Systems & Control Letters 8 (1987) 197–204
North-Holland

How exciting can a signal really be?

I.M.Y. MAREELS *, R.R. BITMEAD, M. GEVERS

Department of Systems Engineering, Research School of Physical Sciences, P.O. Box 4, Australian National University, Canberra, A.C.T. 2601, Australia

C.R. JOHNSON ** Jr

School of Electrical Engineering, Cornell University, Ithaca, NY, USA

R.L. KOSUT

Integrated Systems Inc., Palo Alto, CA and Department of Electrical Engineering, Stanford University, Stanford, CA, USA

M.A. POUBELLE

Turramurra Institute of Technological Sciences, N.S.W., Australia

Received 12 May 1986
Revised 13 August 1986

Abstract: The rate of parameter convergence in a number of adaptive estimation schemes is related to the smallest eigenvalue of the average information matrix determined by the regression vector. Using a very simple example, we illustrate that the input signals that maximize this minimum eigenvalue may be quite different from the input signals that optimize more classical input design criteria, e.g. D-optimal criterion.

Keywords: Exponential convergence, Persistence of excitation, Experiment design.

1. Preamble

The concept of 'persistently exciting' (PE) signals has invaded the adaptive systems literature at an exponential rate. Currently, many papers on adaptive estimation or adaptive control contain long derivations proving that there exists some $T > 0$, some $t_0 \geqslant 0$, and some $\alpha > 0$, $\beta > 0$ such that a certain regression vector $\phi(t)$ satisfies the following condition:

$$\beta I \geqslant \frac{1}{T}\int_t^{t+T} \phi(\tau)\phi^{\mathrm{T}}(\tau)\,\mathrm{d}\tau \geqslant \alpha I \quad \text{for all } t \geqslant t_0. \tag{1.1}$$

This is the celebrated persistency of excitation condition. The regression vector $\phi(t)$ can take many forms, depending on the problem, but the following form is typical:

$$\phi^{\mathrm{T}}(t) = \frac{1}{(2+\gamma)^{n-1}}\left[u(t) \quad \dot{u}(t) \quad \cdots \quad u^{(n-1)}(t) \quad y(t) \quad \dot{y}(t) \quad \cdots \quad y^{(n-1)}(t)\right]. \tag{1.2}$$

We have assumed a single input single output (SISO) system for simplicity, with input $u(t)$ and output $y(t)$; γ is a positive constant and n is the order of the system.

It is beyond the reach of this short technical note to dwell on the many occurrences of the persistency of excitation condition in adaptive estimation and adaptive control theory, but for those readers unfamiliar

* Research Assistant with the National Fund for Scientific Research, Belgium, whose support is acknowledged.
** Supported by NSF Grants No. 85-13400.

0167-6911/87/$3.50

with this field let us just say that this condition is often appealed to to establish the exponential convergence of a linear time-varying error system. This insures the exponential convergence of all internal variables to their desired values in the idealized case (constant system, exact model matching, etc.) and their boundedness in certain non-ideal cases (time-varying parameters, unmodelled dynamics, etc.). And so it goes.

The simplest and most informative occurrence of the PE condition is in the analysis of gradient algorithms for the estimation of a parameter vector. The error equations have the form

$$\dot{\theta}(t)=\varepsilon\phi(t)e(t) \tag{1.3}$$

where $\phi(t)$ is the parameter estimation error, $\varepsilon>0$ is the adaption gain, $\phi(t)$ is the regression vector and $e(t)=-\phi^{\mathrm{T}}(t)\ \theta(t)$ is an error signal. It can be shown that, subject to ϕ satisfying (1.1), (1.3) is uniformly exponentially convergent to zero. Further, if additionally ε is small (actually $\varepsilon\beta T\ll 1$) then the convergence rate of (1.3) is bounded below by

$$k=\varepsilon\lambda_{\min}(R)=\mathrm{O}(\varepsilon^2)\quad\text{where}\quad R=\lim_{T\to\infty}\frac{1}{T}\int_0^T\phi(\tau)\phi^{\mathrm{T}}(\tau)\,\mathrm{d}\tau$$

assuming this limit exists. That is,

$$|\theta(t+\tau)|\leqslant K\,\mathrm{e}^{-k\tau}|\theta(t)|$$

for all t and τ and some $1\leqslant K<\infty$ fixed, with k being approximately linear in ε [1,2,3].

One is then drawn to ask how to maximize the convergence rate of the error system (1.3) by manipulating designer variables – specifically ε and $u(t)$. In most adaptive situations the algorithm gain ε is constrained to be small relative to the regressor magnitude by the requirements of noise rejection – the variance of the parameter error in adaptive filtering is typically proportional to $\varepsilon\beta^{1/2}$ [4] – so that the small ε assumption concurs with engineering dictates. The meaningful subproblem then is: Given that ε is already small (to keep $\varepsilon\beta^{1/2}$ small), and that if $\varepsilon\beta T$ is small, the convergence rate is proportional to $\varepsilon\alpha$, how can we choose $u(t)$ to maximize α and/or α/β?

This is clearly an optimal input design type question. Input design was more fashionable a decade ago in the system identification literature, where persistency of excitation also originated. However, most of the effort was aimed at maximizing the determinant of the information matrix (this is called D-optimality), rather than its minimum eigenvalue, or the inverse of its condition number. In this note we examine the very simple case:

$$\phi(t)=\begin{vmatrix}u(t)\\ y(t)\end{vmatrix},\qquad H(s)=\frac{b}{s+a},\qquad y(t)=H(s)u(t),$$

and we solve the optimal input design problem for three different criteria. We seek the input $u(t)$ that maximizes, respectively, $\lambda_{\min}(R)$, $\lambda_{\min}(R)/\lambda_{\max}(R)$, and for comparison purposes, det R. Given the connections we have established with the convergence rate of an adaptive algorithm, we shall show that maximizing the determinant leads to a rather poor input design. The main reason for its popular use is probably the simplicity of computation of the optimal input.

Some of these issues of experiment design in an adaptive systems context have been raised before [5] simply to emphasize the connection.

One reason for our interest in this question arises from experimental attempts to generate PE signals for simple linear systems, with an adequate 'richness' of the regression vector $\phi(t)$ leading to a particular minimum convergence rate of the error variables. It is often thought that only an academic researcher with a very twisted mind could generate signals that will violate the PE condition (e.g. try $u(t)=\cos t^2$ going through a low pass filter). As it turns out, the problem is not to generate a $\phi(t)$ that satisfies (1.1) for some $\alpha>0$, but to obtain an α that is large enough to produce a reasonable convergence rate for the adaptive algorithm. In other words: How does one turn exponentially slow convergence into exponentially fast convergence? Finally, we wish to mention that we are by no means the first to discover that exponential convergence can be exceedingly slow, and some authors have conjectured that the slow convergence was probably due to a poor choice of input signal (see e.g. [6]). Given the practical importance of the question we raise, it is surprising that almost no attempts have been made to answer it. The purpose of this note is to give some very preliminary answers based on the analysis of the simplest possible case. We believe that our results provide a lot of insight, at least for us, which may help crack the more general case.

2. Example

In this section we consider the situation of a two dimensional regression vector:

$$\Phi(t) = (u(t) \quad (Hu)(t))^{\mathrm{T}}, \quad t \in \mathbb{R}, \tag{2.1}$$

consisting of the input $u(t)$ and the filtered signal $(Hu)(t)$ – where H is a strictly stable, causal, linear time invariant operator with transfer function

$$H(s) = \frac{b}{s+a}, \quad s \in \mathbb{C},\ a > 0. \tag{2.2}$$

This is typical for the adaptive identification or control of a first order plant. The design variable is the input $u(t)$, which we want to select so as to guarantee 'optimal' performance of the adaptive system. Under the mild assumption that the input allows the definition of a power spectrum [8], this boils down to investigating the properties of the matrix

$$R = \lim_{T \uparrow \infty} \frac{1}{T} \int_0^T \phi(t) \phi(t)^{\mathrm{T}}\, \mathrm{d}t. \tag{2.3}$$

The input's power spectrum is defined as

$$S(\omega) = \int_{-\infty}^{+\infty} r(\tau)\, \mathrm{e}^{\mathrm{j}\omega\tau}\, \mathrm{d}\tau, \quad \omega \in \mathbb{R}, \tag{2.4}$$

where $r(\tau)$ is by assumption Fourier transformable, and is defined via

$$r(\tau) = \lim_{T \uparrow \infty} \frac{1}{T} \int_\alpha^{\alpha+T} u(t) u(t+\tau)\, \mathrm{d}t, \quad \tau, \alpha \in \mathbb{R}, \tag{2.5}$$

where the limit exists uniformly in α (for $u(t)$ defined on $\mathbb{R}^+$, $\alpha \geqslant \min(0, -\tau)$). Under mild conditions the spectrum uniquely determines the input.

Obviously (2.5) and (2.2) imply the existence of R (2.3), moreover R is given by

$$R = \begin{vmatrix} \int_{-\infty}^{+\infty} S(\omega)\, \mathrm{d}\omega & \int_{-\infty}^{+\infty} S(\omega) \operatorname{Re} H(\mathrm{j}\omega)\, \mathrm{d}\omega \\ \int_{-\infty}^{+\infty} S(\omega) \operatorname{Re} H(\mathrm{j}\omega)\, \mathrm{d}\omega & \int_{-\infty}^{+\infty} S(\omega) \,|\, H(\mathrm{j}\omega) \,|^2\, \mathrm{d}\omega \end{vmatrix} \tag{2.6}$$

where in this case

$$\operatorname{Re} H(\mathrm{j}\omega) = \frac{b}{a} \frac{1}{(\omega^2/a^2) + 1}, \tag{2.7}$$

$$|\, H(\mathrm{j}\omega) \,|^2 = \frac{b^2}{a^2} \frac{1}{(\omega^2/a^2) + 1}. \tag{2.8}$$

We compare the following three input selection criteria:

Selection Criteria. Over the class of input functions ($u(t)$, $t \in \mathbb{R}^+$), which have a power spectrum (as defined in (2.4)–(2.5)) and which satisfy the constraint

$$0 < \int_{-\infty}^{+\infty} S(\omega)\, d\omega \leqslant 1, \tag{2.9}$$

maximize, either
(C1) $\det(R)$, or
(C2) $\lambda_{\min}(R)$, or
(C3) $\lambda_{\min}(R)/\lambda_{\max}(R)$.

Solution. Define

$$\alpha = \left\{\int_{-\infty}^{+\infty} S(\omega) \frac{1}{(\omega^2/a^2)+1} d\omega\right\} \Big/ \left\{\int_{-\infty}^{+\infty} S(\omega)\, d\omega\right\} \tag{2.10}$$

and

$$M(\alpha) = \begin{vmatrix} 1 & \alpha\dfrac{b}{a} \\ \alpha\dfrac{b}{a} & \alpha\dfrac{b^2}{a^2} \end{vmatrix}.^{1} \tag{2.11}$$

Surprisingly, we have that

$$R = M(\alpha) \int_{-\infty}^{+\infty} S(\omega)\, d\omega \tag{2.12}$$

and

$$0 \leqslant \alpha \leqslant 1. \tag{2.13}$$

Therefore, the optimal input functions according to (C1) or (C2) satisfy (2.9) with equality, whilst for (C3) the magnitude of the total input power is immaterial. Consequently, the optimal inputs are characterized as:

$$\text{C1-optimal:} \quad \int_{-\infty}^{+\infty} S(\omega) \frac{1}{(\omega^2/a^2)+1} d\omega = \alpha_1^*, \qquad \int_{-\infty}^{+\infty} S(\omega)\, d\omega = 1; \tag{2.14}$$

$$\text{C2-optimal:} \quad \int_{-\infty}^{+\infty} S(\omega) \frac{1}{(\omega^2/a^2)+1} d\omega = \alpha_2^*, \qquad \int_{-\infty}^{+\infty} S(\omega)\, d\omega = 1; \tag{2.15}$$

$$\text{C3-optimal:} \quad \int_{-\infty}^{+\infty} S(\omega) \frac{1}{(\omega^2/a^2)+1} d\omega = \alpha_3^*\beta, \qquad \int_{-\infty}^{+\infty} S(\omega)\, d\omega = \beta \leqslant 1; \tag{2.16}$$

where the $\alpha_i^* \in (0, 1)$ maximize respectively $\det M(\alpha)$, $\lambda_{\min} M(\alpha)$ and $\lambda_{\min} M(\alpha)/\lambda_{\max} M(\alpha)$ over $\alpha \in (0, 1)$; and β is any number in (0, 1].

After some simple calculations, we arrive at

$$\alpha_1^* = \tfrac{1}{2}, \tag{2.17}$$

$$\alpha_2^* = \frac{2}{4 + (b^2/a^2)} \in (0, \tfrac{1}{2}), \tag{2.18}$$

$$\alpha_3^* = \frac{1}{2(b^2/a^2) + 1} \in (0, 1). \tag{2.19}$$

[1] This matrix M should not be confused with the M matrix of Poubelle et al. [7].

Table 1

Criterion	ω^*	det R	$\lambda_{\min}(R)$	$\lambda_{\min}(R)/\lambda_{\max}(R)$
C1	a	$\frac{1}{4}\frac{b^2}{a^2}$	$\frac{1}{2}\left(\left(1+\frac{b^2}{2a^2}\right)-\left(1+\frac{1}{4}\frac{b^4}{a^4}\right)^{1/2}\right)$	$\dfrac{1+\frac{b^2}{2a^2}-\left(1+\frac{1}{4}\frac{b_4}{a^4}\right)^{1/2}}{1+\frac{b^2}{2a^2}+\left(1+\frac{1}{4}\frac{b^4}{a^4}\right)^{1/2}}$
C2	$a\left(1+\frac{b^2}{2a^2}\right)^{1/2}$	$\dfrac{2\left(2+\frac{b^2}{a^2}\right)\frac{b^2}{a^2}}{\left(4+\frac{b^2}{a^2}\right)^2}$	$\dfrac{\frac{b^2}{a^2}}{4+\frac{b^2}{a^2}}$	$\frac{4a^2}{b^2}+2$
C3	$a\left(2\frac{b^2}{a^2}\right)^{1/2}$	$\dfrac{2\frac{b^4}{a^4}}{\left(2\frac{b^2}{a^2}+1\right)^2}$	$\dfrac{\left(3\frac{b^2}{a^2}+1\right)-\left(\frac{b^4}{a^4}+6\frac{b^2}{a^2}+1\right)^{1/2}}{2\left(2\frac{b^2}{a^2}+1\right)}$	$\dfrac{3\left(\frac{b^2}{a^2}+1\right)-\left(\frac{b^4}{a^4}+6\frac{b^2}{a^2}+1\right)^{1/2}}{\left(3\frac{b^2}{a^2}+1\right)+\left(\frac{b^4}{a^4}+6\frac{b^2}{a^2}+1\right)^{1/2}}$

Equations (2.14)–(2.19) characterize all 'optimal' solutions. In order to get some more insight, we verify whether there exist optimal inputs of the form

$$u(t)=\sqrt{2}\cos\omega^* t,\quad t\in\mathbb{R}^+,\ \omega^*\in\mathbb{R}^+, \tag{2.20}$$

with power spectrum

$$S(\omega)=\tfrac{1}{2}(\delta(\omega-\omega^*)+\delta(\omega+\omega^*)). \tag{2.21}$$

Solving for ω^* we find respectively for C1, C2 and C3:

$$\omega_1^*=a, \tag{2.22}$$

$$\omega_2^*=a\left(1+\frac{b^2}{2a^2}\right)^{1/2}, \tag{2.23}$$

$$\omega_3^*=a\left(2\frac{b^2}{a^2}\right)^{1/2}, \tag{2.24}$$

For this type of input (2.20) we collected in Table 1 the relevant quantities (det R, $\lambda_{\min}(R)$, $\lambda_{\min}(R)/\lambda_{\max}(R)$), as a function of b and a. In Table 2, the same quantities are displayed for $b/a=1$.

For the purely sinusoidal input $u(t)$ (2.20), the dependence of the determinant of R on the frequency ω is displayed in Figure 1. (Notice that the determinant is normalized by the D.C. gain squared.) The minimum eigenvalue and the condition number are displayed as functions of frequency respectively in

Table 2
(D.C. gain = 1)

Criterion	ω^*	det R	$\lambda_{\min}(R)$	$\lambda_{\min}(R)/\lambda_{\max}(R)$
C1	a	$\frac{1}{4}$	$\frac{3-\sqrt{5}}{4}$	$\frac{3-\sqrt{5}}{3+\sqrt{5}}$
C2	$a\sqrt{\frac{3}{2}}$	$\frac{6}{25}$	$\frac{1}{5}$	$\frac{1}{6}$
C3	$a\sqrt{2}$	$\frac{2}{9}$	$\frac{2-\sqrt{2}}{3}$	$\frac{2-\sqrt{2}}{2+\sqrt{2}}$

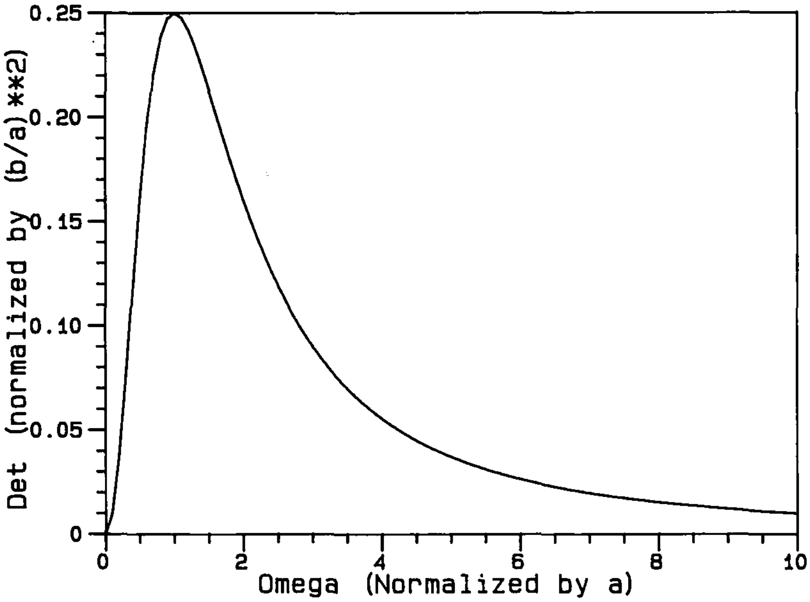

Fig. 1. Determinant.

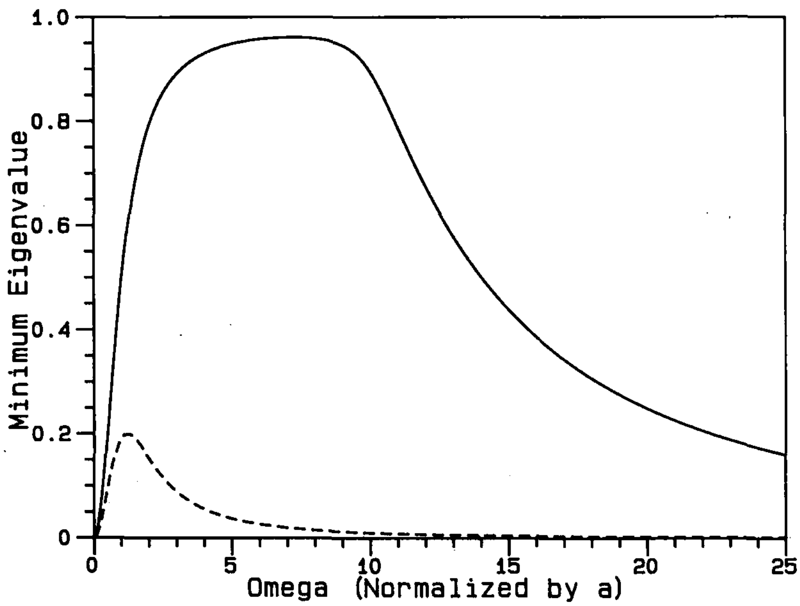

Fig. 2. Minimum eigenvalue.

Figure 2 and 3. The full line corresponds to a D.C. gain of 10, whilst the dotted line corresponds to a D.C. gain of 1.

3. Discussion

Consideration of input sequences that maximize the minimum eigenvalue of R (as in criterion (C2) of the preceding section) is encouraged in the introduction. This point is argued persuasively, and with more

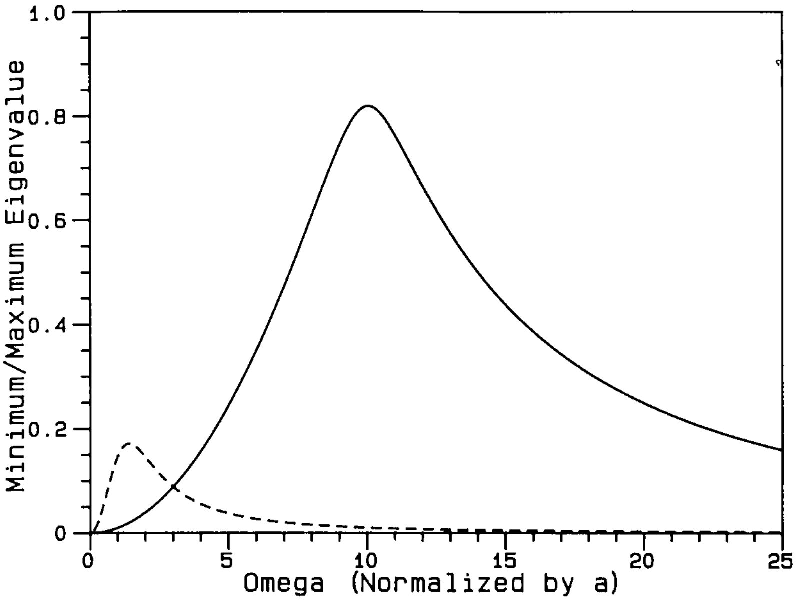

Fig. 3. Inverse of condition number.

detail, in [5]. What is not examined in [5] is the difference in the subsequent input choice relative to the more common objective of determinant maximization [8]. As stated earlier, the purpose of this note is to draw attention to this difference via examination of a simple example.

1. The most immediate observation is that maximizing the minimum eigenvalue of the information matrix yields a different 'optimal' input sequence from the one derived by maximizing the determinant or the ratio of the minimum and maximum eigenvalues. This is not really surprising as it concerns quite different optimization criteria.

2. In the first order example of Section 2, the frequency ω^* of the selected sinusoid is the breakpoint (or 3dB) frequency of the plant in (2.2) with determinant maximization; while ω^* is larger for minimum eigenvalue and minimum-to-maximum eigenvalue ratio maximization objectives. In fact, as the D.C. gain (b/a) of the plant increases, so do the selected input frequencies for the minimum and minimum/maximum eigenvalue maximization criteria.

3. One interpretation of the tradeoff inherent with ω^* selection for minimum eigenvalue maximization is its tendency to make R in (2.12) or equivalently $M(\alpha)$ in (2.11) equal the identity matrix by attempting to keep the plant gain close to one while simultaneously attempting to achieve a 90° phase shift to null the off-diagonal terms on average. Table 2 indicates the compromise between these conflicting objectives when $b/a = 1$. This interpretation also explains why the input frequency that maximizes $\lambda_{\min}(R)$ increases as the plant D.C. gain increases.

4. We should also note the nonuniqueness of the 'optimal' u, unless, as in our example, the input power is constrained and $\{u\}$ is assumed to be composed of a number of sinusoids. We refer to [8, Chapter 6] for further comment.

5. One extension of 'optimal' input selection would be to incorporate a measure of sensitivity to imprecisely known plant parameters. Such plant model imprecision is actually the motivation for identification procedures and the associated input selection. The example in Section 2 clearly indicates that the 'optimal' input, by the various criteria, is a function of the 'unknown' plant parameters.

6. We note that in this example the sensitivity of the $\lambda_{\min}$ design criterion is better than for the determinant criterion as is clear from Figures 1 and 2. Further, for all three criteria the penalty for using higher than optimal frequency appears less than that for using a lower frequency than optimal, and the sensitivity for the $\lambda_{\min}$ criterion in this example improves with increasing D.C. gain. This criterion is the one of prime interest for the convergence rate.

Finally, as discussed in [5], optimal input design questions are perhaps better posed in an adaptive estimation context than in an off-line identification situation. This is because, as the adaptive identifier learns more about the system, the input signals can be adjusted according to the relevant criterion. The insensitivity to imprecise knowledge of system parameters is then clearly advantageous and of relevant concern.

4. Conclusions

We have argued that, in the case of slow adaptation, the smallest eigenvalue and the condition number of the average information matrix determined by the regression vector should be considered as input design criteria in order to maximize the rate of exponential convergence. We have then performed this optimal input design in the simplest possible case, which allows a complete description of all optimal solutions. One should be very careful in extrapolating the conclusions of this simple example to more general situations.

More attention has to be paid to the effects of modelling errors, noise spectra, etc. which tend to complicate matters even further. Including observations of this sort leads one to concentrate the energy of the input signal in low frequency regimes [1,9] which as our results indicate, causes an attendant decrease in the level persistency of excitation which in its turn may accentuate the effect of unmodelled dynamics [1]. Given this dilemma, and its practical importance in designing well performing (robust) adaptive control/estimation algorithms, we believe that this problem deserves much more attention but its resolution will probably depend strongly on the particular circumstances.

References

[1] B.D.O. Anderson, R.R. Bitmead, C.R. Johnson Jr, P.V. Kokotovic, R.L. Kosut, I.M.Y. Mareels, L. Praly and B.D. Riedle, *Stability of Adaptive Systems: Passivity and Averaging Analysis* (MIT Press, Cambridge, MA, 1986).

[2] A.P. Morgan and K.S. Narendra, On the uniform asymptotical stability of certain linear non-autonomous differential equations, *SIAM J. Control Optim.* **15** (1) (1977) 5–24.

[3] R.L. Kosut, B.D.O. Anderson and I.M.Y. Mareels, Stability theory for adaptive systems: methods of averaging and persistency of excitation, *Proc. 24th IEEE Conf. Decision and Control*, Florida (Dec. 1985) pp. 478–483; *IEEE Trans. Automat. Control*, to appear.

[4] B. Widrow, J.M. McCool, M.G. Larimore and C.R. Johnson Jr, Stationary and nonstationary learning characteristics of the LMS adaptive filter, *Proc. IEEE* **45** (8) (1976) 1151–1162.

[5] R.R. Bitmead, B.D.O. Anderson and T.S. Ng, Convergence rate determination for gradient-based adaptive estimators, *Automatica* **22** (1986).

[6] K.S. Narendra, Stable identification schemes in: R.K. Mehra and D.G. Lainiotis, Eds., *System Identification: Advances and Case Studies* (Academic Press, New York, 1976).

[7] M.A. Poubelle, I.R. Peterson, M.R. Gevers, and R.R. Bitmead, A miscellany of results on an equation of Count J.F. Riccati, *IEEE Trans. Automat. Control* **31** (1986).

[8] G.C. Goodwin and R.L. Payne, *Dynamic System Identification: Experiment Design and Data Analysis* (Academic Press, New York, 1977).

[9] B.D. Riedle and P.V. Kokotovic, A stability-instability boundary for disturbance-free slow adaptation with unmodelled dynamics, *IEEE Trans. Automat. Control* **30** (10) (1985) 1027–1030.