Frequency Domain Analysis of Tracking and Noise Performance of Adaptive Algorithms

Brett Ninness* Juan Carlos Gómez †

Abstract

In this paper, an analysis of the tracking performance of several adaptive algorithms is carried out for the case of model structures with fixed pole positions. Such structures have recently been proposed as an efficient generalisation of the common FIR model structure. The focus of this work is to analyse the tradeoff between noise sensitivity and tracking ability in the frequency domain by illustrating how it is influenced by such things as input and noise spectral densities, step size, and the choice of the fixed pole locations.

Technical Report EE9656, Department of Electrical and Computer Engineering, University of Newcastle,AUSTRALIA. EDICS Number: SP 2.6.4

1 Introduction

This paper is inspired by recent results [12] that quantify the parameter space performance of adaptive algorithms, by work suggesting the utility of analysis of adaptive algorithms in the frequency domain [6, 20, 2, 3], by work suggesting such analysis be simplified by considering high model orders [14, 15, 9], and by recent work suggesting novel model structures for adaptive algorithms [26, 24].

These model structures are generalisations of the popular FIR structure, but are more flexible in that the poles in the model structure need not all be fixed at the origin. As pointed out in [26, 24], exploiting this flexibility can accrue many advantages in terms of estimation accuracy, while still retaining the desirable convergence properties enjoyed by adaptive FIR schemes.

Following the suggestions in [6, 20, 2, 3], this paper provides a frequency domain analysis of the performance of these 'generalised FIR' methods in order to make explicit how the

*This work was supported by the Australian Research Council and the Centre for Industrial Control Science. This author is with the Department of Electrical and Computer Engineering, University of Newcastle, Australia and can be contacted at email:brett@ee.newcastle.edu.au or FAX: +61 49 21 69 93

†This author gratefully acknowledges the support of the Centre for Industrial Control Science.

tradeoff between noise sensitivity and tracking ability is influenced by input and noise spectral densities, choice of step size, and (what is a main focus of this paper) the choice of fixed pole position. There are close relations between this work and those of Gunnarsson and Ljung [10, 14, 9] who studied adaptive FIR algorithms in the frequency domain. Specifically [10, 9] provide the main idea of this paper which is to simplify error expressions by considering large model order; the results of this paper specialise to some of those in [10, 9] when all the poles are chosen at the origin.

Aside from these ideas, there are two main tools employed in this paper, the most important one being the use of recent results by Guo and Ljung [12] that provide approximations to the parameter covariance for a general class of adaptive algorithms under mild assumptions. The second main tool is to re-parameterise the fixed denominator model structure [26, 24] into an orthonormal form studied in [16] in order to facilitate the theoretical analysis. In fact, a main result of this paper is that this orthonormal parameterisation is an intrinsic part of estimation using fixed denominator model structures. This arises since for recursive least squares (RLS) and Kalman filtering algorithms, the tracking error can be quantified in terms of the orthonormal form whether or not the model structure is originally cast in this form. The nature of this quantification is that it directly shows how pole choices not at the origin affect the tracking and noise performance in the frequency domain.

2 Problem Formulation

This paper considers situations where an observed input sequence $\{u_k\}$ is related to an observed output sequence $\{y_k\}$ according to

$$y_k = G_k(q)u_k + \nu_k \tag{1}$$

where $\{\nu_k\}$ is a zero mean white noise process with variance $\mathbf{E}\{\nu_k^2\} = \sigma_\nu^2 < \infty$ and

$$G_k(q) = \sum_{n=1}^{\infty} g_k(n)q^{-n}$$

is a possibly time varying linear system with impulse response $\{g_k(n)\} \in \ell_2$. It is assumed that $\{u_k\}$ is a realisation of a stationary stochastic process with covariance function $R_u(\tau) = \mathbf{E}\{u_k u_{k-\tau}\}$ and associated spectral density

$$\Phi_u(\omega) = \sum_{\tau=-\infty}^{\infty} R_u(\tau)e^{-j\omega\tau}$$

and that $\{u_k\}$ is weakly uncorrelated with $\{\nu_k\}$ in the sense that $|\mathbf{E}\{u_k\nu_{k-\tau}\}| \to 0$ as $\tau \to \infty$. It is also assumed that $\Phi_u(\omega) > 0$ and that $\Phi_u(\omega)$ has a finite dimensional spectral factorisation.

At issue is the estimation of the (assumed unknown) time varying dynamics $G_k(q)$ by means of the observations $\{u_k\}$ and $\{y_k\}$. There are many approaches to this problem, but a common theme [21, 7, 13] is to express the dependence (1) in a linear regression form

$$y_k = \phi_k^T \theta_k + \nu_k \tag{2}$$

where the 'regression vector' ϕ_k depends on measurements of $\{u_t\}$ and $\{y_t\}$ up until $t = k$ and $\theta_k \in \mathbf{R}^p$ is a vector of p parameters in a model structure $G(q, \theta_k)$ that attempts to describe the true dynamics $G_k(q)$. An estimate of $G_k(q)$ is then obtained as $G(q, \widehat{\theta}_k)$ where the estimate $\widehat{\theta}_k$ is obtained recursively via

$$\widehat{\theta}_{k+1} = \widehat{\theta}_k + L_k(y_k - \phi_k^T \widehat{\theta}_k), \quad \mu \in (0, 1) \tag{3}$$

where L_k is a gain vector that may be computed in various ways. A common choice for this gain vector is

$$L_k = \mu \phi_k, \quad \mu \in (0, 1) \tag{4}$$

in which case (3) is known as the 'gradient' or 'least mean square' (LMS) algorithm. Another common choice is

$$L_k = P_k \phi_k \tag{5}$$

where P_k satisfies

$$P_k = \frac{1}{\lambda} \left\{ P_{k-1} - \frac{P_{k-1}\phi_k \phi_k^T P_{k-1}}{\lambda + \phi_k^T P_{k-1} \phi_k} \right\}; \quad \lambda = 1 - \mu, \mu \in (0, 1) \tag{6}$$

initialised with some positive definite P_0 and with the ensuing algorithm being known as 'Recursive Least Squares' (RLS). The constant λ is known as the 'forgetting factor'. Finally, if the time variation of the parameters θ_k are modeled via a random walk as

$$\theta_{k+1} = \theta_k + \rho \, w_k \tag{7}$$

where w_k is a stationary zero mean vector white noise process with $\mathbf{E}\{w_k w_k^T\} = Q$, then the update law

$$L_k = \frac{\mu P_{k-1} \phi_k}{\sigma^2 + \mu \phi_k^T P_{k-1} \phi_k} \tag{8}$$

where P_k satisfies the Riccati equation

$$P_k = P_{k-1} - \mu \frac{P_{k-1}\phi_k \phi_k^T P_{k-1}}{\sigma^2 + \mu \phi_k^T P_{k-1} \phi_k} + \mu \Sigma \tag{9}$$

with $\Sigma > 0$ and symmetric is known as the Kalman Filter. It is optimal in a mean-square sense under Gaussian assumptions if $\mu = \rho, \sigma = \sigma_\nu$ and $\Sigma = Q$. In the sequel it is assumed that w_k is weakly uncorrelated with $\{u_k\}$ in the sense that $|\mathbf{E}\{u_k w_{k-\tau}\}| \to 0$ as $\tau \to \infty$.

When employing any of these adaptive schemes, a central question is the accuracy of the estimate $G(q, \widehat{\theta}_k)$ as a description of $G_k(q)$. The most common way of assessing this

is to examine the accuracy of $\widehat{\theta}_k$ itself [21, 7]. This may be achieved by defining θ_k as the true parameter vector that allows the model structure to exactly describe the underlying time varying dynamics as $G(q,\theta_k) = G_k(q)$ and by defining the estimation error $\widetilde{\theta}_k$ as

$$\widetilde{\theta}_k \triangleq \theta_k - \widehat{\theta}_k. \tag{10}$$

Of course, in general the model structure $G(q,\theta_k)$ will be too parsimonious to exactly describe the true dynamics $G_k(q)$ for any value of θ_k, and so there will be no 'true' parameter θ_k. However, as pointed out in [14, 15], under some criterion (for example, L_2 error) there is a 'best' value θ_k that can be used in calculating $\widetilde{\theta}_k$, so that in fact

$$G_k(q) = G(q,\theta_k) + \Delta_p(q)$$

where $\Delta_p(q) \to 0$ as $p \to \infty$. Since all the analysis in this paper will involve deriving expressions that are asymptotic in p, there is no point in continually including the $\Delta_p(q)$ in expressions, and so (following [14, 15]) it is ignored at the beginning. The validity of this strategy is confirmed by simulation example in §8, and the rate at which $\Delta_p(e^{j\omega})$ tends to zero with increasing p is established via theorem 6.1 in §6.

In any event, substituting (10) into the general update equation (3) gives that this error satisfies the following difference equation [12]

$$\widetilde{\theta}_{k+1} = \left(I - L_k\phi_k^T\right)\widetilde{\theta}_k + \rho\,\omega_k - L_k\nu_k. \tag{11}$$

The quality of an adaptive estimation scheme can then be quantified by using (11) to calculate the covariance $\mathbf{E}\{\widetilde{\theta}_k\widetilde{\theta}_k^T\}$ as a measure of estimation accuracy. Unfortunately, as pointed out in [10, 12, 9], the exact expression for this covariance will be very complicated except in very special circumstances. The main result of [12] which will be central to the analysis of this paper is that under the stated assumptions, $\mathbf{E}\{\widetilde{\theta}_k\widetilde{\theta}_k^T\}$ may be approximated by Π_k given by the deterministic difference equation

$$\Pi_{k+1} = (I - \mu S_k R)\Pi_k(I - \mu S_k R)^T + \mu^2\sigma_\nu^2 S_k R S_k + \rho^2 Q \tag{12}$$

where $R \triangleq \mathbf{E}\left\{\phi_k\phi_k^T\right\}$ and S_k is defined as

LMS:

$$S_k = I, \tag{13}$$

RLS:

$$S_k = (1+\mu)S_{k-1} - \mu S_{k-1}RS_{k-1}; \quad S_0 = P_0, \tag{14}$$

Kalman Filter:

$$S_k = S_{k-1} - \mu S_{k-1}RS_{k-1} + \frac{\mu}{\sigma^2}\Sigma; \quad S_0 = \frac{1}{\sigma^2}P_0, \tag{15}$$

In [12] the quality of this approximation is quantified as

$$\left\| \mathbf{E}\left\{ \widetilde{\theta}_k \widetilde{\theta}_k^T \right\} - \Pi_k \right\| \leq \kappa(\mu)$$

where $\kappa(\mu)$ is a bounded function that tends to zero as μ tends to zero.

However, as argued in [10, 6, 15, 3], in many cases the interest is not in the accuracy in parameter space, but the accuracy in how close the estimated model $G(q, \widehat{\theta}_k)$ is to the true system $G_k(q)$ in terms of the error

$$\widetilde{G}_k(e^{j\omega}) \triangleq G_k(e^{j\omega}) - G(e^{j\omega}, \widehat{\theta}_k)$$

in the estimated frequency response. In this paper, model structures $G(q, \theta_k)$ are considered for which the estimated frequency response depends linearly on the estimated parameters as

$$G(e^{j\omega}, \widehat{\theta}_k) = \Gamma_p^T(e^{j\omega})\widehat{\theta}_k \tag{16}$$

where

$$\Gamma_p(q) \triangleq [\mathcal{B}_0(q), \mathcal{B}_1(q), \cdots, \mathcal{B}_{p-1}(q)]^T \tag{17}$$

is a vector of p rational transfer functions $\mathcal{B}_n(q)$. For example, $\mathcal{B}_n(q) = q^{-n}$ corresponds to an FIR model structure.

Using (12),(16) and (17), an approximate frequency domain quantification of adaptive performance may then be taken as

$$\mathbf{E}\left\{ |\widetilde{G}_k(e^{j\omega})|^2 \right\} = \Gamma_p^\star(e^{j\omega}) \mathbf{E}\left\{ \widetilde{\theta}_k \widetilde{\theta}_k^T \right\} \Gamma_p(e^{j\omega}) \approx \Gamma_p^\star(e^{j\omega}) \Pi_k \Gamma_p(e^{j\omega}) \tag{18}$$

where $\cdot^\star$ denotes 'conjugate transpose'. Unfortunately, again this expression will in general be of a very complicated nature. The main contribution of this paper will be to follow the lead of [10, 15, 9] and derive simple approximations for (18) that are increasingly accurate for increasing model order p. These simplified expressions make clear how factors such as measurement noise variance, input spectral density, and (what is the novel part of this work) choice of fixed pole location affect $\mathbf{E}\{|\widetilde{G}_k(e^{j\omega})|^2\}$.

In order to proceed with this, it is necessary to be more explicit as to the exact nature of the model structures $G(q, \theta_k)$ being considered, and hence be more specific on the exact formulation of the transfer functions $\{\mathcal{B}_n(q)\}$.

3 Model Structures

The model structures examined in this paper have recently been proposed and examined in an adaptive filtering context by Williamson and co–workers in a series of papers [26, 24, 23, 25] where they have been termed 'fixed pole adaptive filters'. They are formulated as

$$G(q, \theta_k') = \left[\prod_{n=0}^{p-1} (q - \xi_n) \right]^{-1} \sum_{n=0}^{p-1} \theta_k'(n) q^n \tag{19}$$

where the poles $\{\xi_n\}$ are fixed according to prior information about the likely pole positions of the true time varying system $G_k(q)$. A special case of this structure arises when all the poles $\{\xi_n\}$ are chosen at the origin in which case (19) is an FIR model structure.

However, empirical evidence [26, 4, 18] supports the fact that in an adaptive filtering context, a significant improvement in estimation accuracy is possible by avoiding poles all fixed at the origin, and instead distributing them in the unit disk so as to be as close as possible to the true poles of $G_k(q)$. Theoretical analysis supporting these empirical observations is provided via theorem 6.1 in §6.

Additionally, although the model structure (19) is an IIR structure, it does not imply the convergence and noise induced bias difficulties often associated with adaptive IIR filtering [19]. Instead, it leads to adaptive algorithms enjoying all the same guaranteed convergence properties known to hold for adaptive FIR schemes [26].

However, in spite of the pleasant properties enjoyed by the model structure (19), its generality (as compared to an FIR structure) makes frequency domain analysis of adaptive algorithms much more difficult. To be more specific, with an FIR structure, the frequency response of the estimated model can be considered to be a linear combination of the 'basis functions' $\{1, e^{-j\omega}, \cdots, e^{-j(p-1)\omega}\}$ used in classical Fourier analysis. Indeed, (as shown in appendix A) the expression (18) is the Cèsaro mean of the Fourier series of a certain function, and there is a large body of work studying the convergence of such Cèsaro means [5].

Furthermore, since these FIR 'basis functions' enjoy the group structure $e^{-j\omega n}e^{-j\omega m} = e^{-j\omega(m+n)}$, then for large k the parameter covariance matrix approximation Π_k is Toeplitz, and by drawing on the wealth of literature on such matrices [8, 1] it is possible to determine the function for which (18) is a (partial) Cèsaro mean Fourier reconstruction. These are the main tools used in [10, 15].

Unfortunately, for the more general model structure (19), all these properties are lost, making the instructive frequency domain expressions presented in [10, 15] less straightforward to derive. The contribution of this paper is to overcome these difficulties by replacing the model structure (19) with the following formulation

$$G(q,\theta_k) = \sum_{n=0}^{p-1} \theta_k(n)\mathcal{B}_n(q) \tag{20}$$

where

$$\mathcal{B}_n(q) = \left(\frac{\sqrt{1-|\xi_n|^2}}{q-\xi_n}\right)\prod_{k=0}^{n-1}\left(\frac{1-\overline{\xi_k}q}{q-\xi_k}\right). \tag{21}$$

This model structure (20) is easily cast in the linear regression form (2) with

$$\begin{aligned}\phi_k &\triangleq \Gamma_p^T(q)u_k = [\mathcal{B}_0(q)u_k, \mathcal{B}_1(q)u_k, \cdots, \mathcal{B}_{p-1}(q)u_k],\\ \theta_k &\triangleq [\theta_k(0), \theta_k(1), \cdots, \theta_k(p-1)]^T.\end{aligned}$$

In this case, the quantity $R = \mathbf{E}\left\{\phi_k\phi_k^T\right\}$ can be calculated using Parseval's theorem as

$$R = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Gamma_p(e^{j\omega})\,\Gamma_p^\star(e^{j\omega})\,\Phi_u(\omega)\,\mathrm{d}\omega, \tag{22}$$

and this latter formulation will be particularly useful in the sequel for the purposes of characterising estimation errors in terms of spectral densities. Note also that the generation of the regression vector ϕ_k corresponding to both model structures (19) and (20) may be unified by the formulation

$$\phi_{k+1} = A\phi_k + Bu_k, \tag{23}$$

where for both structures (19) and (20) the matrix A is invariant with eigenvalues at $\{\xi_0, \xi_1, \cdots \xi_{p-1}\}$, but the vector B is different.

A key point is that since the poles of the model structure (20) and (19) are identical, then they are equivalent in the sense that for some nonsingular $J \in \mathbf{R}^{p\times p}$, the parameter vectors θ_k in (20) and $\theta'_k = J\theta_k$ in (19) describe exactly the same transfer function. As well, with initialisation $P_0 = J^{-1}P'_0J^{-T}$ consistent with this linear re–parameterisation, the RLS updates (5),(6) are invariant to the re–parameterisation in the sense that $\widehat{\theta}'_k = J\widehat{\theta}_k$ so that frequency response estimates are identical: $G(e^{j\omega}, \widehat{\theta}'_k) = G(e^{j\omega}, \widehat{\theta}_k)$. This same property also applies to the Kalman Filtering update law (8),(9) provided the compatibility $\Sigma = J^{-1}\Sigma' J^{-1}$ is also maintained.

However, an important feature of the 'basis functions' $\{\mathcal{B}_n(q)\}$ in (20) is that they are orthonormal with respect to the inner product

$$\langle \mathcal{B}_n, \mathcal{B}_m \rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathcal{B}_n(e^{j\omega})\overline{\mathcal{B}_m(e^{j\omega})}\,\mathrm{d}\omega = \delta(n-m)$$

and this crucial property allows Cèsaro mean convergence results to be extended to generalised Fourier expansions with respect to the basis $\{\mathcal{B}_n(e^{j\omega})\}$ (this is done in theorem A.1). This key result allows frequency domain results specific to FIR model structures [10, 14] to be extended to the wider class of fixed denominator model structures (20). Since (as just discussed) the RLS and Kalman Filtering algorithms are frequency domain invariant to linear model re–parameterisations, this strategy of analysing an orthonormal structure provides results for any fixed denominator model structure such as (19).

Unfortunately, the LMS algorithm is not invariant under linear re–parameterisations [6, 22], and so any subsequent results for this algorithm will only pertain to the orthonormal model structure (20). This is not considered a significant limitation, since as detailed in §7, for the LMS algorithm the orthonormal structure (20) is attractive from the viewpoint of numerical robustness and enhanced convergence rate under white noise excitation.

4 Transient analysis

In this section, the transient behavior of the frequency response estimation error is studied by using the approximation (18) substituted into (12) to obtain

$$\begin{aligned}\mathbf{E}\left\{|\widetilde{G}_{k+1}(e^{j\omega})|^2\right\} \;\approx\; & \Gamma_p^\star(e^{j\omega})(I-\mu S_k R)\Pi_k(I-\mu S_k R)^T\Gamma_p(e^{j\omega}) + \\ & \mu^2\sigma_\nu^2\Gamma_p^\star(e^{j\omega})S_kRS_k\Gamma_p(e^{j\omega}) + \rho^2\Gamma_p^\star(e^{j\omega})\,Q\,\Gamma_p(e^{j\omega}).\end{aligned} \tag{24}$$

Now, as mentioned at the end of §2, this is a very complicated expression from which it is difficult to extract useful design insights. However, if the model order p is assumed large, then very significant simplifications arise in (24). In order to present them, it is first necessary to define the quantity

$$\gamma_p(\omega) \triangleq \sum_{n=0}^{p-1} |\mathcal{B}_n(e^{j\omega})|^2. \tag{25}$$

It is common knowledge that the sensitivity of the accuracy of estimated models to measurement noise is proportional to the model order p, and this has been rigorously justified for FIR model structures [15, 10]. The quantity $\gamma_p(\omega)$ will serve to capture how this phenomenon generalises to the fixed–denominator model structures (19),(20) in a frequency dependent fashion, as illustrated in figure 1. Note from this diagram, that for $\xi_n = 0$ so that (19) is FIR, then $\gamma_p(\omega) = p$.

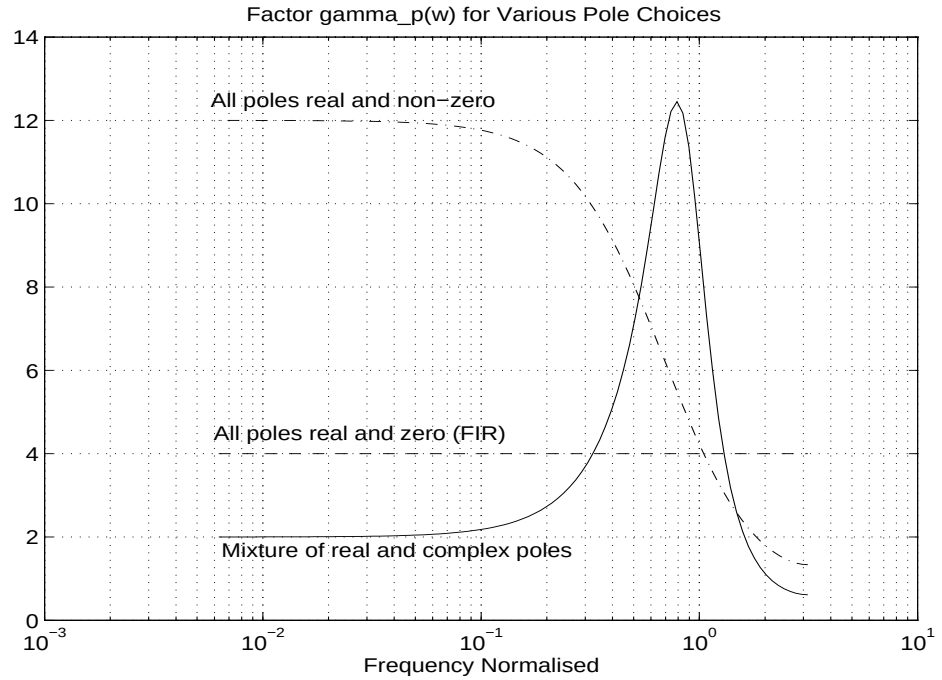

Figure 1: *Plot of* $\gamma_p(\omega) = \sum_{k=0}^{p-1} |\mathcal{B}_k(e^{j\omega})|^2$ *for* $p = 4$ *and for various choices of pole location.*

It is also necessary to define a quantity which reflects the time varying nature of the system $G_k(q)$ in the frequency domain, and is also commensurate with the parameter space model for this (7). Following [10], this may be achieved by using the linear relationship (16) in combination with (7) to obtain the model for the time variation as

$$G_{k+1}(e^{j\omega}) = G_k(e^{j\omega}) + \rho\, \Gamma_p^T(e^{j\omega}) w_k$$

so that

$$\mathbf{E}\left\{|G_{k+1}(e^{j\omega}) - G_k(e^{j\omega})|^2\right\} \triangleq \rho^2 \delta_p(\omega) = \rho^2 \Gamma_p^\star(e^{j\omega}) Q \Gamma_p(e^{j\omega}). \tag{26}$$

Now, in the sequel, simplifications will be obtained by considering growing model order p, and to facilitate this it is necessary (as in [10]) to assume that as p grows, Q also grows in such a way as to always be positive definite, but of bounded norm. In this case, the limit

$$\delta(\omega) \triangleq \lim_{p\to\infty} \frac{\delta_p(\omega)}{\gamma_p(\omega)}$$

exists and is non-zero. Given these definitions, the following theorem provides a simple frequency domain characterisation of the tracking characteristics of the LMS adaptation scheme when using the general fixed denominator model structure (19).

Theorem 4.1. *For the LMS algorithm and the model structure (20), then using the approximation (12)*

$$\lim_{p\to\infty} \left| \frac{1}{\gamma_p(\omega)} \mathbf{E}\left\{ |\widetilde{G}_{k+1}(e^{j\omega})|^2 \right\} - \left(\frac{[1-\mu\Phi_u(\omega)]^2}{\gamma_p(\omega)} \mathbf{E}\left\{ |\widetilde{G}_k(e^{j\omega})|^2 \right\} + \mu^2\sigma_\nu^2\Phi_u(\omega) + \rho^2\delta(\omega) \right) \right| = 0$$

Proof. Using the formulation (22) together with the notation (A.1) in (24) with $S_k = I$ gives

$$\begin{aligned} \frac{1}{\gamma_p(\omega)} \mathbf{E}\left\{ |\widetilde{G}_{k+1}(e^{j\omega})|^2 \right\} &= \frac{1}{\gamma_p(\omega)} \Gamma_p^\star(e^{j\omega}) M_p (1-\mu\Phi_u) \Pi_k M_p (1-\mu\Phi_u) \Gamma_p(e^{j\omega}) + \\ &\quad \frac{1}{\gamma_p(\omega)} \mu^2\sigma_\nu^2 \Gamma_p^\star(e^{j\omega}) M_p(\Phi_u) \Gamma_p(e^{j\omega}) + \frac{\rho^2\delta_p(\omega)}{\gamma_p(\omega)}. \end{aligned}$$

Use of Theorem A.1 and lemma B.1 then gives the result. □

The interpretation of this result is that for large model order p, the frequency domain tracking performance of the LMS algorithm for model structures with fixed poles may be approximated by the dynamics of the following first order difference equation

$$\mathbf{E}\left\{ |\widetilde{G}_{k+1}(e^{j\omega})|^2 \right\} \approx [1-\mu\Phi_u(\omega)]^2 \mathbf{E}\left\{ |\widetilde{G}_k(e^{j\omega})|^2 \right\} + \mu^2\sigma_\nu^2\Phi_u(\omega)\gamma_p(\omega) + \rho^2\delta(\omega)\gamma_p(\omega). \quad (27)$$

This is similar to the results presented in [10, 9, 6], and indeed, for the FIR cases studied there, the expression (27) with $\xi_n = 0$ so that $\gamma_p(\omega) = p$ reduces to the expressions in those works. However, there is a significant difference between (27) and the expressions of [10, 9, 6], in that the term $\gamma_p(\omega)$ in (27) captures how the tracking error is affected by the choice of poles in the model structures (19) and (20).

To be more specific, as in [10, 9], (27) shows that for the LMS algorithm the tracking error decays like $[1-\mu\Phi_u(\omega)]^{2k}$, so that tracking is likely to be better at frequencies where $\mu\Phi_u(\omega)$ is large. However, unlike the expressions in [10, 9], (27) shows that the 'noise driving' term $\sigma_\nu^2\mu^2\Phi_u(\omega)\gamma_p(\omega)$, which in a trade-off fashion is also large at frequencies where $\mu\Phi_u(\omega)$ is large, is modulated by the term $\gamma_p(\omega)$ which depends on the chosen pole positions as shown in figure 1.

Of course, an important question is the accuracy of the approximation (27) for the low model orders likely to be used in practice. This is examined in §8 where for small enough step size μ, (27) is shown to faithfully predict the frequency domain convergence properties of the LMS algorithm – see figure 8.

5 Steady State Analysis

Using similar techniques as employed in the previous section, it is also possible to quantify the steady state behavior of the frequency response estimation error. The latter is defined to be the error $\mathbf{E}\{|\widetilde{G}_k(e^{j\omega})|^2\}$ for large k, and in order for it to be investigated, it is necessary to examine the behavior of the solution Π_k of (12) for large k; that is

$$\lim_{k\to\infty} \Pi_k \triangleq \Pi.$$

Of course, for this limit to exist (and indeed for the approximation (27) to hold) it is necessary that the adaptive algorithm (3) be stable. This is established in [11] where under the stated assumptions of this paper, for any of the choices (4),(5) or (6) the algorithm (3) is shown to be exponentially stable with $\|\Pi\|_2 < \infty$.

With this stability in hand, Π may be evaluated by determining the steady state solutions $S \triangleq \lim_{k\to\infty} S_k$ of (13)–(15) and then substituting them into (27) before examining its steady own state solution with terms of order $\mu^2\Pi$ discarded [12]. The results of this strategy are as follows.

LMS: Here $S = I$ so that Π is the solution of the Lyapunov equation

$$\Pi R + R\Pi = \mu\sigma_\nu^2 R + \frac{\rho^2}{\mu} Q. \tag{28}$$

RLS: Here $S = R^{-1}$ so that Π is given by

$$\Pi = \frac{\mu\sigma_\nu^2}{2} R^{-1} + \frac{\rho^2}{2\mu} Q. \tag{29}$$

Kalman Filter: This case is more difficult. S is the solution of

$$\sigma^2 SRS = \Sigma$$

so that Π is given by the solution of

$$SR\Pi + \Pi RS = \frac{\mu\sigma_\nu^2}{\sigma^2}\Sigma + \frac{\rho^2}{\mu} Q. \tag{30}$$

For the special case of $\Sigma = Q$ this system has solution

$$\Pi = \frac{\sigma^2}{2}\left(\mu\frac{\sigma_\nu^2}{\sigma^2} + \frac{\rho^2}{\mu}\right) S. \tag{31}$$

Using these characterisations of Π together with (18) it is possible to quantify the steady state estimation error in the frequency domain as

$$\mathbf{E}\left\{|\widetilde{G}(e^{j\omega})|^2\right\} \triangleq \lim_{k\to\infty} \mathbf{E}\left\{|\widetilde{G}_k(e^{j\omega})|^2\right\} \approx \Gamma_p^\star(e^{j\omega})\Pi\Gamma_p(e^{j\omega}). \tag{32}$$

As per the previous section, although this provides a quantifiable performance measure, the resulting expression is so complicated that it is difficult to extract useful design insight from it. However, again following the lead of the previous section, if the model order is assumed large, then the expression (32) can be significantly simplified so as to clearly illustrate how the estimation error is related to step size, measurement noise energy, input spectral densities, and perhaps most interestingly, the choice of fixed pole position. The effect of this latter choice is quantified by the term $\gamma_p(\omega)$. For the case of the LMS algorithm, the simplified error quantification is as follows.

Theorem 5.1. *For the LMS algorithm and the model structure (20), then for large model order p,*

$$\lim_{p\to\infty} \frac{1}{\gamma_p(\omega)} \mathbf{E}\left\{|\widetilde{G}(e^{j\omega})|^2\right\} = \frac{1}{2}\left[\mu\sigma_\nu^2 + \frac{\rho^2\delta(\omega)}{\mu\Phi_u(\omega)}\right]$$

Proof. Using the formulation (22) together with the notation (A.1) in (28) gives that in the limit as $k \to \infty$

$$\frac{\Gamma_p^\star(e^{j\omega})\Pi M_p(\Phi_u)\Gamma_p(e^{j\omega})}{\gamma_p(\omega)} + \frac{\Gamma_p^\star(e^{j\omega})M_p(\Phi_u)\Pi\Gamma_p(e^{j\omega})}{\gamma_p(\omega)} = \frac{\mu\sigma_\nu^2\Gamma_p^\star(e^{j\omega})M_p(\Phi_u)\Gamma_p(e^{j\omega})}{\gamma_p(\omega)} + \frac{\rho^2\delta_p(\omega)}{\mu\gamma_p(\omega)}$$

Taking the limit of both sides as $p \to \infty$ while using Theorem A.1 and Lemma B.2 then gives the result. □

The interpretation of this theorem is that for large model order p, and after the algorithm has converged (large k)

$$\mathbf{E}\left\{|\widetilde{G}_k(e^{j\omega})|^2\right\} \approx \frac{\gamma_p(\omega)}{2}\left[\mu\sigma_\nu^2 + \frac{\rho^2\delta(\omega)}{\mu\Phi_u(\omega)}\right]. \tag{33}$$

For the case of all the poles $\{\xi_n\}$ in the model structure $G(q,\theta)$ chosen at the origin, then $\gamma_p(\omega) = p$ and the above expression specialises to that derived in [10, 9]. However, as illustrated in figure 1, the factor $\gamma_p(\omega)$ in the expression (33) shows how pole choices other than FIR influence the frequency domain estimation error. As well, (33) illustrates that for time invariant systems ($\rho = 0$), then the error is proportional to the step size μ and the measurement noise variance σ_ν^2, while for time varying systems ($\rho \neq 0$), another error component arises, is inversely proportional to step size, and is also inversely proportional to input spectral density $\Phi_u(\omega)$.

Of course, a fundamental question is the reliability of using the approximation (33) for practically useful (and hence finite) model orders, given that it is obtained from a result that is asymptotic in p. The most suitable way to deal with this issue would be to quantify the convergence rate in theorem 5.1. This appears to be extremely difficult. Instead, the approach used in [10] is taken wherein the validity of (33) for finite p is examined via a simulation study. This is done in §8 , where it is shown (see figures 4 and 6) that for a tenth order model, (33) is quite an accurate approximation; in fact, as shown in figure 5, this holds even for as low as a fourth order model.

This same sort of analysis can also be performed for the RLS algorithm, with the results as follows.

Theorem 5.2. *For the RLS algorithm and the model structure (19) or (20), then*

$$\lim_{p\to\infty}\frac{1}{\gamma_p(\omega)}\mathbf{E}\left\{|\widetilde{G}(e^{j\omega})|^2\right\}=\frac{1}{2}\left[\frac{\mu\sigma_\nu^2}{\Phi_u(\omega)}+\frac{\rho^2}{\mu}\delta(\omega)\right]\quad ;\mu=1-\lambda.$$

Proof. Using the formulation (22) together with the notation (A.1) in (29) gives that

$$\frac{1}{\gamma_p(\omega)}\mathbf{E}\left\{|\widetilde{G}(e^{j\omega})|^2\right\}=\frac{\Gamma_p^\star(e^{j\omega})\Pi\Gamma_p(e^{j\omega})}{\gamma_p(\omega)}=\frac{\mu\sigma_\nu^2}{2\gamma_p(\omega)}\Gamma_p^\star(e^{j\omega})M_p^{-1}(\Phi_u)\Gamma_p(e^{j\omega})+\frac{\rho^2\delta_p(\omega)}{2\mu\gamma_p(\omega)}$$

Taking the limit of both sides as $p\to\infty$ while using Theorem B.1 then gives the result. □

Following the example of the previous theorem, the interpretation of this theorem is that for large model order p, and after the algorithm has converged (large k)

$$\mathbf{E}\left\{|\widetilde{G}_k(e^{j\omega})|^2\right\}\approx\frac{\gamma_p(\omega)}{2}\left[\frac{\mu\sigma_\nu^2}{\Phi_u(\omega)}+\frac{\rho^2}{\mu}\delta(\omega)\right]. \tag{34}$$

Comparing this approximation to (33) illustrates a fundamental difference between the steady state behavior of LMS and RLS in terms of how the input spectral density affects the noise and tracking performance. Specifically, (34) shows that for stationary systems ($\rho=0$), the RLS estimation error is inversely proportional to the input spectral density $\Phi_u(\omega)$, while (33) shows that the LMS estimation error is invariant to the size of this spectral density; see figure 6 for simulation study validation of this phenomenon. Conversely when $\rho\neq 0$, (33) shows that for LMS the tracking ability increases with increasing input spectral density, whilst for RLS the tracking ability is invariant to this factor, and only depends on step size.

These observations, in the context of FIR model structures have already been made in [10, 9], and as per the LMS case, when all the poles are chosen at the origin and hence $\gamma_p(\omega)=p$, then (34) is identical to the expressions for RLS steady state error presented in [10, 9]. However, again as per the LMS case, the inclusion of the frequency dependent factor $\gamma_p(\omega)$ in (34) shows how the choice of fixed denominator pole position in the model structure $G(q,\theta)$ affects the estimation error in the frequency domain (figure 1).

The question of the validity of using an asymptotic result as a finite data and model order approximation in (34) again arises, and again this is dealt with in §8 via a simulation study; for example in figure 5, (34) is shown to be quite accurate even for only a fourth order model and 300 data points.

To complete the analysis, the following theorem quantifies Kalman Filter performance via the same strategy of considering large model order. However, as already mentioned, there are particular difficulties in solving for the steady state parameter covariance Π, and this leads to the treatment of only a specialised case in which $Q=\Sigma$.

Theorem 5.3. *For the Kalman Filtering algorithm, the model structure (19) or (20) and under the assumption that $Q=\Sigma$, then*

$$\lim_{p\to\infty}\frac{1}{\gamma_p(\omega)}\mathbf{E}\left\{|\widetilde{G}(e^{j\omega})|^2\right\}=\frac{1}{2}\left(\mu\frac{\sigma_\nu^2}{\sigma^2}+\frac{\rho^2}{\mu}\right)\sqrt{\frac{\sigma^2\delta(\omega)}{\Phi_u(\omega)}}$$

Proof. Substituting (31) into (30) with $\Sigma = Q$ gives that Π is the solution of

$$\Pi R \Pi = \frac{\sigma^2}{4}\left(\mu\frac{\sigma_\nu^2}{\sigma^2} + \frac{\rho^2}{\mu}\right)^2 Q. \tag{35}$$

Define two positive definite, $p \times p$ dimensional real matrices A_p and B_p to be asymptotically equivalent $A_p \sim B_p$ if they leave quadratic forms with $\Gamma_p(e^{j\omega})$ invariant:

$$A_p \sim B_p \Leftrightarrow \lim_{p\to\infty} \frac{1}{\gamma_p(\omega)}\Gamma_p^\star(e^{j\omega})(A_p - B_p)\Gamma_p(e^{j\omega}) = 0.$$

Using the definition $\alpha^2 \triangleq \sigma^2/4(\mu\sigma_\nu^2/\sigma^2 + \rho^2/\mu)^2$ and the notation (A.1), the matrix $\alpha M_p(\sqrt{\delta/\Phi_u})$ is an asymptotically equivalent solution to Π given by (35) since using the representation (22)

$$\frac{1}{\gamma_p(\omega)}\Gamma_p^\star(e^{j\omega})\left(\Pi R\Pi - \alpha^2 M_p(\sqrt{\delta/\Phi_u})M_p(\Phi_u)M_p(\sqrt{\delta/\Phi_u})\right)\Gamma_p(e^{j\omega}) =$$

$$\frac{\alpha^2}{\gamma_p(\omega)}\Gamma_p^\star(e^{j\omega})\left(M_p(\delta) - M_p(\sqrt{\delta/\Phi_u})M_p(\Phi_u)M_p(\sqrt{\delta/\Phi_u})\right)\Gamma_p(e^{j\omega}) + \tag{36}$$

$$\frac{\alpha^2}{\gamma_p(\omega)}\Gamma_p^\star(e^{j\omega})\left(Q - M(\delta)\right)\Gamma_p(e^{j\omega}). \tag{37}$$

Now, considering the term (36), by theorem A.1 and Lemma C.4

$$\lim_{p\to\infty}\frac{1}{\gamma_p(\omega)}\Gamma_p^\star(e^{j\omega})\left(M_p(\delta) - M_p(\sqrt{\delta/\Phi_u})M_p(\Phi_u)M_p(\sqrt{\delta/\Phi_u})\right)\Gamma_p(e^{j\omega}) =$$

$$\delta(\omega) - \sqrt{\frac{\delta(\omega)}{\Phi_u(\omega)}}\Phi_u(\omega)\sqrt{\frac{\delta(\omega)}{\Phi_u(\omega)}} = 0.$$

Considering the term (37), using the definition of $\delta(\omega)$ and theorem A.1

$$\lim_{p\to\infty}\frac{1}{\gamma_p(\omega)}\Gamma_p^\star(e^{j\omega})\left(Q - M(\delta)\right)\Gamma_p(e^{j\omega}) = \delta(\omega) - \delta(\omega) = 0.$$

Therefore, since $\Pi \sim \alpha M_p(\sqrt{\delta(\omega)/\Phi_u(\omega)})$, then again by theorem A.1

$$\begin{aligned}\lim_{p\to\infty}\frac{1}{\gamma_p(\omega)}\mathbf{E}\left\{|\widetilde{G}(e^{j\omega})|^2\right\} &= \lim_{p\to\infty}\frac{1}{\gamma_p(\omega)}\Gamma_p^\star(e^{j\omega})\Pi\Gamma_p(e^{j\omega}) \\ &= \lim_{p\to\infty}\frac{\alpha}{\gamma_p(\omega)}\Gamma_p^\star(e^{j\omega})M_p(\sqrt{\delta/\Phi_u})\Gamma_p(e^{j\omega}) = \alpha\sqrt{\frac{\delta(\omega)}{\Phi_u(\omega)}}.\end{aligned}$$

□

In sympathy with previous results, the interpretation of this theorem is that for large model order p and after the algorithm has converged (large k) then

$$\mathbf{E}\left\{|\widetilde{G}_k(e^{j\omega})|^2\right\} \approx \frac{\gamma_p(\omega)}{2}\left(\mu\frac{\sigma_\nu^2}{\sigma^2}+\frac{\rho^2}{\mu}\right)\sqrt{\frac{\sigma^2\delta(\omega)}{\Phi_u(\omega)}}. \tag{38}$$

In terms of how input spectral density affects noise sensitivity and tracking ability, (38) shows that Kalman Filter based algorithms sit between the LMS and RLS algorithms in that instead of being affected separately, both noise and tracking performance are affected equally (but to a lesser extent due to the $\sqrt{\cdot}$ operation–see figure 6) by the size of the input spectral density $\Phi_u(\omega)$. Again, the accuracy of the approximation (38) is validated experimentally in §8 to show that in fact it is meaningful for low model orders; see figure 5.

Note that the approximations (33), (34) and (38) can also be used to calculate the optimal step size μ which will minimise the noise sensitivity at a particular frequency, with the results being $\mu = (\rho/\sigma_\nu)\sqrt{\delta(\omega)/\Phi_u(\omega)}$, $\mu = (\rho/\sigma_\nu)\sqrt{\delta(\omega)\Phi_u(\omega)}$, and $\mu = \rho$ for the LMS, RLS and Kalman Filter cases respectively. For all of these cases, the minimal noise variance is $\mathbf{E}\{|\widetilde{G}_k(e^{j\omega})|^2\} = \rho\,\sigma_\nu\gamma_p(\omega)\sqrt{\delta(\omega)/\Phi_u(\omega)}$.

The ubiquity of the term $\gamma_p(\omega)$ in all these error quantifications shows that the orthonormal parameterisation (20),(21) is more than just an essential tool for the analysis of general fixed denominator model structures. Instead, the orthonormal 'basis functions' $\{\mathcal{B}_n(q)\}$ appear as an intrinsic part of adaptive estimation with any fixed denominator model structure $G(q,\theta)$.

For example, for RLS and Kalman Filtering schemes, then in steady–state whether or not the fixed denominator model structure is ab–initio cast in the orthonormal form (20), the complete contribution of the fixed pole choice to the frequency domain error properties is captured by the term $\gamma_p(\omega)$ which via (25) is itself completely described by the orthonormal basis $\{\mathcal{B}_n(q)\}$.

In other words, for any model structure which permits the state space description (23) for the regressor update, the RLS and Kalman Filter frequency domain error quantification depends only on the eigenvalues $\{\xi_n\}$ of the matrix A (and is independent of the choice of the vector B) and is quantified via the orthonormal basis functions $\{\mathcal{B}_n(q)\}$ in $\gamma_p(\omega)$.

6 Modeling accuracy

A key premise behind the idea of using a fixed denominator model structure such as (19) or its orthonormal re–parameterisation (20) is the belief that modeling accuracy may be improved (over the FIR case) by the appropriate choice of fixed poles $\{\xi_n\}$, see [26, 24, 23, 25].

In these latter works, and in others [4, 18] considerable empirical evidence is presented to support this intuitively reasonable motivation. The purpose of this section is to re-enforce this via theoretical analysis, the key component of which is the following theorem.

Theorem 6.1. *Suppose $G(z)$ has partial fraction expansion*

$$G(z) = \sum_{i=0}^{r-1} \frac{\alpha_i}{z - \gamma_i}$$

where all the poles satisfy $|\gamma_i| < 1$. Put $\widehat{G}_p(z)$ as the best L_2 approximation to $G(z)$ with respect to any model structure $G(z, \theta)$ with poles fixed at $\{\xi_0, \xi_1, \cdots, \xi_{p-1}\}$. Then

$$|G(e^{j\omega}) - \widehat{G}_p(e^{j\omega})| \leq \sum_{i=0}^{r-1} \left| \frac{\alpha_i}{e^{j\omega} - \gamma_i} \right| \prod_{k=0}^{p-1} \left| \frac{\gamma_i - \xi_k}{1 - \overline{\xi_k}\gamma_i} \right|. \tag{39}$$

Proof. See [17]. □

This indicates that if in fact one pole ξ_k is included for each pole γ_i in the true system, then the modeling error is zero. This intuitively reasonable fact has already been pointed out in [26]. However, theorem 6.1 goes further than this by showing that the nature of the dependence of the modeling error to the error between the true poles $\{\gamma_i\}$ and the fixed denominator choices $\{\xi_k\}$ is governed by the term

$$\prod_{k=0}^{p-1} \left| \frac{\gamma_i - \xi_k}{1 - \overline{\xi_k}\gamma_i} \right|.$$

This illustrates that the modeling error decreases geometrically in model order p, and at a rate determined by the distance $|\gamma_i - \xi_k|$ between the true poles $\{\gamma_i\}$ and approximations of them $\{\xi_k\}$. Clearly then, there is scope for dramatic increase in modeling accuracy by moving beyond the FIR choice of all poles at the origin in an attempt to decrease $|\gamma_i - \xi_k|$.

While this result lends theoretical support to the strategy of trying to 'guess' the poles γ_i of the system being estimated, it is necessary to also point out that under assumptions of a stationary system ($\rho = 0$), then with step size μ tending to zero it is possible to show [9] that for all of the LMS, RLS and Kalman Filtering algorithms

$$\lim_{k\to\infty} \widehat{\theta}_k = \arg\min_{\theta \in \mathbf{R}^p} \int_{-\pi}^{\pi} |G(e^{j\omega}) - G(e^{j\omega}, \theta)|^2 \Phi_u(\omega)\, \mathrm{d}\omega \quad \text{w.p.1}$$

so that theorem 6.1 quantifies the achieved asymptotic under–modeling induced error when the input excitation $\{u_k\}$ is white.

7 Convergence Rates and Numerical Conditioning

The convergence properties of the general learning algorithm (3) are determined by the autonomous part of (11) as

$$\widetilde{\theta}_{k+1} = (I - L_k \phi_k^T)\widetilde{\theta}_k.$$

For the LMS algorithm where $L_k = \mu\phi_k$, then this suggests [3, 21] that the convergence (averaged over an ensemble of input realisations) is determined by the choice of μ combined with the eigenvalues of $R = \mathbf{E}\{\phi_k\phi_k^T\}$. Specifically, in order to guarantee stability it is necessary that the step–size μ be limited as

$$\frac{1}{\mu} > \lambda_{\max}(R)$$

in which case the parameter space convergence in the direction of the eigenvector associated with $\lambda_{\min}(R)$ is limited to be no faster than the rate of decay of

$$\left[1 - \frac{\lambda_{\min}(R)}{\lambda_{\max}(R)}\right]^k.$$

Therefore, in the interests of LMS convergence speed, it is desirable to choose a model structure $G(q,\theta)$ such that the ensuing regressor ϕ_k implies the minimum possible condition number for R.

For the case of white input spectrum $\Phi_u(\omega) = \text{constant}$, the orthonormal model structure (20) achieves this, and is therefore optimal (in the white input convergence sense) among the class of all fixed denominator model structures, of which the more 'natural' or obvious choice (19) is a member.

This optimality is illustrated via simulation in figure 2. The top (slowly converging) plot is the observed mean square error (averaged over 500 simulations with different input and measurement noise realisations) obtained when using the LMS algorithm with the 'natural' fixed–denominator model structure (19). The bottom (quickly converging) plot is the observed mean square error using the orthonormal model structure (20) and the same LMS algorithm, but with μ changed to keep the steady state error invariant to the change in model structure; the convergence rate improvement obtained by using the orthonormal structure is clear. In both cases the model structure was third order with poles chosen at $\xi_0 = 0.4$, $\xi_1 = 0.85$, $\xi_2 = 0.6$ and the true system $G_k(q)$ was time invariant with true poles at $\gamma_0 = 0.9$ and $\gamma_1 = 0.37$. The output was corrupted by white Gaussian distributed noise of variance $\sigma_\nu^2 = 0.01$, and the input was white Gaussian distributed noise of variance $\Phi_u(\omega) = 10$.

For the case of non–white $\Phi_u(\omega)$, this convergence–rate optimality of the orthonormal model structure (20) is lost, but a feature retained by the orthonormal structure (20) is that a bound on the numerical conditioning of R in terms of the properties of $\Phi_u(\omega)$ may be derived.

Lemma 7.1.

$$\min_{\omega\in[-\pi,\pi]} \Phi_u(\omega) \le \lambda(R) \le \max_{\omega\in[-\pi,\pi]} \Phi_u(\omega).$$

Proof. Take $x \in \mathbf{R}^p$ arbitrary such that $x^T x = 1$. Then using the formulation (22) for R gives

$$x^T R x = \frac{1}{2\pi}\sum_{r=0}^{p-1}\sum_{k=0}^{p-1} x_r\overline{x_k}\int_{-\pi}^{\pi} \mathcal{B}_r(e^{j\omega})\overline{\mathcal{B}_k(e^{j\omega})}\Phi_u(\omega)\,\mathrm{d}\omega, \tag{40}$$

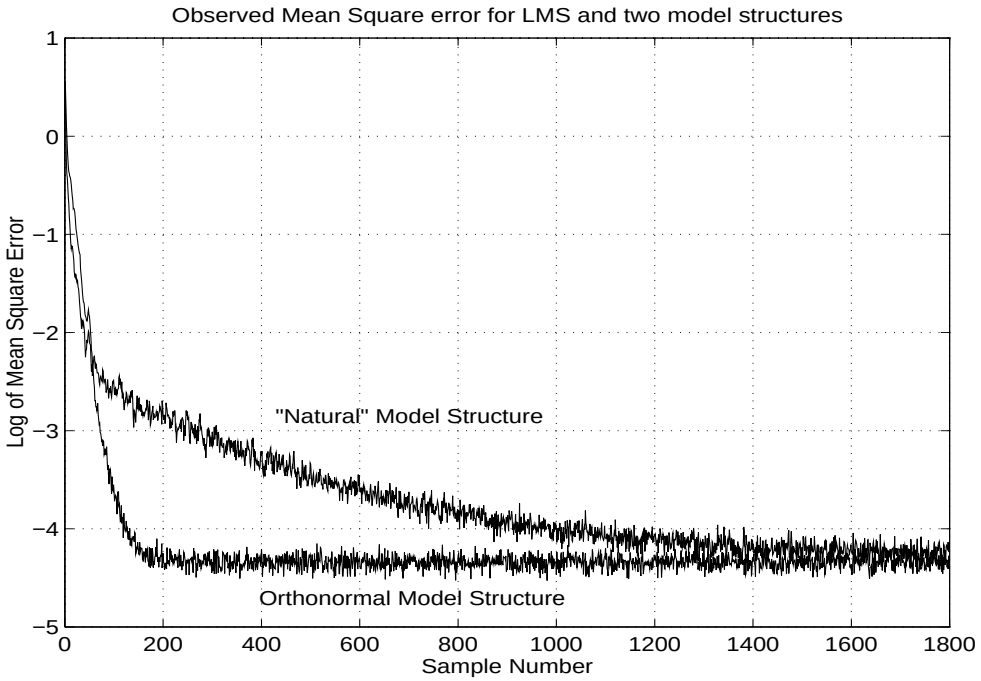

Figure 2: *Sample mean square error rates (averaged over 500 realisations) for LMS algorithm with (top plot) 'natural' fixed denominator model structure (19) and (bottom plot) orthonormal model structure (20).*

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_u(\omega)\left|\sum_{r=0}^{p-1}x_r\mathcal{B}_r(e^{j\omega})\right|^2 \mathrm{d}\omega, \tag{41}$$

$$\leq \max_{\omega\in[-\pi,\pi]}\frac{\Phi_u(\omega)}{2\pi}\sum_{r=0}^{p-1}\sum_{k=0}^{p-1}x_r x_k\int_{-\pi}^{\pi}\mathcal{B}_r(e^{j\omega})\overline{\mathcal{B}_k(e^{j\omega})}\,\mathrm{d}\omega \tag{42}$$

$$= \max_{\omega\in[-\pi,\pi]}\Phi_u(\omega). \tag{43}$$

But since R is symmetric then

$$\max_{x^Tx=1} x^TRx = \lambda_{\max}(R).$$

Using a similar argument to under-bound the eigenvalue then completes the lemma. □

This implies an upper bound on the on–average convergence rate of

$$\left[1-\frac{\min_{\omega\in[-\pi,\pi]}\Phi_u(\omega)}{\max_{\omega\in[-\pi,\pi]}\Phi_u(\omega)}\right]^k. \tag{44}$$

An interesting question is how closely this worst case bound is reached, which is equivalent to asking how conservative the bounds are in the previous lemma. In keeping with the

spirit of this paper of providing simplifications by considering high model orders, this can be answered by considering the infinite dimensional matrix $R_\infty \triangleq \lim_{p\to\infty} R$. In this case, the formulation (22) shows that R_∞ is an operator $\ell_2 \to \ell_2$, so that the eigenvalues of the finite dimensional matrix R, generalise to the continuous spectrum $\lambda(R_\infty)$ of the operator R_∞ defined as [1]

$$\lambda(R_\infty) = \{\lambda \in \mathbf{R} : \lambda I - R_\infty \text{ is not invertible}\}.$$

This spectrum can be characterised as follows.

Lemma 7.2.

$$\lambda(R_\infty) = \text{Range}\{\Phi_u(\omega)\}.$$

Proof. Take $x \in \ell_2$ arbitrary. Then using the formulation (22), and the knowledge [16] that $\text{Span}\{\mathcal{B}_n\}$ is dense in $H_2([-\pi,\pi])$ gives that R_∞ is not just an operator $\ell_2 \to \ell_2$, but can also be considered as the following Toeplitz operator $H_2 \to H_2$ with symbol $\Phi_u(\omega)$:

$$\begin{array}{rcl} R_\infty : H_2([-\pi,\pi]) & \to & H_2([-\pi,\pi]) \\ \displaystyle\sum_{n=0}^{\infty} x_n \mathcal{B}_n(e^{j\pi}) & \mapsto & \displaystyle\sum_{n=0}^{\infty} \langle f, \mathcal{B}_n\rangle\, \mathcal{B}_n(e^{j\omega}), \quad f(\omega) \triangleq \Phi_u(\omega) \sum_{n=0}^{\infty} x_n \mathcal{B}_n(e^{j\pi}). \end{array}$$

However, it is well known that the spectrum of the Toeplitz operator from $H_2 \to H_2$ is equal to the range of its symbol [1]. □

This indicates that for high model orders, the worst case bound on the spectrum of R given in Lemma 7.1 is likely to be achieved, and hence the worst case bound (44) on convergence rate is also likely to be achieved.

8 Simulation Examples

In this section, the utility of the previous theoretical analysis will be demonstrated via several simulation studies. In all cases, it is supposed that there is an underlying continuous time system with transfer function

$$G(s) = \frac{1}{(s+1)(10s+1)}$$

from which input–output data is collected by sampling every one second. To begin with, the case of stationary systems will be treated, but later time variations away from $G(s)$ will be considered. The input $\{u_k\}$ is stationary and Gaussian with spectral density

$$\Phi_u(\omega) = \frac{10}{1.25 - \cos\omega},$$

and the observed output $\{y_k\}$ is subject to white Gaussian corruption $\{\nu_k\}$ of variance $\sigma_\nu^2 = 0.01$. Based on this observed data, an attempt is made to estimate the discrete time system

$$G(q) = \text{ZOH}\left\{\frac{1}{(s+1)(10s+1)}\right\} = \frac{0.0355q + 0.0247}{(q-0.9048)(q-0.3679)} \tag{45}$$

via the model structure (20) with poles $\{\xi_n\}$ chosen to correspond to continuous time guesses of 0.2 and 0.25 radians per second. Note that these poles, being far from either of the true poles at 0.1 and 1 rad/s, are particularly bad guesses. They have been chosen to dispel any suspicion in the sequel that the high accuracy of the approximations (33),(34) and (38) illustrated in figures 4–8 derive from unreasonable prior knowledge or idealised conditions.

All three algorithms, the LMS with $\mu = 0.001$, RLS with $\lambda = 0.999$ and $P_0 = I$, and the Kalman Filter with $\mu = 0.001, \Sigma = 0.1, P_0 = I$ and $\sigma^2 = 0.01$ were employed with a tenth order model structure ($p = 10$). The parameter space convergence results are shown in figures 3, the fast convergence illustrating that these examples do not represent a case of unreasonably slow adaptation. Again, this choice is made to illustrate the robustness of the theoretical analysis to the violation of certain assumptions (small μ) that it is performed under.

These estimation experiments were performed five hundred times with different realisations for the input and measurement noise. This allowed the true frequency domain estimation error $\mathbf{E}\{|\widetilde{G}_k(e^{j\omega})|^2\}$ to be estimated by calculating its sample value as an average over the 500 realisations. This is plotted as the solid line in figures 4–6. The dash–dot lines in these figures are the approximations (33), (34) and (38) derived from theorems 5.1–5.3.

To be more specific, in the left hand diagram of figure 4, the LMS approximation (33) is profiled against the true average error distribution with respect to frequency, and appears to be highly accurate. This is in spite of the approximation (33) being derived from the asymptotic in p result in theorem 5.1, but being applied in this simulation to only a $p = 10$'th order model.

Similarly, in the right hand diagram of figure 4, the RLS approximation (34) is profiled against the true error distribution, and is again quite accurate. Finally, in the left hand diagram of figure 5, the same validity of (38) as an approximant for the error distribution of the Kalman Filter algorithm is demonstrated.

To illustrate just how robust the approximations can be to the use of a low model order, the case for the RLS algorithm and a model order of only $p = 4$ and a data length of only 300 samples is shown in the right hand diagram of figure 5. The approximation (dash–dot) line still appears to be a highly informative indication of the true variability (solid line).

In figure 6, all these tenth order monte–carlo simulation results are compared to their theoretical approximants (33), (34) and (38) in one diagram. The top plots are for the LMS algorithm, the middle ones are for the Kalman filter and the bottom ones are for RLS. The fact that the RLS variability is approximately an order of magnitude lower than that for LMS is supported theoretically by examining the approximants (33) and (34). Specifically, (33) indicates that for $\rho = 0$, the LMS steady state error is unaffected by input spectral

density $\Phi_u(\omega)$, while (34) illustrates that for RLS, this same steady state error is inversely proportional to $\Phi_u(\omega)$. Therefore, for stationary systems, the ratio of LMS error to RLS error should be equal to $\Phi_u(\omega)$, which observed in 6.

As well, comparing the theoretical expressions (38) and (34) indicates that the Kalman Filter variability should be larger than the RLS variability by a factor $(1+\rho^2/\mu^2)\sqrt{\Phi_u/\sigma_\nu^2}$, and this is also supported by the larger observed variability for the Kalman filter in figure 6. Notice also the slower roll–off at high frequencies of the Kalman filter error as compared to the RLS or LMS error. This is due to the presence of the $\sqrt{\cdot}$ term in the Kalman filter error expression (38) which is not present in the LMS or RLS expressions (33) or (34).

Turning away from steady–state (large k) performance, the LMS (with μ decreased to $\mu = 0.0003$ and $\Phi_u(\omega) = (1.57 - 1.75\cos\omega)^{-1}$) transient analysis approximation in equation (27) was also tested via simulation with the results shown in figure 8. In this figure, the variance $\mathbf{E}\{|\widetilde{G}_k(e^{j\omega})|^2\}$ was estimated as the sample variance calculated over 500 simulations with different input and measurement noise realisations. This estimate is shown as the solid lines. The top plot with the faster decay rate is the error at $\omega = 0.1$ rad/s, and the bottom plot with the slower decay rate is the error at $\omega = 0.4$ rad/s. The dash–dot lines in these two diagrams are the estimates for this transient response derived from (27). Again, the agreement between the observed error and the theoretical prediction is quite close in spite of the fact that the approximation (27) is derived from an asymptotic in p result and then applied to a small ($p = 10$) model order. In particular, note that (27) is able to clearly explain and predict the slower convergence at the higher frequency where the input spectral density is smaller.

Finally, the validity of the approximations (33), (34) and (38) for the case of non–stationary plants ($\rho \neq 0$) was tested by starting with the system (45) and then perturbing it according to the random walk model (26) with Q chosen so as to imply a $\rho^2\delta_p(\omega)$ shown in the right hand diagram of figure 7. In the left hand diagram of that figure is shown (solid line) the sample mean square variability of the RLS estimate in steady state ($k = 800$) versus that predicted (dash–dot line) via the theoretically derived approximation (34). As in previous simulations, a tenth order model was used, and also as in previous simulations the agreement between sample observation and theoretical approximation is good.

If fact, counter–intuitively the agreement between observation and theory appears better in the time varying plant case than in the previous time invariant case. This can be explained by noting that in the time invariant case, the poles $\{\xi_k\}$ in the model were deliberately chosen to be far from the true plant poles in order to test the robustness of the approximations (33), (34) and (38). In the time varying case under the model (45) this is not possible, so that there is no under–modeling component in the results shown in figure 7.

9 Conclusions

This paper has provided an analysis of the frequency domain tracking and steady–state error for various adaptive estimation schemes. A key contribution of the paper was to

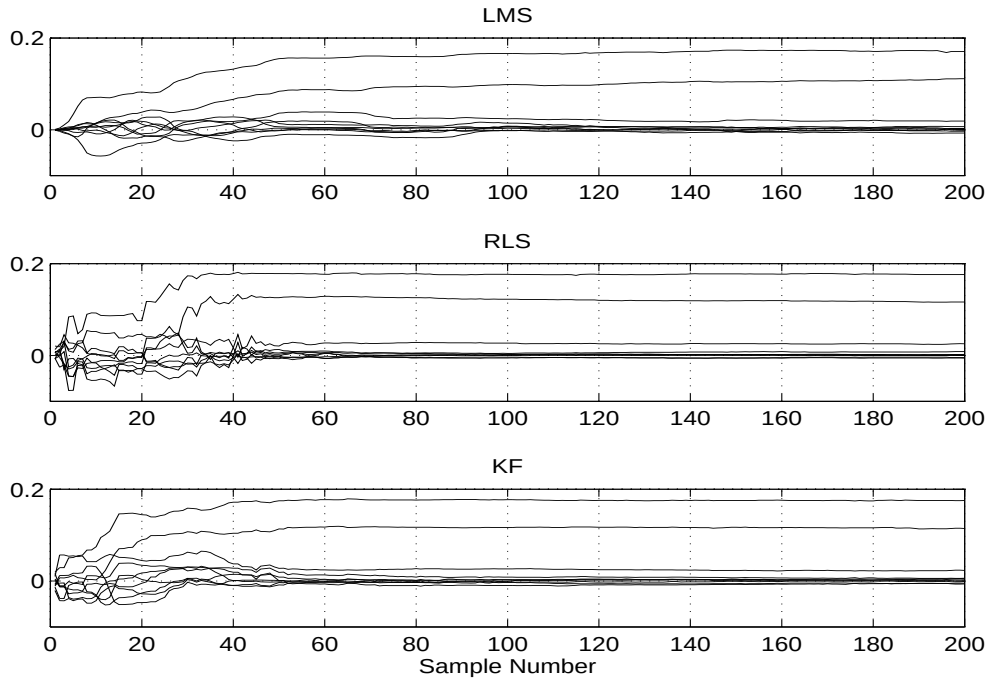

Figure 3: *Parameter space convergence for LMS, RLS and Kalman Filter.*

extend certain results already known for FIR model structures wherein all poles are fixed at the origin, to more general model structures where the poles may be placed arbitrarily, so long as they are stable. It was shown that the entire effect of the choice of fixed pole position on the frequency domain steady–state error depends on one term (called $\gamma_p(\omega)$). For FIR models, this term is simply p, the model order, but for fixed pole choices not at the origin (FIR), the term becomes frequency dependent in a manner that is influenced solely by the choice of poles.

The key tool in providing these results was to re–parameterise with respect to a particular orthonormal basis that generalises the classical trigonometric basis. For the case of RLS and Kalman Filtering algorithms, the orthonormal basis was shown to be an intrinsic part of the estimation problem, since whether or not the model structure used is cast in terms of this orthonormal basis, the frequency domain error is characterised by the term $\gamma_p(\omega)$ which depends wholly on the orthonormal parameterisation.

The validity of using results derived for infinite model order in a finite and small model order setting was examined via simulation, where for tenth (and even fourth) order models, the results which are exact for infinite model order where shown to provide very informative approximations for finite model order.

Appendices

A Generalised Fourier Series Convergence

Throughout these appendices a $p \times p$ real symmetric matrix $M_p(f)$ is defined by a real valued function $f(\omega)$ and a vector $\Gamma_p(\omega) \triangleq \Gamma_p(e^{j\omega})$ as defined in (17) by

$$M_p(f) \triangleq \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma_p(\sigma)\Gamma_p^\star(\sigma) f(\sigma)\, \mathrm{d}\sigma. \tag{A.1}$$

Such matrices will be termed 'generalised Toeplitz matrices' since when all the poles $\{\xi_k\}$ are chosen at the origin this formulation is a well known parameterisation of a Toeplitz matrix [8]. As well, in this special case of all the poles chosen at the origin, then the quadratic form

$$\frac{1}{\gamma_p(\omega)}\Gamma_p^\star(\omega) M_p(f) \Gamma_p(\omega) = \frac{1}{\gamma_p(\omega)} \int_{-\pi}^{\pi} |K_p(\omega,\sigma)|^2 f(\sigma)\, \mathrm{d}\sigma$$

where

$$K_p(\omega,\sigma) \triangleq \sum_{k=0}^{p-1} \overline{\mathcal{B}_k(\sigma)} \mathcal{B}_k(\omega), \quad \gamma_p(\omega) \triangleq K_p(\omega,\omega)$$

specialises to the Cèsaro mean

$$\sum_{k=1-p}^{p-1} \left(1 - \frac{|k|}{p}\right) c_k e^{j\omega k} = \frac{1}{p} \int_{-\pi}^{\pi} |K_p(\omega,\sigma)|^2 f(\sigma)\, \mathrm{d}\sigma \tag{A.2}$$

of the Fourier co-efficients $\{c_k\}$ of $f(\sigma)$ defined by

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\sigma) e^{-j\sigma k}\, \mathrm{d}\sigma$$

where

$$|K_p(\omega,\sigma)|^2 = \frac{\sin^2 (p+1)(\omega-\sigma)/2}{(p+1)\sin^2(\omega-\sigma)/2}$$

is known as the Fejér Kernel. These Cèsaro mean reconstructions (A.2) are commonly employed since they are guaranteed to converge to $f(\sigma)$ if the latter is continuous; an ordinary Fourier series construction is not necessarily convergent in this case [5].

In the following, the key technical result upon which the analysis of this paper rests is to generalise this Cèsaro mean convergence to generalised Fourier series $f(\sigma)$ with respect to the orthonormal basis $\{\mathcal{B}_k(e^{j\omega})\}$. The above mentioned convergence result for the classical Fourier basis $\mathcal{B}_k(e^{j\omega}) = e^{jk\omega}$ emerges as a special case by choosing all the poles $\{\xi_k\}$ at the origin.

Theorem A.1. *Suppose $f(\omega)$ is a real valued and continuous function on $[-\pi,\pi]$. Then provided*

$$\sum_{k=0}^{\infty} (1 - |\xi_k|) = \infty$$

the following limit result holds

$$\lim_{p\to\infty} \frac{\Gamma_p^\star(\omega) M_p(f) \Gamma_p(\omega)}{\gamma_p(\omega)} = f(\omega)$$

Proof.

$$\begin{aligned}
\Gamma_p(\omega)^\star M_p(f)\Gamma_p(\omega) &= \sum_{m=0}^{p-1}\sum_{n=0}^{p-1} \overline{\mathcal{B}_m(e^{j\omega})}\mathcal{B}_n(e^{j\omega})[M_p(f)]_{m,n} \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\sigma) \sum_{m=0}^{p-1}\sum_{n=0}^{p-1} \overline{\mathcal{B}_m(e^{j\omega})}\mathcal{B}_n(e^{j\omega})\mathcal{B}_m(e^{j\sigma})\overline{\mathcal{B}_n(e^{j\sigma})}\,\mathrm{d}\sigma \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\sigma)\,|K_p(\omega,\sigma)|^2\,\mathrm{d}\sigma.
\end{aligned}$$

Therefore, for any $\delta > 0$

$$\begin{aligned}
\frac{1}{2\pi}\left|\frac{\Gamma_p^\star(\omega) M_p(f)\Gamma_p(\omega)}{\gamma_p(\omega)} - f(\omega)\right| &= \frac{1}{2\pi\gamma_p(\omega)}\left|\Gamma_p^\star(\omega) M_p(f)\Gamma_p(\omega) - \gamma_p(\omega) f(\omega)\right| \\
&= \frac{1}{2\pi\gamma_p(\omega)}\left|\int_{-\pi}^{\pi} (f(\sigma)-f(\omega))\,|K_p(\omega,\sigma)|^2\,\mathrm{d}\sigma\right| \\
&\le \frac{1}{2\pi\gamma_p(\omega)}\left|\int_{\sigma\in[\omega-\delta,\omega+\delta]} (f(\sigma)-f(\omega))\,|K_p(\omega,\sigma)|^2\,\mathrm{d}\sigma\right| + \\
&\quad \frac{1}{2\pi\gamma_p(\omega)}\left|\int_{\sigma\notin[\omega-\delta,\omega+\delta]} (f(\sigma)-f(\omega))\,|K_p(\omega,\sigma)|^2\,\mathrm{d}\sigma\right|.
\end{aligned}$$

Now, $f(\omega)$ is continuous, so for δ sufficiently small

$$|f(\sigma) - f(\omega)| \le \epsilon \ \text{ on } [\omega-\delta, \omega+\delta].$$

and hence

$$\frac{1}{2\pi\gamma_p(\omega)}\left|\int_{\sigma\in[\omega-\delta,\omega+\delta]} (f(\sigma)-f(\omega))\,|K_p(\omega,\sigma)|^2\,\mathrm{d}\sigma\right| \le \frac{\epsilon}{2\pi\gamma_p(\omega)}\int_{-\pi}^{\pi} |K_p(\omega,\sigma)|^2\,\mathrm{d}\sigma = \epsilon.$$

Also, since f is continuous on compact $[-\pi,\pi]$ then f is bounded by some $M/2 < \infty$. Therefore

$$\frac{1}{2\pi\gamma_p(\omega)}\left|\int_{\sigma\notin[\omega-\delta,\omega+\delta]} (f(\sigma)-f(\omega))\,|K_p(\omega,\sigma)|^2\,\mathrm{d}\sigma\right| \le \frac{M}{2\pi\gamma_p(\omega)}\int_{\sigma\notin[\omega-\delta,\omega+\delta]} |K_p(\omega,\sigma)|^2\,\mathrm{d}\sigma.$$

This gives

$$\frac{1}{2\pi}\left|\frac{\Gamma_p^\star(\omega) M_p(f)\Gamma_p(\omega)}{\gamma_p(\omega)} - f(\omega)\right| \le \epsilon + \frac{M}{2\pi\gamma_p(\omega)}\int_{\sigma\notin[\omega-\delta,\omega+\delta]} |K_p(\omega,\sigma)|^2\,\mathrm{d}\sigma$$

However, by Lemma C.1

$$|K_p(\omega,\sigma)| \leq \frac{2}{|e^{j(\sigma-\omega)}-1|} = \frac{1}{|\sin(\sigma-\omega)/2|}$$

so that

$$\int_{\sigma\notin[\omega-\delta,\omega+\delta]} |K_p(\omega,\sigma)|^2 \,\mathrm{d}\sigma \leq \frac{2\pi}{\sin^2\delta/2}.$$

Also, since $|e^{j\omega}-\xi_k| \leq 1+|\xi_k|$

$$\gamma_p(\omega) = \sum_{k=0}^{p-1} \frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2} \geq \frac{1}{2}\sum_{k=0}^{p-1}(1-|\xi_k|) \tag{A.3}$$

so that

$$\frac{1}{\gamma_p(\omega)}\int_{\sigma\notin[\omega-\delta,\omega+\delta]} |K_p(\omega,\sigma)|^2 \,\mathrm{d}\sigma \leq \frac{\pi}{\sin^2\delta/2}\left(\sum_{k=0}^{p-1}(1-|\xi_k|)\right)^{-1}$$

which tends to zero. Using the fact that ϵ is arbitrary then gives the result. □

B Convergence of Quadratic Forms

Theorem B.1. *Suppose $f(\omega) \in L_2([-\pi,\pi])$ is positive definite and has finite dimensional spectral factorisation. Then provided*

$$\sum_{k=0}^{\infty}(1-|\xi_k|) = \infty$$

the following limit result holds

$$\lim_{p\to\infty} \frac{\Gamma_p(\mu)^\star M_p^{-1}(f)\Gamma_p(\omega)}{\gamma_p(\omega)} = f^{-1}(\omega)$$

Proof.

$$\frac{\Gamma_p(\mu)^\star M_p(f)^{-1}\Gamma_p(\omega)}{\gamma_p(\omega)} = \frac{\Gamma_p(\mu)^\star M_p(1/f)\Gamma_p(\omega)}{\gamma_p(\omega)} + \frac{\Gamma_p(\mu)^\star M_p(f)^{-1}[I-M_p(f)M_p(1/f)]\Gamma_p(\omega)}{\gamma_p(\omega)}$$

Now by construction, the elements of the vector $\Gamma(\omega)$ are bounded in magnitude by some finite number K. Similarly, by Lemma C.2 the elements of the vector $M_p(f)^{-1}\Gamma_p(\omega)$ can also be bounded by some finite number K. In this case, using Lemma C.3 gives that for some $|\eta|<1$

$$\left|\Gamma_p(\omega)^\star M_p(f)^{-1}[I-M_p(f)M_p(1/f)]\Gamma_p(\omega)\right| \leq \sum_{m=0}^{p-1}\sum_{n=0}^{p-1}\left|\left[\Gamma_p(\omega)^\star M_p(f)^{-1}\right]_m\right| \times$$

$$\begin{aligned} & \left| [M_p(f)M_p(1/f)]_{m,n} - [M_p(1)]_{m,n} \right| \left| [\Gamma_p(\omega)]_n \right|, \\ \leq \; & K^3 \sum_{m=0}^{p-1} \sum_{n=0}^{p-1} (\eta^{p-m} + \eta^m)(\eta^{p-n} + \eta^n), \\ = \; & K^3 \left(\frac{1-\eta^p}{1-\eta} \right)^2 (\eta^p + \eta)^2 < \infty. \end{aligned}$$

But by (A.3)

$$\frac{1}{2} \sum_{k=0}^{p-1} (1 - |\xi_k|) \leq \gamma_p(\omega)$$

so that under the conditions of the Lemma

$$\lim_{p\to\infty} \frac{\Gamma_p(\omega)^\star M_p(f)^{-1}[I - M_p(f)M_p(1/f)]\Gamma_p(\omega)}{\gamma_p(\omega)} = 0.$$

Therefore, by using Theorem A.1

$$\lim_{p\to\infty} \frac{\Gamma_p(\omega)^\star M_p(f)^{-1}\Gamma_p(\omega)}{\gamma_p(\omega)} = \lim_{p\to\infty} \frac{\Gamma_p(\omega)^\star M_p(1/f)\Gamma_p(\omega)}{\gamma_p(\omega)} = f^{-1}(\omega).$$

□

Lemma B.1. *Suppose $Q_p \in \mathbf{R}^{p\times p}$ is symmetric, positive definite and $\|Q_p\|_2 < \infty$ for all p. Suppose that $f(\omega)$ is real valued, continuous on $[-\pi, \pi]$ and has finite dimensional spectral factorisation. Suppose that*

$$\sum_{k=0}^{\infty} (1 - |\xi_k|) = \infty.$$

Then

$$\lim_{p\to\infty} \frac{1}{\gamma_p(\omega)} \Gamma_p^\star(\omega) M_p(f) Q_p M_p(f) \Gamma_p(\omega) = f^2(\omega) \lim_{p\to\infty} \frac{1}{\gamma_p(\omega)} \Gamma_p^\star(\omega) Q_p \Gamma_p(\omega).$$

Proof. For simplicity of notation, define the function $g(\sigma) \triangleq f(\sigma) - f(\omega)$. Then

$$\frac{1}{\gamma_p(\omega)} \left| \Gamma_p^\star(\omega) M_p(f) Q_p M_p(f) \Gamma_p(\omega) - f^2(\omega) \Gamma_p^\star(\omega) Q_p \Gamma_p(\omega) \right| \leq$$

$$\frac{1}{\gamma_p(\omega)} \left| \Gamma_p^\star(\omega) M_p(g) Q_p M_p(g) \Gamma_p(\omega) \right| + \frac{|f(\omega)|}{\gamma_p(\omega)} \left| \Gamma_p^\star(\omega) [Q_p M_p(g) + M_p(g) Q_p] \Gamma_p(\omega) \right|. \quad \text{(B.1)}$$

Now, considering the first term in this upper bound,

$$\begin{aligned} \frac{1}{\gamma_p(\omega)} \left| \Gamma_p^\star(\omega) M_p(g) Q_p M_p(g) \Gamma_p(\omega) \right| \; & \leq \; \frac{\|Q_p\|_2}{\gamma_p(\omega)} \left| \Gamma_p^\star(\omega) M_p^2(g) \Gamma_p(\omega) \right| \\ & \leq \; \frac{\|Q_p\|_2}{\gamma_p(\omega)} \left| \Gamma_p^\star(\omega) M_p(g^2) \Gamma_p(\omega) \right| + \\ & \quad \frac{\|Q_p\|_2}{\gamma_p(\omega)} \left| \Gamma_p^\star(\omega) [M_p^2(g) - M_p(g^2)] \Gamma_p(\omega) \right|. \end{aligned}$$

Now, by theorem A.1

$$\lim_{p\to\infty}\frac{1}{\gamma_p(\omega)}\Gamma_p^\star(\omega)M_p(g^2)\Gamma_p(\omega)=g^2(\omega)=0.$$

and since by construction the elements of $\Gamma_p(\omega)$ are bounded in magnitude by some finite number K, then by lemma C.2 for some $|\eta|<1$

$$\begin{aligned}\frac{\|Q_p\|_2}{\gamma_p(\omega)}\left|\Gamma_p^\star(\omega)[M_p^2(g)-M_p(g^2)]\Gamma_p(\omega)\right| &\le \sum_{m=0}^{p-1}\sum_{n=0}^{p-1}\left|[\Gamma_p(\omega)^\star]_m\right|\times\\ &\quad\left|\left[M_p^2(g)\right]_{m,n}-[M_p(g^2)]_{m,n}\right|\left|[\Gamma_p(\omega)]_n\right|\\ &\le K^3\sum_{m=0}^{p-1}\sum_{n=0}^{p-1}(\eta^{p-m}+\eta^m)(\eta^{p-n}+\eta^n)\\ &= K^3\left(\frac{1-\eta^p}{1-\eta}\right)^2(\eta^p+\eta)^2<\infty.\end{aligned}$$

However,

$$\frac{1}{2}\sum_{k=0}^{p-1}(1-|\xi_k|)\le\gamma_p(\omega)$$

so that under the conditions of the theorem

$$\lim_{p\to\infty}\frac{\|Q_p\|_2}{\gamma_p(\omega)}\left|\Gamma_p^\star(\omega)[M_p^2(g)-M_p(g^2)]\Gamma_p(\omega)\right|=0.$$

Using Lemma B.2 to deal with the remaining term in the upper bound (B.1) then completes the proof. □

Lemma B.2. *Suppose $Q_p\in\mathbf{R}^{p\times p}$ is symmetric, positive definite and $\|Q_p\|_2<\infty$ for all p. Suppose that $f(\omega)$ is real valued and continuous on $[-\pi,\pi]$. Suppose that*

$$\sum_{k=0}^{\infty}(1-|\xi_k|)=\infty.$$

Then

$$\lim_{p\to\infty}\frac{1}{\gamma_p(\omega)}\Gamma_p^\star(\omega)M_p(f)Q_p\Gamma_p(\omega)=\lim_{p\to\infty}\frac{1}{\gamma_p(\omega)}\Gamma_p^\star(\omega)Q_pM_p(f)\Gamma_p(\omega)=f(\omega)\lim_{p\to\infty}\frac{1}{\gamma_p(\omega)}\Gamma_p^\star(\omega)Q_p\Gamma_p(\omega).$$

Proof. For simplicity of notation, define the function $g(\sigma)\triangleq f(\sigma)-f(\omega)$. Then

$$\frac{1}{\gamma_p(\omega)}\left|\Gamma_p^\star(\omega)M_p(f)Q_p\Gamma_p(\omega)-f(\omega)\Gamma_p^\star(\omega)Q_p\Gamma_p(\omega)\right|=\frac{1}{\gamma_p(\omega)}\left|\Gamma_p^\star(\omega)M_p(g)Q_p\Gamma_p(\omega)\right|.$$

Now define the functions $g^+(\sigma) \triangleq \max[g(\sigma), 0]$ and $g^-(\sigma) \triangleq \min[g(\sigma), 0]$ so that $g(\sigma) = g^+(\sigma) + g^-(\sigma)$ and hence

$$\begin{aligned}\frac{1}{\gamma_p(\omega)}\left|\Gamma_p^\star(\omega)Q_pM_p(g)\Gamma_p(\omega)\right| &\leq \frac{1}{\gamma_p(\omega)}\left|\Gamma_p^\star(\omega)Q_pM_p(g^+)\Gamma_p(\omega)\right| + \\ &\quad \frac{1}{\gamma_p(\omega)}\left|\Gamma_p^\star(\omega)Q_pM_p(g^-)\Gamma_p(\omega)\right|. \qquad \text{(B.2)}\end{aligned}$$

Considering only the first term in this upper bound, note that provided $f(\omega)$ is not equal to a constant (in which case the Lemma is trivial since $M_p(1) = I$), then for $x \in \mathbf{R}^p$ arbitrary

$$x^T M_p(g^+)x = \frac{1}{2\pi}\int_{-\pi}^{\pi} |x^T\Gamma_p(\sigma)|^2 g^+(\sigma)\,\mathrm{d}\sigma > 0$$

so that $M_p(g^+)$ is positive definite and hence

$$\begin{aligned}\frac{1}{\gamma_p(\omega)}\left|\Gamma_p^\star(\omega)Q_pM_p(g^+)\Gamma_p(\omega)\right| &= \frac{1}{\gamma_p(\omega)}\left|\Gamma_p^\star(\omega)M_p^{1/2}(g^+)[M_p^{-1/2}(g^+)Q_pM_p^{1/2}(g^+)]M_p^{1/2}(g^+)\Gamma_p(\omega)\right| \\ &\leq \frac{\|Q_p\|_2}{\gamma_p(\omega)}\left|\Gamma_p^\star(\omega)M_p(g^+)\Gamma_p(\omega)\right|.\end{aligned}$$

Finally, using Theorem A.1

$$\lim_{p\to\infty}\frac{1}{\gamma_p(\omega)}\Gamma_p^\star(\omega)M_p(g^+)\Gamma_p(\omega) = g^+(\omega) = 0.$$

Using the same argument for the remaining term in the upper bound (B.2) then completes the proof. □

C Technical Lemmata

Lemma C.1. *Define the Blaschke product*

$$\varphi_p(z) \triangleq \prod_{k=0}^{p-1}\frac{1-\overline{\xi_k}z}{z-\xi_k}.$$

Then

$$K_p(\omega,\sigma) \triangleq \sum_{n=0}^{p-1}\overline{\mathcal{B}_n(\sigma)}\mathcal{B}_n(e^{j\omega}) = \frac{1-\overline{\varphi_p(e^{j\sigma})}\varphi_p(e^{j\omega})}{e^{j(\sigma-\omega)}-1}. \qquad \text{(C.3)}$$

Proof. See [17]. □

Lemma C.2. *If $f, g \in L_2([-\pi,\pi])$ have finite dimensional spectral factorisations then there exists $|\eta| < 1$ and $K < \infty$ such that*

$$\left|[M_p(f)M_p(g)]_{m,n} - [M_p(fg)]_{m,n}\right| \leq K(\eta^{p-m}+\eta^m)(\eta^{p-n}+\eta^n).$$

Proof. See [17]. □

Lemma C.3. *Suppose $f \in L_2([-\pi,\pi])$ has a finite dimensional spectral factorisation. Then $\exists K < \infty$ which is independent of p such that*

$$\left|[M_p(f)\Gamma_p(\omega)]_\ell\right| < K, \quad \left|\left[M_p^{-1}(f)\Gamma_p(\omega)\right]_\ell\right| < K.$$

Proof. See [17]. □

Lemma C.4. *Suppose $f, g \in L_2([-\pi,\pi])$ have finite dimensional spectral factorisations. Then provided $\sum_{k=0}^{\infty}(1-|\xi_k|) = \infty$, using the $\sim$ notation defined in the proof of theorem 5.3*

$$M_p(f)M_p(g)M_p(f) \sim M_p(f^2g) \quad \textit{as } p \to \infty.$$

Proof. Using Lemmas C.2 and C.3

$$\Gamma_p^\star(\omega)\left[M_p(f)M_p(g)M_p(f) - M_p(f^2g)\right]\Gamma_p(\omega) =$$

$$\Gamma_p^\star(\omega)M_p(f)\left[M_p(g)M_p(f) - M_p(fg)\right]\Gamma_p(\omega) + \Gamma_p^\star(\omega)\left[M_p(f)M_p(fg) - M_p(f^2g)\right]\Gamma_p(\omega)$$

$$\begin{aligned}
&\leq \sum_{m=0}^{p-1}\sum_{n=0}^{p-1}\left|[\Gamma_p(\omega)^\star M_p(f)]_m\right|\left|[M_p(g)M_p(f)]_{m,n} - [M_p(fg)]_{m,n}\right|\left|[\Gamma_p(\omega)]_n\right| + \\
&\quad \sum_{m=0}^{p-1}\sum_{n=0}^{p-1}\left|[\Gamma_p(\omega)^\star]_m\right|\left|[M_p(f)M_p(fg)]_{m,n} - [M_p(f^2g)]_{m,n}\right|\left|[\Gamma_p(\omega)]_n\right| \\
&\leq K^3\sum_{m=0}^{p-1}\sum_{n=0}^{p-1}(\eta^{p-m}+\eta^m)(\eta^{p-n}+\eta^n), \\
&= K^3\left(\frac{1-\eta^p}{1-\eta}\right)^2(\eta^p+\eta)^2 < \infty.
\end{aligned}$$

Noting that $2\gamma_p(\omega) \geq \sum_{k=0}^{p-1}(1-|\xi|)$ then completes the proof. □

References

[1] A. Böttcher and B. Silbermann, *Invertibility and Asymptotics of Toeplitz Matrices*, Akadamie–Verlag, Berlin, 1983.

[2] C. Richard Johnson Jr., *On the interaction of adaptive filtering, identification and control*, IEEE Signal Processing Magazine, 12 (1995).

[3] ——, *Adaptive Filtering in Communication Systems*, School of Electrical Engineering, Cornell University, Ithaca, NY, Version 1.1 1993.

[4] G. Davidson and D. Falconer, *Reduced complexity echo cancellation using orthonormal functions*, IEEE Transactions on Circuits and Systems, 38 (1991), pp. 20–28.

[5] R. Edwards, *Fourier Series, a Modern Introduction.*, vol. 2 of Graduate texts in mathematics number 64., Springer Verlag, 1979.

[6] B. Egardt, C. Richard Johnson Jr., L. Ljung, and G. A. Williamson, *Adaptive system performance in the frequency domain*, in Proceedings of the 4th IFAC Symposium on Adaptive Systems, Control and Signal Processing, Grenoble, France, 1992, pp. 33–39.

[7] G. Goodwin and K. Sin, *Adaptive Filtering Prediction and Control*, Prentice-Hall, Inc., New Jersey, 1984.

[8] U. Grenander and G. Szegö, *Toeplitz Forms and their Applications*, University of California Press, Berkeley, 1958.

[9] S. Gunnarsson, *Frequency Domain aspects of Modeling and Control in Adaptive Systems*, PhD thesis, Department of Electrical Engineering, Linköping University, S-581 83 Linköping, Sweden, 1988.

[10] S. Gunnarsson and L. Ljung, *Frequency domain tracking characteristics of adaptive algorithms*, IEEE Transactions on Acoustics, Speech, Signal Processing, 37 (1989), pp. 1072–1089.

[11] L. Guo and L. Ljung, *Exponential stability of general tracking algorithms*, IEEE Transactions on Automatic Control, TAC-40 (1995), pp. 1376–1387.

[12] ——, *Performance analysis of general tracking algorithms*, IEEE Transactions on Automatic Control, TAC-40 (1995), pp. 1388–1402.

[13] L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, Inc., New Jersey, 1987.

[14] L. Ljung and S. Gunnarsson, *Adaptation and tracking in system identification–a survey*, Automatica, 26 (1990), pp. 7–21.

[15] L.Ljung and Z.D.Yuan, *Asymptotic properties of black-box identification of transfer functions*, IEEE Transactions on Automatic Control, AC-30 (1985), pp. 514–530.

[16] B. Ninness and F. Gustafsson, *A unifying construction of orthonormal bases for system identification*, IEEE Transactions on Automatic Control, (To Appear, March 1997).

[17] B. Ninness, H. Hjalmarsson, and F. Gustafsson, *Bias and variance analysis of estimation using fixed denominator model structures*, Technical Report EE9617, Department of Electrical and Computer Engineering, University of Newcastle, Australia. Submitted to IEEE Transactions on Automatic Control., (1996).

[18] H. Perez and S. Tsujii, *A system identification algorithm using orthogonal functions*, IEEE Transactions on Signal Processing, 38 (1991), pp. 752–755.

[19] P. Regalia, *Adaptive IIR Filtering in Signal Processing and Control*, Marcel Dekker, 1995.

[20] Z. Reznic, C. Richard Johnson Jr., and Fernando López de Victoria, *Frequency domain interpretation of LMS convergence of a fractionally spaced equalizer*, IEEE Signal Processing Letters, 3 (1996), pp. 206–208.

[21] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms*, Prentice Hall, 1995.

[22] G. A. Williamson, *Globally convergent adaptive filters with infinite impulse responses*, in Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing, 1993, pp. Volume III, 543–546.

[23] ——, *Linear in the parameters identification for classes of systems*, in Proceedings of the 32nd IEEE Conference on Decision and Control, 1993, pp. 2607–2612.

[24] ——, *Tracking random walk systems with vector space adaptive filters*, IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing, 42 (1995), pp. 543–547.

[25] G. A. Williamson and C. Richard Johnson Jr., *Some effects of parameterization change in system identification*, in Proceedings of the American Control Conference, 1992, pp. 1268–1269.

[26] G. A. Williamson and S. Zimmermann, *Globally convergent adaptive IIR filters based on fixed pole locations*, IEEE Trans. Signal Processing, 44 (1996), pp. 1418–1427.

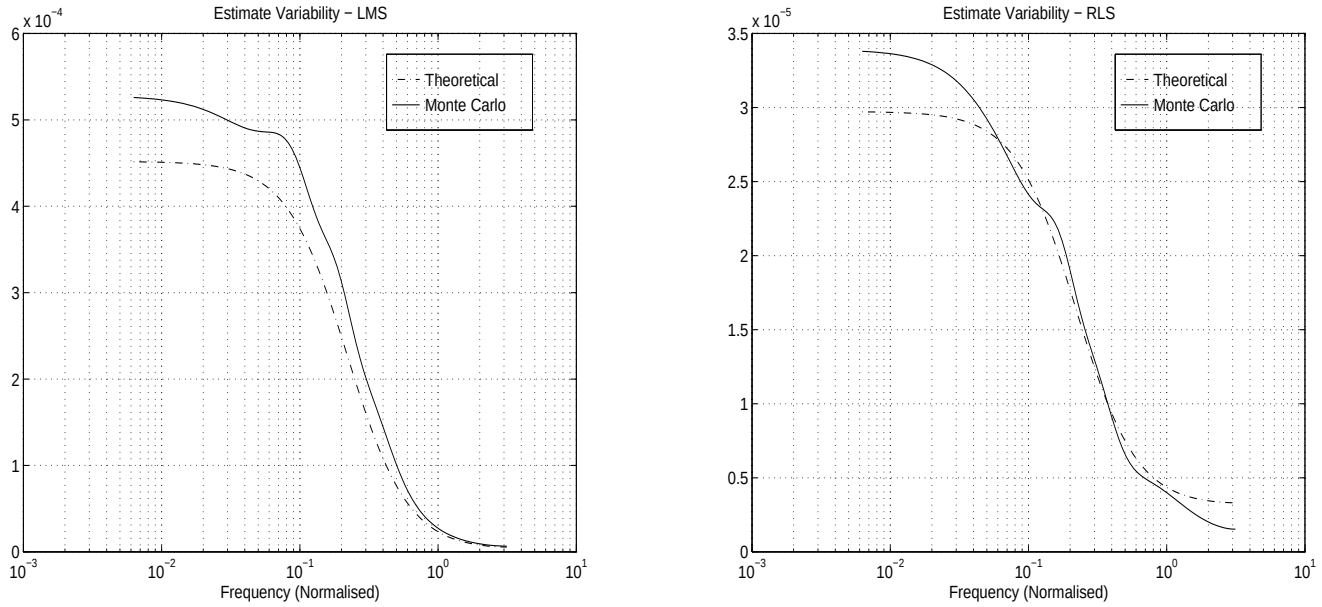

Figure 4: *Comparison of Sample Variability (over 500 experiments) vs theoretically derived approximations. On the left is the LMS algorithm vs the approximation (33). On the right is the RLS algorithm vs the approximation (34). In all cases a 10th order model and 800 data samples were used.*

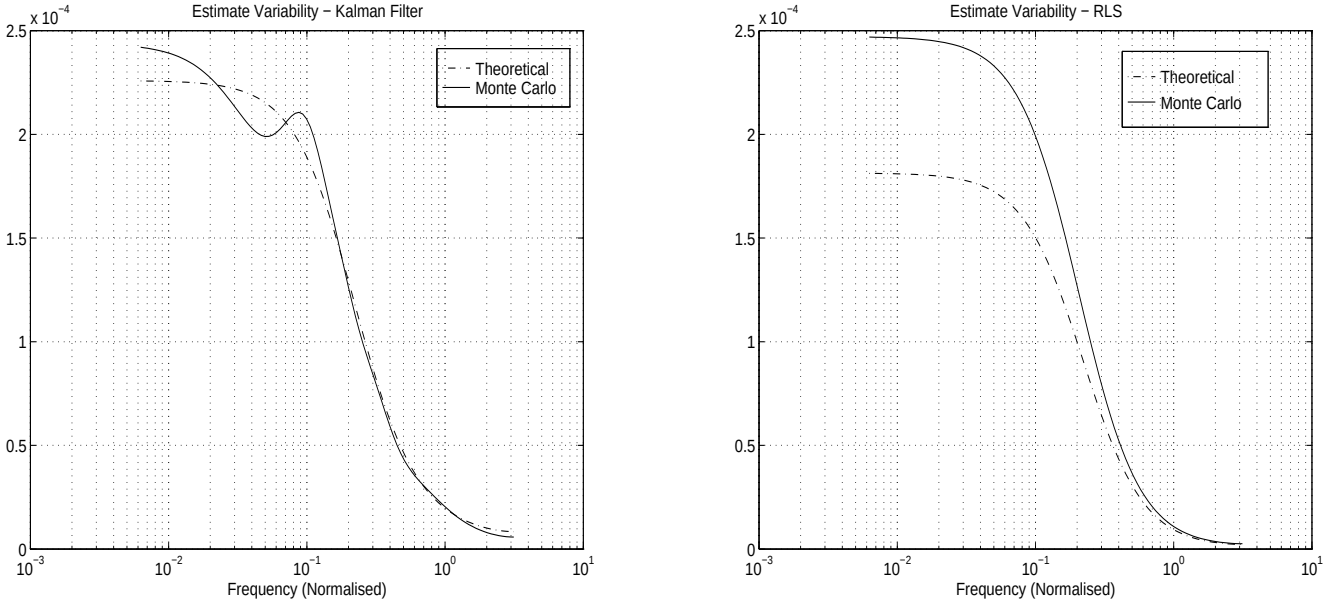

Figure 5: *Comparison of Sample Variability (over 500 experiments) vs theoretically derived approximations. On the left is the Kalman Filter algorithm vs the approximation (38). On the right is the RLS algorithm vs the approximation (34). For the Kalman Filter on the left, 10th order model and 800 data samples were used. For RLS on the right, only 300 data samples and only a 4th order model were used.*

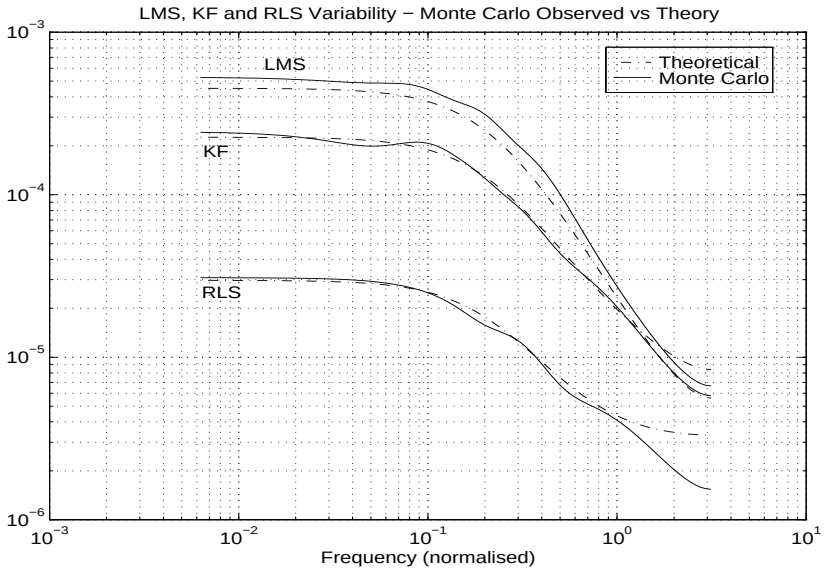

Figure 6: *Comparison of observed variability (solid line) vs theory (dash–dot line) for Kalman Filter (top), LMS (middle) and RLS (bottom).*

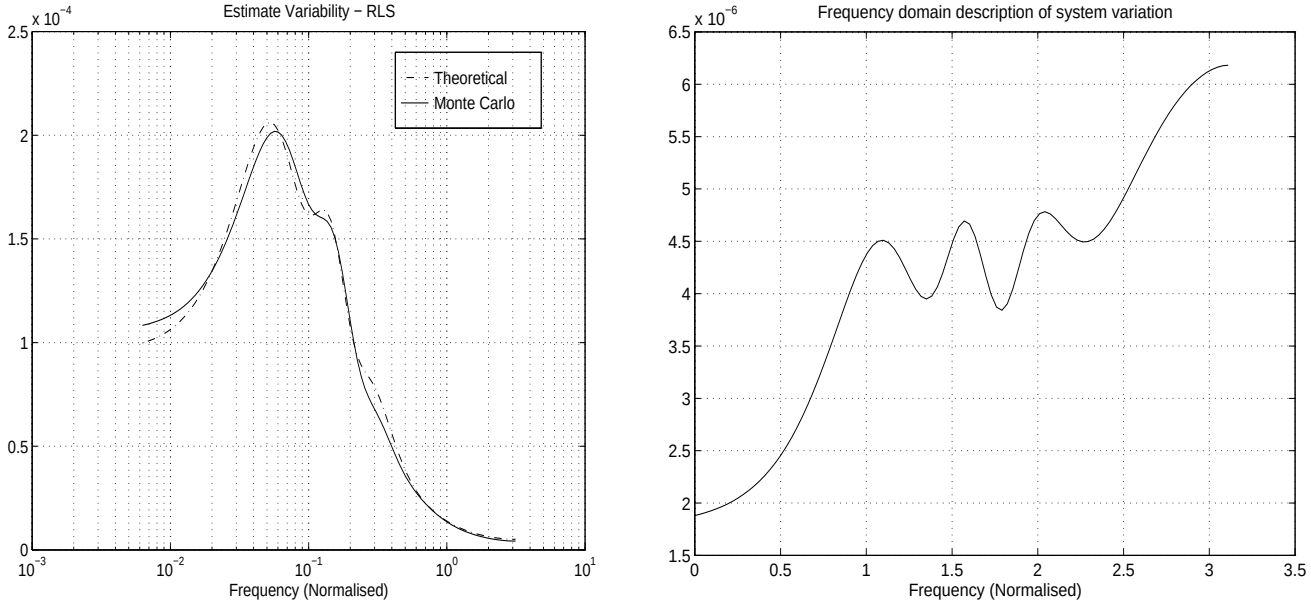

Figure 7: *Comparison of Sample Variability (over 500 experiments) vs theoretically derived approximation for RLS algorithm and non–stationary plant. On the left is the sample estimation error after $k = 800$ iterations (solid line) compared to the theoretical approximation (34) as a dash–dot line. On the right is the non-stationarity modeled in the frequency domain by $\rho^2 \delta_p(\omega)/\mu$.*

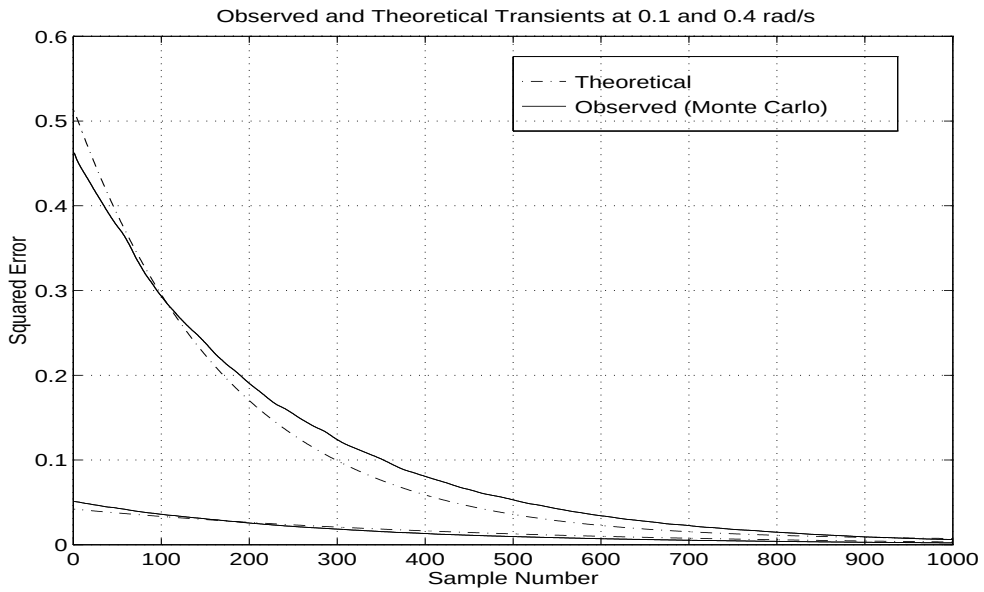

Figure 8: *Transient behavior of LMS at two different frequencies - observed (Monte Carlo Average) vs Theoretical prediction via approximation (24). Top plot is error* $\mathbf{E}\{|\widetilde{G}_k(e^{j\omega})|^2\}$ *at* $\omega = 0.1$ *rad/s. Bottom plot is error* $\mathbf{E}\{|\widetilde{G}_k(e^{j\omega})|^2\}$ *at* $\omega = 0.4$ *rad/s.*