QUANTIFYING THE ACCURACY OF ADAPTIVE TRACKING ALGORITHMS

Brett Ninness and Juan Carlos Gómez

Electrical and Computer Engineering
University of Newcastle,
Callaghan 2308, Australia
brett@ee.newcastle.edu.au

ABSTRACT

The use of adaptive algorithms such as Kalman Filtering, LMS and RLS together with FIR model structures is very common and extensively analysed. In the interests of improved performance an extension of the FIR structure has been proposed in which the fixed poles are not all at the origin, but instead are chosen by prior knowledge to be close to where the true poles are. Existing FIR analysis would indicate that the noise and tracking properties of such a scheme are invariant to the choice of fixed pole location. This paper establishes both numerically and theoretically that in fact this is not the case. Instead, the dependence of fixed pole location is made explicit by deriving frequency domain expressions that are obtained by using new results on generalised Fourier series and generalised Toeplitz matrices.

1. INTRODUCTION

In the context of estimating a vector of n FIR taps θ using an adaptive algorithm, a very well known result that has proved to be of great intuitive utility is that the variability of the FIR frequency response estimate $G(e^{j\omega}, \widehat{\theta})$ ($\widehat{\theta}$ is the estimate of θ) is approximately given by [3]

$$\mathrm{Var}\{G(e^{j\omega}, \widehat{\theta})\} \approx n\frac{\mu\sigma_\nu^2\kappa}{[\Phi_u(\omega)]^r} \qquad (1)$$

where κ is some constant, μ is the step size, σ_ν^2 is the white measurement noise variance, $\Phi_u(\omega)$ is the input excitation spectral density, and $r = 1$ for Recursive Least Squares, $r = 1/2$ for Kalman Filtering and $r = 0$ for the LMS algorithm.

Recently, workers [8] have suggested the use of 'extended FIR' structures wherein all the poles are not necessarily at the origin, but instead some (or all) are chosen to be as close as possible to where the true poles are believed to lie.

In assessing the validity of this scheme, a natural question arises as to how (1) should be modified to describe $\mathrm{Var}\{G(e^{j\omega}, \widehat{\theta})\}$ where now θ is a vector of 'generalised' FIR taps. The most obvious course is to conclude that these new model structures are really just the old FIR ones with an input $\{u_t\}$ pre-filtered by an all-pole filter $F(q) = 1/D_n(q)$ where $D_n(q) = \prod_{k=0}^{n-1}(q - \xi_k)$ with $\{\xi_k\}$ being the user chosen guesses as to the true pole locations. This would imply that the variability of the FIR 'numerator' part is then given by the expression (1) with the substitution

$$\Phi_u(\omega) \mapsto |F(e^{j\omega})|^2\Phi_u(e^{j\omega}) = \frac{\Phi_u(\omega)}{|D_n(e^{j\omega})|^2} \qquad (2)$$

made. The frequency domain variability of the whole model structure, being the FIR numerator part divided by the frequency response of the fixed denominator part, should then be (1) with the substitution (2) and then divided by $|D_n(e^{j\omega})|^2$. Clearly the $|D_n(e^{j\omega})|^2$ terms will cancel, and the conclusion will ensue that the variability of $\mathrm{Var}\{G(e^{j\omega}, \widehat{\theta})\}$ is invariant to the choice of fixed pole location.

This can be tested on a simple example wherein the true system is

$$G(q) = \frac{0.1548q + 0.0939}{(q - 0.6065)(q - 0.3679)}$$

and an $n = 10$'th order model is fitted using RLS when the input $\{u_t\}$ has spectral density $\Phi_u(\omega) = 10(1.25 - \cos\omega)^{-1}$, the output measurements $\{y_t\}$ are corrupted by white noise of variance $\sigma_\nu^2 = 0.01$, and the algorithm is run for $N = 2000$ data samples. In this case, the true variability can be estimated by the sample average over 200 Monte-Carlo simulations with different input and noise realisations. This can then be compared to the approximation (1).

For the case of all the $\{\xi_k\}$ being at the origin (so that a true FIR structure is employed), then the results of such a comparison are shown in the top plot of figure 1 - (1) being the dash-dot line, and the Monte-Carlo estimate of true variability being the solid line. The agreement is excellent.

However, if all the poles are chosen away from the origin, five at $\xi_k = 0.2$ and five at $\xi_k = 0.8$ then when examining the theoretical prediction (1) and the true variability as shown in the bottom plot of figure 1, the agreement between the two has disappeared

Nevertheless, for this latter case, note the good agreement between the true variability and the dashed line, which is a plot of an improved approximation presented in this paper. This new approximation is the old one (1) with the model order term n replaced with a frequency dependent function $\gamma_n(\omega)$. That is, the main results of the paper are firstly that the previous reasoning, which tried to employ the pre-existing result (1) is misleading in its conclusion that

This work was supported by the Australian Research Council and the Centre for Integrated Dynamics and Control.

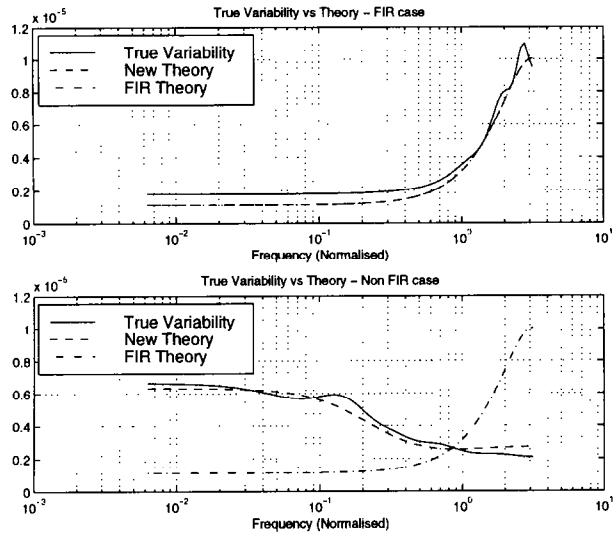

Figure 1: *Comparison of (solid line) true variability to (dashed line) FIR based theoretical approximation (1). Top plot is case of all poles at origin (FIR case), bottom plot is case of all poles away from origin. Dash-dot line is the improved approximation presented in this paper (Th 4.1).*

$\mathrm{Var}\{G(e^{j\omega},\widehat{\theta})\}$ is insensitive to fixed pole locations and secondly the approximation

$$\mathrm{Var}\{G(e^{j\omega},\widehat{\theta})\} \approx \gamma_n(\omega)\frac{\mu\sigma_\nu^2\kappa}{[\Phi_u(\omega)]^r} \tag{3}$$

acts as a generalisation of (1) which in the case of fixed poles not at the origin, can be much more accurate than (1). The function $\gamma_n(\omega)$ is purely a function of the choice of pole position $\{\xi_k\}$, and for the special choice of all of these at the origin (FIR), then $\gamma_n(\omega) = n$ as a special case. The development of this new approximation depends on the use of several new results which are generalisations of classical Fourier series and Toeplitz matrix properties to the case of the underlying orthonormal basis being generalised from the usual trigonometric one.

2. PROBLEM FORMULATION

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

$$y_t = G_t(q)u_t + \nu_t \tag{4}$$

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

$$G_t(q) = \sum_{n=1}^{\infty} g_t(n)q^{-n}$$

is a possibly time varying linear system with impulse response $\{g_t(n)\} \in \ell_2$. It is assumed that $\{u_t\}$ is a realisation of a stationary stochastic process with covariance function $R_u(\tau) = \mathbf{E}\{u_t u_{t-\tau}\}$ and associated spectral density $\Phi_u(\omega) = \sum_{\tau=-\infty}^{\infty} R_u(\tau)e^{-j\omega\tau}$ and that $\{u_t\}$ is weakly uncorrelated with $\{\nu_t\}$ in the sense that $|\mathbf{E}\{u_t\nu_{t-\tau}\}| \to 0$ 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 $G_t(q)$ by means of the observations $\{u_t\}$ and $\{y_t\}$. There are many approaches to this problem, but a common theme [2] is to express the dependence (4) in a linear regression form $y_t = \phi_t^T\theta_t + \nu_t$ where the 'regression vector' ϕ_t depends on measurements of $\{u_t\}$ and $\{y_t\}$ up until $t = k$ and $\theta_t \in \mathbf{R}^n$ is a vector of n parameters in a model structure $G(q,\theta_t)$ that attempts to describe the true dynamics $G_t(q)$. An estimate of $G_t(q)$ is then obtained as $G(q,\widehat{\theta}_t)$ where the estimate $\widehat{\theta}_t$ is obtained recursively via

$$\widehat{\theta}_{t+1} = \widehat{\theta}_t + L_t(y_t - \phi_t^T\widehat{\theta}_t), \quad \mu \in (0,1) \tag{5}$$

where L_t is a gain vector that may be computed in various ways. A common choice for this gain vector is $L_t = \mu\phi_t$, $\mu \in (0,1)$ in which case (5) is known as the 'gradient' or 'least mean square' (LMS) algorithm. Another common choice is $L_t = P_t\phi_t$ where P_t satisfies

$$P_t = \frac{1}{\lambda}\left\{P_{t-1} - \frac{P_{t-1}\phi_t\phi_t^T P_{t-1}}{\lambda + \phi_t^T P_{t-1}\phi_t}\right\}$$

with $\lambda = 1-\mu, \mu \in (0,1)$ and P_t is initialised with some positive definite P_0 and with the ensuing algorithm being known as 'Recursive Least Squares' (RLS). Finally, if the time variation of the parameters θ_t are modeled via a random walk as $\theta_{t+1} = \theta_t + \rho\, w_t$ where w_t is a stationary zero mean vector white noise process with $\mathbf{E}\{w_t w_t^T\} = Q$, then the update law

$$L_t = \frac{\mu P_{t-1}\phi_t}{\sigma^2 + \mu\phi_t^T P_{t-1}\phi_t} \tag{6}$$

where P_t satisfies the Riccati equation

$$P_t = P_{t-1} - \mu\frac{P_{t-1}\phi_t\phi_t^T P_{t-1}}{\sigma^2 + \mu\phi_t^T P_{t-1}\phi_t} + \mu\Sigma \tag{7}$$

with $\Sigma > 0$ and symmetric is known as the Kalman Filter.

When employing any of these adaptive schemes, a central question is the accuracy of the estimate $G(q,\widehat{\theta}_t)$ as a description of $G_t(q)$. The most common way of assessing this is to examine the accuracy of $\widehat{\theta}_t$ itself [2]. This may be achieved by defining θ_t as the true parameter vector that allows the model structure to exactly describe the underlying time varying dynamics as $G(q,\theta_t) = G_t(q)$ and by defining the estimation error $\widetilde{\theta}_t$ as $\widetilde{\theta}_t \triangleq \theta_t - \widehat{\theta}_t$.

The quality of an adaptive estimation scheme can then be quantified by using $\mathbf{E}\{\widetilde{\theta}_t\widetilde{\theta}_t^T\}$ as a measure of estimation accuracy. Unfortunately, as pointed out in [3, 4], the exact expression for this covariance will be very complicated except in very special circumstances. The main result of [4] which will be central to the analysis of this paper is that in spite of this, and under the stated assumptions and the definition $R \triangleq \mathbf{E}\{\phi_t\phi_t^T\}$ then $\mathbf{E}\{\widetilde{\theta}_t\widetilde{\theta}_t^T\}$ may be approximated by a time varying matrix Π_t which in the steady state case converges to a positive definite symmetric matrix Π given by

LMS: The solution of the Lyapunov equation

$$\Pi R + R\Pi = \mu\sigma_\nu^2 R + \frac{\rho^2}{\mu}Q. \tag{8}$$

RLS:

$$\Pi = \frac{\mu\sigma_\nu^2}{2}R^{-1} + \frac{\rho^2}{2\mu}Q. \tag{9}$$

Kalman Filter: For the special case of $\Sigma = Q$:

$$\Pi = \frac{\sigma^2}{2}\left(\mu\frac{\sigma_\nu^2}{\sigma^2} + \frac{\rho^2}{\mu}\right)S, \quad \sigma^2 SRS = \Sigma. \tag{10}$$

However, as argued in [3, 1], in many cases the interest is not in the accuracy in parameter space, but the accuracy in how close the estimated model $G(q,\widehat{\theta}_t)$ is to the true system $G_t(q)$ in terms of the error $\widetilde{G}_t(e^{j\omega}) \triangleq G_t(e^{j\omega}) - G(e^{j\omega},\widehat{\theta}_t)$ in the estimated frequency response. In this paper, model structures $G(q,\theta_t)$ are considered for which the estimated frequency response depends linearly on the estimated parameters as $G(e^{j\omega},\widehat{\theta}_t) = \Gamma_n^T(e^{j\omega})\widehat{\theta}_t$ so that $\phi_t = \Gamma_n(q)u_t$ where

$$\Gamma_n(q) \triangleq [\mathcal{B}_0(q), \mathcal{B}_1(q), \cdots, \mathcal{B}_{p-1}(q)]^T \tag{11}$$

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

In this case an approximate frequency domain quantification of adaptive performance may then be taken as

$$\begin{aligned}\mathbf{E}\left\{|\widetilde{G}_t(e^{j\omega})|^2\right\} &= \Gamma_n^*(e^{j\omega})\mathbf{E}\left\{\widetilde{\theta}_t\widetilde{\theta}_t^T\right\}\Gamma_n(e^{j\omega}),\\ &\approx \Gamma_n^*(e^{j\omega})\Pi_t\Gamma_n(e^{j\omega})\end{aligned} \tag{12}$$

where $\cdot^*$ 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 [3] and derive simple approximations for (12) such as (3).

3. MODEL STRUCTURES

The model structures examined in this paper have recently been proposed by Williamson and Zimmermann [8] where they have been termed 'fixed pole adaptive filters'. They are formulated as

$$G(q,\theta_t') = \left[\prod_{k=0}^{n-1}(q-\xi_k)\right]^{-1}\sum_{k=0}^{n-1}\theta_t'(k)q^k \tag{13}$$

A special case of this structure arises when all the poles $\{\xi_n\}$ are chosen at the origin in which case (13) is an FIR model structure.

However, empirical evidence [8] supports the fact that in an adaptive filtering context, a significant improvement in estimation accuracy is possible by avoiding poles $\{\xi_k\}$ 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 $G_t(q)$.

In spite of the pleasant properties enjoyed by the model structure (13), its generality (as compared to an FIR structure) makes frequency domain analysis of adaptive algorithms much more difficult. This has already been foreshadowed by the simulation examples of § 1, but to be more explicit on a theoretical level, the two key ideas of the seminal work of [3] on this topic were to notice that *in the FIR case*

1. Π_t is a Toeplitz matrix which can be formulated, for some spectral density f as $\Pi_t = T_n(f)$ defined as

$$[T_n(f)]_{m,\ell} = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j\omega(m-\ell)}f(\omega)\,\mathrm{d}\omega.$$

2. By recognising this Toeplitz matrix feature, then (12) specifying $\mathbf{E}\{|\widetilde{G}_t(e^{j\omega})|^2\}$ is actually an n'th order Fourier reconstruction of the function $f(\omega)$, and so should be approximately equal to $f(\omega)$.

The reason why these principles have failed in the simulation example shown in figure 1 of § 1 is that the underlying function being Fourier reconstructed depends on the effective input spectral density (for RLS, the function f above is $f = |D_n|^2/\Phi_u$), and if there are fixed poles in the model structure, this is an all-pole filtered quantity, which will be far less smooth (the degree depending on the number of fixed poles not at the origin) than the original spectral density Φ_u. Since the accuracy of the Fourier reconstruction used in step 2 depends crucially on the degree of smoothness, it eventuates that for a high proportion of the fixed poles not at the origin, the approximation (1) breaks down because the underlying Fourier reconstruction used in step 2 has not approximately converged.

The solution proposed in this paper and shown as the improved (dashed line) approximation in figure 1 is to absorb the fixed poles into the model structure, but *still in an orthonormally parameterised way.* In this case, certain new results on generalised Fourier convergence and the algebraic properties of generalised Toeplitz matrices are employed [6] to provide an approximation that is improved since it involves generalised Fourier reconstruction of a function that is invariant to the choice of fixed poles, and hence has fixed smoothness.

To be more specific, the strategy employed here involves replacing the model structure (13) with the following *orthonormal* formulation

$$G(q,\theta_t) = \sum_{n=0}^{p-1}\theta_t(n)\mathcal{B}_n(q) \tag{14}$$

where

$$\mathcal{B}_n(q) = \left(\frac{\sqrt{1-|\xi_n|^2}}{q-\xi_n}\right)\prod_{t=0}^{n-1}\left(\frac{1-\overline{\xi_t}q}{q-\xi_t}\right) \tag{15}$$

Since the poles of the model structure (14) and (13) are identical, then they are equivalent in the sense that for some nonsingular $J \in \mathbf{R}^{p\times p}$, the parameter vectors θ_t in (14) and $\theta_t' = J\theta_t$ in (13) describe exactly the same transfer function. As well, with initialisation $P_0 = J^{-1}P_0'J^{-T}$ consistent with this linear re–parameterisation, the RLS update equations are invariant to the re–parameterisation in the sense that $\widehat{\theta}_t' = J\widehat{\theta}_t$ so that frequency response estimates are identical: $G(e^{j\omega},\widehat{\theta}_t') = G(e^{j\omega},\widehat{\theta}_t)$. This same property also applies to the Kalman Filtering update law (6),(7) provided the compatibility $\Sigma = J^{-1}\Sigma'J^{-T}$ is also maintained.

The above 'basis functions' $\{\mathcal{B}_k(z)\}$ are orthonormal in the sense that

$$\langle\mathcal{B}_n,\mathcal{B}_m\rangle\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathcal{B}_n(e^{j\omega})\overline{\mathcal{B}_m(e^{j\omega})}\,\mathrm{d}\omega = \begin{cases}1 & ;m=n\\ 0 & ;m\neq n\end{cases}$$

In this case, the idea of Toeplitz matrices is generalised in [6] to one in which a matrix $M_n(f)$ is defined by a spectral density $f(\omega)$ as

$$[M_n(f)]_{m,t} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathcal{B}_m(e^{j\omega})\overline{\mathcal{B}_t(e^{j\omega})}f(\omega)\,\mathrm{d}\omega$$

is considered, in which case, taking the RLS case as an example, it is possible to show [5] that $\Pi_t \approx 1/2M_n(\mu\sigma_\nu^2/\Phi_u)$. Continuing by substituting $\Gamma_n(q) = [\mathcal{B}_0(q), \cdots, \mathcal{B}_{n-1}(q)]^T$ then provides that from (12)

$$\frac{1}{\gamma_n(\omega)}\mathbf{E}\{|\widetilde{G}_t(e^{j\omega})|^2\} \approx \frac{1}{2\gamma_n(\omega)}\Gamma_n^*(e^{j\omega})M_n\left(\frac{\mu\sigma_\nu^2}{\Phi_u}\right)\Gamma_n(e^{j\omega})$$

where

$$\gamma_n(\omega) \triangleq \sum_{k=0}^{p-1}|\mathcal{B}_k(e^{j\omega})|^2. \tag{16}$$

Finally, in [6] Fourier analysis is generalised in such a way to establish that

$$\lim_{n\to\infty}\frac{1}{2\gamma_n(\omega)}\Gamma_n^*(e^{j\omega})M_n\left(\frac{\mu\sigma_\nu^2}{\Phi_u}\right)\Gamma_n(e^{j\omega}) = \frac{\mu\sigma_\nu^2}{2\Phi_u(\omega)}$$

so that if one assumes that this convergence holds for finite n, the extended approximation (3) shown as the dashed line in figure 1 eventuates. Furthermore, the complete contribution of the fixed pole location choice is captured by the term $\gamma_n(\omega)$, which for several choices of $\{\xi_k\}$ is shown in figure 2.

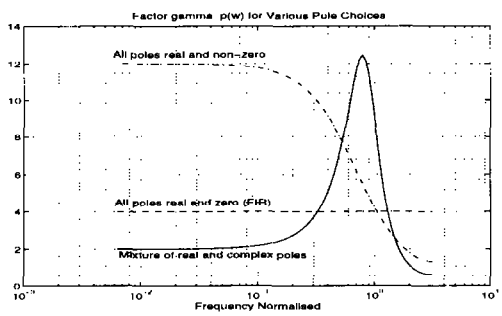

Figure 2: *Plot of $\gamma_n(\omega)$ for various choices of pole location.*

4. EXTENDED VARIANCE EXPRESSIONS

Having presented an informal overview for the RLS case of the methods used to provide the improved approximation of this paper, the results are now stated more comprehensively as (the quantity $\delta(\omega) \triangleq \mathbf{E}\{|G_{t+1}(e^{j\omega}) - G_t(e^{j\omega})|^2\}$ is used to quantify the time variation of $G_t(q)$ in the frequency domain-see [5] for details)

Theorem 4.1. *For the LMS algorithm and the model structure (14), then*

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

For the RLS algorithm and the model structure (13) or (14), then

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

For the Kalman Filtering algorithm, the model structure (13) or (14) and under the assumption that $Q = \Sigma$, then

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

Proof. Derived by using the parameters space approximations (8), (9) and (10) together with the generalised Fourier and Toeplitz results of [6]. See [5] for details. □

5. CONCLUSION

This paper has presented only an overview of results that are more fully developed in [5], where it is shown that Theorem 4.1 applies for any model structure that allows the regressors to be generated as

$$\phi_{t+1} = A\phi_t + Bu_t,$$

where $A \in \mathbf{R}^{n\times n}$ is any matrix with eigenvalues at locations $\{\xi_0, \cdots \xi_{n-1}\}$ and B is any n dimensional vector. As well, in [7] the results are further extended to cases where the poles are not fixed, such as ARX and ARMAX modeling. In this case it is shown that the presence of fixed noise model zeroes (which is equivalent to pre-filtering the data $\{u_t\}$ and $\{y_t\}$) implies that improved accuracy approximations to $\mathrm{Var}\{G(e^{j\omega}, \widehat{\theta})\}$ can also be generated along the lines of this paper by replacing n with $\gamma_n(\omega)$.

6. REFERENCES

[1] B. Egardt *et al.*, Adaptive system performance in the frequency domain. In *Proc. 4th IFAC Symp. Adap. Sys. Control and Sig. Proc.*, pages 33–39, 1992.

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

[3] S. Gunnarsson and L. Ljung. Frequency domain tracking characteristics of adaptive algorithms. *IEEE Trans. ASSP.*, 37(7):1072–1089, 1989.

[4] L. Guo and L. Ljung. Performance analysis of general tracking algorithms. *IEEE Trans. Auto. Control*, 40(8):1388–1402, 1995.

[5] B. Ninness and J.C. Gómez. Frequency domain analysis of tracking and noise properties of adaptive algorithms. *To appear, IEEE Trans. Sig. Proc.*, 1996. Available via `http://www.ee.newcastle.edu.au/users/staff/brett/`

[6] B. Ninness, H. Hjalmarsson, and F. Gustafsson. Generalised Fourier and Toeplitz results for rational orthonormal bases. *Tech. Rep. EE9740, Dept. of Elec. Eng. Uni. Newcastle, Australia. Submitted to SIAM J. Cont. and Optim.*, 1997. Available via `http://www.ee.newcastle.edu.au/users/staff/brett/`

[7] B. Ninness, H. Hjalmarsson, and F. Gustafsson. The Fundamental Role of Orthonormal Bases in System Identification. *Tech. Rep. EE9739, Dept. of Elec. Eng. Uni. Newcastle, Australia. Submitted to IEEE Tran. Auto. Contr.*, 1997. Available via `http://www.ee.newcastle.edu.au/users/staff/brett/`

[8] G.A. Williamson and S. Zimmermann. Globally convergent adaptive IIR filters based on fixed pole locations. *IEEE Trans. Signal Proc.*, 44(6):1418–1427, 1996.