{1-\frac{\mu}{2}\cdot\frac{\left(\psi_i(n)\right)^2+\left(\psi_j(n)\right)^2}{\boldsymbol{\psi}(n)^T\,\boldsymbol{\psi}(n)}\right\}\right] \tag{82}$$
$$>0$$

where the strict inequality is due to Lemma 10. Conversely

$$1-\gamma_{ij}<\mu\left(A_{ii}+A_{jj}\right)=\mu\ E\left[\frac{\left(\psi_i(n)\right)^2+\left(\psi_j(n)\right)^2}{\boldsymbol{\psi}(n)^T\,\boldsymbol{\psi}(n)}\right]\leq\mu<2 \tag{83}$$

Taking (82) and (83) together, $0<1-\gamma_{ij}<2$, or, $\left|\gamma_{ij}\right|<1$, completing the proof.

Lemma 17 states that the off-diagonal terms of $\mathbf{C}(n)$ converge to zero as time increases. We can now focus on the diagonal entries of $\mathbf{C}(n)$. Define a vector of the diagonal entries of $\mathbf{C}(n)$:

$$\boldsymbol{\Omega}(n)\equiv\left[C_{11}(n),C_{22}(n),\ldots,C_{(dQ+1)(dQ+1)}(n)\right]^T \tag{84}$$

From (65),

$$\boldsymbol{\Omega}(n+1)=\mathbf{F}\boldsymbol{\Omega}(n)+\mu^2\varepsilon^*\mathbf{h} \tag{85}$$

where, from Lemma 9 and (75),

$$\mathbf{F}=\text{diag}\ \left\{\left(1-2\mu A_{ii}\right)\right\}+\mu^2\mathbf{G} \tag{86}$$

and $\mathbf{h}\equiv\left[H_{11},H_{22},\ldots,H_{(dQ+1)}\right]^T$.

Lemma 18: All entries of $\mathbf{F}$ are non-negative.

Proof: Again, we consider off-diagonal and diagonal entries separately. For $i \neq j$

$$F_{ij} = \mu^2 G_{ij} = \mu^2 E\left[\frac{\left(\psi_i(n)\right)^2 \left(\psi_j(n)\right)^2}{\left(\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)\right)^2}\right] \geq 0 \tag{87}$$

For diagonal entries, consider the fact that

$$E\left[\left(1 - \frac{\mu\left(\psi_i(n)\right)^2}{\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)}\right)^2\right] \geq 0 \tag{88}$$

Expanding (88),

$$\begin{aligned} 0 &\leq E\left[1 - \frac{2\mu\left(\psi_i(n)\right)^2}{\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)} + \frac{\mu^2\left(\psi_i(n)\right)^4}{\left(\boldsymbol{\psi}(n)^T \boldsymbol{\psi}(n)\right)^2}\right] \\ &\leq 1 - 2\mu A_{ii} + \mu^2 G_{ii} \\ &\leq F_{ii} \end{aligned} \tag{89}$$

which completes the proof.

Consider the autonomous part of (85)

$$\tilde{\boldsymbol{\Omega}}(n+1) = \mathbf{F}\, \tilde{\boldsymbol{\Omega}}(n) \tag{90}$$

where $\tilde{\boldsymbol{\Omega}}(0) = \boldsymbol{\Omega}(0)$, noting that from (84) and (33), for $i = 1, \ldots, dQ+1$,

$$\tilde{\Omega}_i(0) = C_{ii}(0) = \left(E\left[\mathbf{L}(0)\mathbf{L}(0)^T\right]\right)_{ii} \geq 0 \tag{91}$$

From (90) and Lemma 18, clearly

$$\tilde{\Omega}_i(n) \geq 0,\ i = 1, \ldots, dQ+1,\ n \geq 0 \tag{92}$$

Now define

$$v(n) \equiv \sum_{i=1}^{dQ+1} \tilde{\mathbf{\Omega}}_i(n) \tag{93}$$

Substituting (86) into (90), premultiplying the result by a row vector of ones, we have

$$\begin{aligned}[1,1,1,\ldots,1]\tilde{\mathbf{\Omega}}(n+1) = [1,1,1,\ldots,1]\tilde{\mathbf{\Omega}}(n) - 2\mu[1,1,1,\ldots,1]\operatorname{diag}\{A_{ii}\}\tilde{\mathbf{\Omega}}(n) \\ +\mu^2[1,1,1,\ldots,1]\mathbf{G}\,\tilde{\mathbf{\Omega}}(n)\end{aligned} \tag{94}$$

Using the fact that $\sum_{j=1}^{dQ+1} G_{ij} = A_{ii}$,

$$\begin{aligned}v(n+1) &= v(n) - 2\mu\left[A_{11}, A_{22}, \ldots, A_{(dQ+1)(dQ+1)}\right]\tilde{\mathbf{\Omega}}(n) \\ &\quad +\mu^2\left[A_{11}, A_{22}, \ldots, A_{(dQ+1)(dQ+1)}\right]\tilde{\mathbf{\Omega}}(n) \\ &= v(n) - \mu(2-\mu)\sum_{j=1}^{dQ+1} A_{jj}\tilde{\mathbf{\Omega}}_j(n)\end{aligned} \tag{95}$$

From (93) and Lemma 10

$$v(n+1) \le \left(1 - \mu(2-\mu)\alpha_1\right)v(n)$$

By constraining μ such that $0 < \mu < 2$

$$\lim_{n\to 0} v(n) = 0 \tag{96}$$

Finally we state the main result of this section.

Theorem 2: Given Assumption 1 - Assumption 4 and $0 < \mu < 2$

$$\lim_{n\to\infty} E\left[\tilde{\mathbf{Q}}(n)\tilde{\mathbf{Q}}(n)^T\right] = \mu^2\varepsilon^*\mathbf{W}^T(\mathbf{I}-\mathbf{F})^{-1}\mathbf{HW}$$

<u>Proof</u>: From the definition of $v(n)$ (93), with (96) and (92)

$$\lim_{n\to 0} \tilde{\mathbf{\Omega}}(n) = \mathbf{0} \tag{97}$$

proving the asymptotic stability of (90), which is the homogenous autonomous part of (85). This implies the BIBO stability of (85). From Lemma 14, the input signal is bounded, thus

$$\lim_{n\to\infty}\mathbf{\Omega}(n)=(\mathbf{I}-\mathbf{F})^{-1}\mu^2\varepsilon*\mathbf{h} \tag{98}$$

Taking (98) together with Lemma 17, $\lim_{n\to\infty}\mathbf{C}(n)=diag\left\{(\mathbf{I}-\mathbf{F})^{-1}\mu^2\varepsilon*\mathbf{h}\right\}$. The definitions of $\mathbf{C}(n)$ and $\mathbf{L}(n)$ give the final result, concluding the proof.

The key contribution of Section 4 comes from our showing (65) (Lemma 12), (74) and (75) (Lemma 16) without requiring $E[\mathbf{y}(n)]=\mathbf{0}$. Expressions (65), (74), and (75) exist in [12], but required $E[\mathbf{y}(n)]=\mathbf{0}$. By demonstrating (65), (74), and (75) without requiring $E[\mathbf{y}(n)]=\mathbf{0}$, the results of [12] are significantly extended.

5 Global Stability

Global Stability has been built into the control structure. Both plant and controller are FIR filters, thus BIBO stable if their coefficients are bounded in magnitude. 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 converge in the mean square to a bounded optimal solution. From an implementation view, the controller parameters can be kept bounded at each time *n* by a simple limiter after the adaptation of (7). By assumption, $\{y^*\}$ is bounded, and thus, $\{u\}$ and $\{y\}$ are bounded as well.

6 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] Rohrs, Charles E. and Berry, Randall A., "A Linear Control Approach to Explicit Rate Feedback in ATM Networks," IEEE Infocom '97 Kobe, Japan v. 3 1997 pp. 277-282

[6] Laberteaux, Kenneth P. and Rohrs, Charles E., "Application Of Adaptive Control To ATM ABR Congestion Control," Globecom '98 Sydney, Australia 1998.

[7] Raj Jain, Shiv Kalyanaraman, Rohit Goyal, Sonia Fahmy, and Fang Lu, "ERICA+: Extensions to the ERICA Switch Algorithm," ATM Forum/95-1346, October 1995. http://www.cis.ohio-state.edu/~jain/atmf/af_erc22.htm

[8] C. Fulton and S. Q. Li , ``UT: ABR Feedback Control with Tracking," Proc. IEEE Infocom'97 Conference, April 1997. http://www.ece.utexas.edu/~sanqi/papers/ut_abr.ps

[9] Yahagi, T. and Lu, J., "On Self-Tuning Control of Nonminimum Phase Discrete-Time Systems Using approximate Inverse Systems," ASME Journal of Dynamic Systems, Measurement, and Control, Vol. 115, March 1993, p. 12-18.

[10] E. Altman, E. Basar, and R. Srikant, "Robust rate control for ABR sources," *INFOCOM'98*, v 1, p 166-173.

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

[12] Tarrab, M. and Feuer, A., "Convergence and Performance Analysis of the Normalized LMS Algorithm with Uncorrelated Gaussian Data," IEEE Transactions on Information Theory, Vol. 34, No. 4, July 1988, p. 680-691.

[13] Haykin, Simon S., *Adaptive Filter Theory*, Prentice Hall, Englewood Cliffs, NJ, 1991.

[14] Stark, H. and Woods, J. W., *Probability, Random Processes, and Estimation Theory for Engineers* 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, c1994.

[15] Peebles Jr., P. Z., *Probability, Random Variables, and Random Signal Principles, 2nd Edition*, McGraw-Hill, New York, NY, c1987.

[16] G. Strang, *Linear Algebra and Its Applications, 3rd Ed*, Harcourt College Publishers, 1988.

[17] K. Laberteaux and C. Rohrs, "On the Convergence of a Direct Adaptive Controller for ATM ABR Congestion Control," Int Conf Comm 2000, June 2000.

[18] K. Laberteaux, *Explicit Rate Congestion Control for Data Networks*, a Dissertation, Department of Electrical Engineering, the University of Notre Dame, available from author at klaberte@iname.com, 2000.