INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING
Int. J. Adapt. Control Signal Process. 2001; **15**:393–413 (DOI: 10.1002/acs.630)

Generalized forgetting functions for on-line least-squares identification of time-varying systems

R. E. Mahony*,† and R. Lozano

Heudiasyc–UTC UMR 6599, Centre de Recherche de Royallieu, BP 20529, 60205 Compiegne Cedex, France

SUMMARY

The problem of on-line identification of a parametric model for continuous-time, time-varying systems is considered via the minimization of a least-squares criterion with a forgetting function. The proposed forgetting function depends on two time-varying parameters which play crucial roles in the stability analysis of the method. The analysis leads to the consideration of a Lyapunov function for the identification algorithm that incorporates both prediction error and parameter convergence measures. A theorem is proved showing finite time convergence of the Lyapunov function to a neighbourhood of zero, the size of which depends on the evolution of the time-varying error terms in the parametric model representation.

1. INTRODUCTION

Historically, the analysis of on-line identification algorithms has been closely linked to the field of adaptive control. Early work in the identification of parametric models of linear systems [1, 2] was motivated by control performance requirements. This work leads to the development of on-line identification algorithms [3, 4] designed to be run in parallel with a control algorithm using the certainty equivalence methodology. Stability for the closed-loop systems requires that the system signals which are used to drive the adaptation are themselves bounded. Unfortunately, this 'boundedness' condition may be broken exactly when certainty equivalence control is applied to the system. Work in the late 1970s [5, 6] (among many) showed that this difficulty could be overcome if the system signals used in the identification algorithms are normalized. Even then, these algorithms were only guaranteed to work effectively when the true parameters of the system were constant (though unknown). In the late 1980s there was a considerable amount of work done to understand the case of time-varying parameters [7–9]. A powerful technique emerging during this period was the use of averaging theory to show that at least for slowly varying plant

* Correspondence to: R.E. Mahony, Heudiasyc-UTC UMR 6599, Centre de Recherche de Royallieu, BP 20529, 60205 Compiegne Cedex, France.
† E-mail: mahony@hds.utc.fr

Received 10 May 1999
Revised 28 January 2000
Accepted 27 June 2000

Published online 22 March 2001

parameters closed-loop stability results for adaptive control algorithms could still be obtained [10]. Work towards understanding the time-varying case is continuing to this day and a recent review of the properties of the most common parameters estimation algorithms can be found in de Mathelin *et al*. [11] with a focus on the problem of time-varying parameters. There are two classical algorithms used for on-line identification of time-varying plants, gradient type algorithms and least-squares algorithms with forgetting. Gradient type identification algorithms have tended to be prefered due to the non-vanishing nature of the gain. However, such algorithms perform poorly in the presence of noise and when the parameter estimates are relatively slowly varying. In contrast least-squares-based algorithms with forgetting preserve many of the desirable features of a full least-squares identification procedure. Least-squares algorithms with constant rate forgetting were introduced in the late 1970s [12]. For discrete time systems some work was done as early as the early 1980s into variable rate forgetting in a least-squares framework [13, 14]. Other authors have addressed this question several times since, however, according to Landau *et al*. [15, p. 500] the efficiency of such schemes have yet to be properly determined. Recent work by the present authors in this direction is in paper [16]. A related area of considerable interest is in the design of directional forgetting algorithms [17–20]. Such algorithms are aimed at avoiding the difficulty associated with forgetting important data when the plant signals are not persistently exciting. At present most such algorithms are derived in discrete-time format and lack a strong interpretation in terms of least-squares error cost.

In this paper, we study the problem of on-line identification of a parametric model for continuous-time, time-varying systems using a least-squares criterion with a forgetting function. The application envisaged is one in which certain *a priori* bounds on the time variation of the plant is available on-line and that this variation is not regular. For example, many industrial processes display periods of relatively constant dynamic response interspersed with occasional periods when the system characteristics change quickly. Such an effect may be caused by a number of factors but is most commonly due to the intermittent arrival of new feed material. Thus, we consider systems where, at certain times it is desirable to forget old data quickly due to rapid variation of the parameters, while at other times a slow forgetting rate is desirable to improve system performance. We search for a well-motivated on-line identification algorithm to deal with this situation within the framework of parametric linear models. Parametric linear models are particularly suitable for applications in adaptive control design since they provide a structure which combines good prediction properties and control design capabilities in the same framework. The forgetting function used depends on two time-varying parameters which can be loosely related to a normalizing factor (similar in motivation to those used in classical on-line identification schemes) and a time-varying forgetting rate. Both these parameters play crucial roles in the evolution of the information matrix and it is a simple matter to derive relationships between the parameters to ensure properties of the information matrix, such as constant trace, constant norm, etc. Indeed, it is important to introduce some relationship between the introduced parameters to ensure that the least-squares cost remains a good measure of the identification error. Studying the stability of the identification method one can see clear relationships emerging between the normalizing factor, the forgetting rate, the norm of the information matrix and the bounds on the time variation in the parameters. The analysis leads to the consideration of a Lyapunov function for the identification algorithm that incorporates both prediction error and parameter convergence measures. Using the proposed Lyapunov function, the convergence and stability properties of the proposed identification procedure are analysed. In particular, we prove a theorem showing finite time convergence of the Lyapunov function to a neighbourhood of zero. This ensures both approximate parameter tracking as well as guaranteeing a small prediction error.

The paper is divided into five sections including the present introductory section. In Section 2 the basic problem considered is presented and the proposed forgetting function framework is introduced. After some properties of the new forgetting function framework are described, several results are derived that are analogues of classical on-line least-squares identification results. Sections 3 and 4 are devoted to stability analysis of the proposed identification algorithm. Section 3 approaches the problem using the classical Lyapunov function that is used in the linear time-invariant case while Section 4 introduces a modified Lyapunov function and shows that this function may be more useful for analysing linear time-varying systems. Finally, the Appendix provides a brief derivation of the linear parametric model for a linear time-varying system and a more detailed discussion of how error estimates for this model may be obtained from *a priori* bounds on the time variation of the parameters.

2. GENERAL THEORY

In this section the formulation of the least-squares cost criterion is introduced with a generalized forgetting factor. It is shown how the parameters of the forgetting factor can be chosen to ensure the validity of the least-squares cost or guarantee certain properties of the information matrix.

Consider a parametric model of a single-input single-output linear time-varying system written in regressor form [21, 22]

$$y_t := \theta_t^{\mathrm{T}} \phi_t + \eta_t \tag{1}$$

Here y_t is the system output, θ_t is a time-varying vector of parameters which define the behaviour of the system, ϕ_t is a vector of (Hurwitz) filtered inputs and outputs of the system and η_t is an error term relating the mismatch of the model (1) to the true system output. A brief presentation of the derivation of the linear parametric model for a linear time-varying (LTV) system is provided in the appendix.

The performance analysis of the proposed identification procedure for time-varying systems is closely linked to the validity of the model (1). Without any *a priori* information (and no restriction) on the variation and magnitude of the parameter vector then the 'linear' parametric model (1) provides little or no predictive power for the system signals y_t and u_t. The framework presented in this paper is one in which the rate of variation of the time-varying parameters is expected to be variable. Thus, the situation considered is one where temporarily the model (1) may not provide a good model of the system dynamics, but that there will be significant periods when the parametric linear structure encoded in (1) can be a good representation of the system dynamics. We look to find a well-motivated identification algorithm which may be used to interpolate smoothly between the desired identification characteristics appropriate to the different phases of parameter variation. In the following analysis we assume the following *a priori* on-line bounds on the variation of the plant parameters and the error η_t:

Assumption I

The true plant parameters θ_t vary differentiably and a time-varying bound $B_t > 0$ is known such that[‡]

$$|\dot{\theta}_t| \leqslant B_t \tag{2}$$

Assumption II

A second time-varying bound $C_t > 0$ is known such that

$$|\eta_t| \leqslant C_t \tag{3}$$

In practice these bounds are best estimated from process knowledge though they may also be thought of as time-varying parameters and tuned on-line according to some criterion. An explicit discussion of how the bound C_t may be estimated from bounds on the time variation of the system parameters $\{|\dot{\theta}|, |\theta^{(2)}|, \ldots, |\theta^{(n)}|\}$ is given in the appendix. It is convenient to keep the two bounds B_t and C_t separate (even though they are related by the error dynamics of the system (39)) since they play different roles in the analysis of the proposed identification algorithm.

Remark 2.1

If the plant parameters are not truly differentiable (or even continuous) the analysis in the sequel may still be applied to an averaged version of the plant parameters. Define

$$\theta_t^{\mathrm{av}} := \frac{1}{T}\int_{t-T}^{t} \theta_\tau \, \mathrm{d}\tau$$

Using θ_t^{av} instead of θ_t in the Lyapunov function (24) leads to an analogous development as presented in Section 3 but with the bound in (2) replaced by

$$|\dot{\theta}_t^{\mathrm{av}}| \leqslant \frac{1}{T}|\theta_t - \theta_{t-T}| = B_t$$

As long as the variation in θ_t is not excessive (for example θ_t is bounded in the mean [23]) then the above expression will provide a useful bound on the variation in θ_t. If T is chosen as sufficiently small, in order that there is a time-scale separation between the evolution of the averaged parameter values and the principal modes of the plant, then the parametric model retains its validity.

Given an estimate $\hat{\theta}_t$ of the system parameters at time t a prediction of the output is given by

$$\hat{y}_t = \hat{\theta}_t^{\mathrm{T}} \phi_t \tag{4}$$

[‡] Here $|x|$ denotes the 2-norm of a vector x or the absolute value of a scalar depending on the context.

This leads naturally to the equation error

$$\varepsilon_t := \hat{y}_t - y_t = \phi_t^{\mathrm{T}}(\hat{\theta}_t - \theta_t) - \eta_t \tag{5}$$

The least-squares cost function with forgetting factor considered is given by

$$E(\hat{\theta}_t) := \int_0^t r(t,\tau)^2 \varepsilon_\tau^2 \, \mathrm{d}\tau \tag{6}$$

where $r(t,\tau)$: $\mathbb{R}^2 \to \mathbb{R}$ is a general 'forgetting factor'. In classical literature the exponential forgetting factor

$$r(t,\tau) := \mathrm{e}^{-\sigma(t-\tau)} \tag{7}$$

is well known and fully understood [12]. Note that for the classical exponential forgetting factor squaring $r(t,\tau)$ is equivalent to re-scaling σ by a factor of two. An important property of (7) is that when $t \gg \tau$ then $r(t,\tau)$ is small. This is the property that ensures that old data, from time τ long before present time t, is discounted in the integral (6) and will not count significantly to the estimate of $\hat{\theta}_t$ at time t.

The proposed *forgetting function* $r(t,\tau)$ is of the following form:

$$r(t,\tau) := \alpha(\tau)\mathrm{e}^{-(\gamma(t)-\gamma(\tau))} \tag{8}$$

where[§] $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ is a given positive function (often written $\alpha(t) = \alpha_t$) which is termed the *normalization function* and

$$\gamma(\tau) = \int_0^\tau \sigma(s)\,\mathrm{d}s \tag{9}$$

for $\sigma : \mathbb{R}_+ \to \mathbb{R}_+$ a second positive function. The function $\sigma(t) = \sigma_t$ is termed the *rate function* while $\gamma(t) = \gamma_t$ is termed the *time-scaling function*.

Remark 2.2

To motivate the choice of $r(t,\tau)$ in (8) consider a constant rate function $\sigma(t) := \sigma > 0$ and observe that $\gamma(\tau) = \sigma\tau$. If one now chooses

$$\alpha(s) = \frac{1}{\sqrt{1 + |\phi_s|^2}} \approx \frac{1}{|\phi_s|} \tag{10}$$

then the least-squares error (6) becomes

$$E(\hat{\theta}_t) := \int_0^t \frac{\mathrm{e}^{-2\sigma(t-\tau)}\varepsilon_\tau^2}{(1 + |\phi_\tau|^2)} \,\mathrm{d}\tau$$

[§] Denote the positive reals by $\mathbb{R}_+ = \{x \in \mathbb{R} | x \geqslant 0\}$.

which is the classical normalized least-squares error function with exponential forgetting [24].

□

As seen in Remark 2.2 the σ function can be thought of as an instantaneous rate of forgetting. The function γ which integrates this rate then acts as a sort of scaled time axis with its instantaneous scaling rate equal to σ_t

$$\dot{\gamma}(\tau) = \sigma(\tau)$$

The role of α_t can also be motivated by considering what happens in classical exponential forgetting framework when $\sigma \gg 0$ is taken very large. In this case, the effective energy (for bounded signals ϕ_t) in the least-squares cost $E(\hat{\theta}_t)$ goes to zero and the information that can be inferred about the error ε_t becomes negligible. If, however, α_t is increased along with σ_t then the effective weighting of very recent data is increased counteracting the tendency of large forgetting rates $\sigma_t \gg 0$ to force $E(\hat{\theta}_t) \to 0$ regardless of the value of $\hat{\theta}_t$. Clearly, a relationship between α_t and σ_t must be chosen which preserves the relevance of the cost E_t, even when σ_t is chosen very large. There are several possible ways to make such a choice presented in this paper. The following choice is based on preserving a 'norm' associated with the forgetting function $r(t, \tau)$. For purely theoretical reasons set

$$\alpha(\tau) = 1, \quad \sigma(\tau) = \tfrac{1}{2}, \qquad \text{for all } \tau \leqslant 0$$

Then, one may define a 'norm', l_t, for the forgetting function $r(t, \tau)$, defined by

$$l_t^2 = \int_{t-\infty}^{t} r(t, \tau)^2 \, \mathrm{d}\tau \tag{11}$$

where $l_0 = 1$ due to the choices of α_τ, σ_τ for negative time. The value l_t is related loosely[¶] to an $\mathscr{L}_2$ norm for $r(t, \tau)$. Consequently, preserving the value of l_t

$$l_t = 1 \quad \text{for all } t \geqslant 0 \tag{12}$$

over time should ensure that the effective energy in the least-squares cost $E_t(\hat{\theta}_t)$ is never zero. It is easily verified that (cf. Lemma 2.3)

$$\frac{\mathrm{d}}{\mathrm{d}t} l_t^2 = \alpha_t^2 - 2\sigma_t l_t^2$$

Thus, choosing

$$\alpha_t^2 = 2\sigma_t \tag{13}$$

[¶] This is a qualitative observation only since the interval of integration depends of time.

where $l_0 = 1$ is assumed without loss of generality, ensures that l_t is constant for all time. In the limiting case, $\alpha_t^2 = 2\sigma_t \to \infty$ the least-squares cost degenerates to an instantaneous cost ε_t^2. In contrast, if α_t is kept constant then $E_t \to 0$ for $\sigma_t \to \infty$.

We mention a number of properties of $r(t, \tau)$ as defined by (8),

$$r(t, t) = \alpha_t e^{(-\gamma_t - \gamma_t)} = \alpha_t \tag{14}$$

$$\lim_{t \to \infty} r(t, \tau) = \alpha_\tau \lim_{t \to \infty} e^{\gamma_\tau - \gamma_t}$$

$$= \alpha_\tau e^{\gamma_\tau} \exp\left(- \int_0^\infty \sigma(s)\, ds \right)$$

Thus $\lim_{t \to \infty} r(t, \tau) \to 0$ if and only if

$$\lim_{t \to \infty} \int_0^t \sigma(s)\, ds \to \infty$$

This second property is important since it is related to a well-defined time scaling $\gamma(t) \to \infty$ as $t \to \infty$. In particular, we will assume without loss of generality that there exists a positive number $\sigma_b > 0$ that acts as a lower bound on σ_t

$$\sigma_t \geqslant \sigma_b, \quad \text{for all } t$$

Thus, $\int_0^t \sigma(s)\, ds \geqslant t\sigma_b \to \infty$ for $t \to \infty$.

To define the least-squares estimate of $\hat{\theta}_t$ based on the cost (6) one needs the *information matrix*

$$P_t := \int_0^t r(t, \tau)^2 \phi_\tau \phi_\tau^T \, d\tau \tag{15}$$

and the vector

$$z_t := \int_0^t r(t, \tau)^2 y_\tau \phi_\tau \, d\tau \tag{16}$$

Then the (or if P_t is not full rank 'a') least-squares estimate $\hat{\theta}_t$ of θ_t is given by solving

$$P_t \hat{\theta}_t = z_t \tag{17}$$

One arrives at (17) by minimizing $E(\hat{\theta}_t)$ from (6) [21]. Recall that $\hat{\theta}_t$ is chosen at time t to be a *time-invariant* model of all the data (including the error term and weighted by the forgetting factor) over the period $[0, t]$. Thus, when θ_t is quickly time varying a good estimate $\hat{\theta}_t$ of θ_t can only be obtained by choosing the forgetting factor to die off quickly. Similarly, if η_t is large then it is expected that $\hat{\theta}_t$ will be poor estimate of θ_t, though, $\hat{y}_t$ may still be a good estimate of y_t! Finally, in closed loop with a certainty equivalence control design it is possible that the regressor vector

ϕ_t may display large transients which will tend to dominate the identification procedure. As long as the regressor does not display finite-time escape behaviour then the estimate $\hat{\theta}_t$ is still well defined and the identification algorithm remains valid. Ill-conditioning of the information matrix and target vector z_t can be avoided if necessary by adjusting the normalizing factor in the generalized forgetting function (cf. Remark 2.2). The proposed algorithm suffers from the same limitations as classical least-squares identification algorithms with respect to 'persistence of excitation' of the signals.

Lemma 2.3

Let P_t be defined by (15) with ϕ_τ generated by the model (1) and the forgetting function $r(t, \tau)$ given by (8). Then for all time that ϕ_t is defined

$$\begin{aligned} &P_t \geqslant 0 \\ &\dot{P}_t = \alpha_t^2 \phi_t \phi_t^{\mathrm{T}} - 2\sigma_t P_t \end{aligned} \tag{18}$$

Proof. First observe that (15) is an integral of a positive-semi-definite matrix and hence P_t is positive semi-definite for all time. Furthermore,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} r(t, s) &= \frac{\mathrm{d}}{\mathrm{d}t}(\alpha_s \mathrm{e}^{-(\gamma_t - \gamma_s)}) \\ &= -\dot{\gamma}_t(\alpha_s \mathrm{e}^{-(\gamma_t - \gamma_s)}) \\ &= -\sigma_t r(t, s) \end{aligned}$$

Thus, taking the derivative of (15) one gets

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} P_t &= r(t,t)^2 \phi_t \phi_t^{\mathrm{T}} + \int_0^t 2r(t, \tau) \frac{\mathrm{d}r(t, s)}{\mathrm{d}t} \phi_\tau \phi_\tau^{\mathrm{T}} \, \mathrm{d}\tau \\ &= \alpha_t^2 \phi_t \phi_t^{\mathrm{T}} - 2\sigma_t P_t \end{aligned} \tag{19}$$

□

In (18) it can be seen what effect the normalization function $\alpha(t) = \alpha_t$ and the rate function $\sigma(t) = \sigma_t$ have on the evolution of the information matrix. Moreover, these functions enter into the evolution equation only in terms of their values at time t. Thus, one may choose α_t and σ_t on-line to engender certain properties for P_t. For example, choosing

$$\sigma_t = \frac{\alpha_t^2 |\phi_t|^2}{2\mathrm{tr}(P_t)} \tag{20}$$

ensures that

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathrm{tr}(P_t) = \mathrm{tr}(\dot{P}_t) = \alpha_t^2 \mathrm{tr}(\phi_t \phi_t^{\mathrm{T}}) - 2 \frac{\alpha_t^2 \phi_t^{\mathrm{T}} \phi_t}{2\mathrm{tr}(P_t)} \mathrm{tr}(P_t) = 0$$

and thus $\mathrm{tr}(P_t) = \mathrm{tr}\, P_0$ (assuming $\mathrm{tr}(P_0) \neq 0$).

One can also normalize for the Frobenius norm $\|P\|_F^2 = \mathrm{tr}(P^T P)$ of P by choosing

$$\alpha_t^2 = \frac{2\sigma_t \|P_t\|_F^2}{\phi_t^T P_t \phi_t} \tag{21}$$

Similar, relationships for the 2-norm of P (at least where it is differentiable) or other functions of P can easily be obtained. Equations (20) and (21) are related to the condition given by (13) which ensures the norm l_t is constant in time. Indeed, in both cases there is a relation of α_t^2 to σ_t, however, when some property of the information matrix is preserved then there is an additional dependence on the data ϕ_t. In the sequel, for reasons of simplicity, we will make use only of the relationship given by (13), however, Equations (20), (21) and (10) may be of interest in other studies.

Remark 2.4

The trade-off between α_t and σ_t can be used to normalize the information matrix P_t but it is not sufficient to stop the matrix going singular in the case where persistence of excitation is missing.

The on-line identification algorithm is given by the time-varying solution $\hat{\theta}_t$ to (17). It is possible, however, to directly derive the dynamics of $\hat{\theta}_t$. The time derivation of $\hat{\theta}_t$ is obtained by deriving (17):

$$\begin{aligned} \dot{P}_t\hat{\theta}_t + P_t\dot{\hat{\theta}}_t &= \frac{\mathrm{d}}{\mathrm{d}t}\int_0^t r(t,\tau)^2 y_{t\tau}\phi_\tau \,\mathrm{d}\tau \\ &= r(t,t)^2 y_t\phi_t + \int_0^t 2r(t,\tau)\frac{\mathrm{d}r(t,\tau)}{\mathrm{d}t} y_\tau\phi_\tau \,\mathrm{d}\tau \\ &= \alpha_t^2 y_t\phi_t - 2\sigma_t z_t \end{aligned}$$

Thus, shifting $\dot{P}_t$ to the right-hand side and substituting for (18) yields

$$\begin{aligned} P_t\dot{\hat{\theta}}_t &= \alpha_t^2 y_t\phi_t - 2\sigma_t z_t - (\alpha_t^2\phi_t\phi_t^T - 2\sigma_t P_t)\hat{\theta}_t \\ &= \alpha_t^2 y_t\phi_t - 2\sigma_t z_t - \alpha_t^2\phi_t\hat{y}_t + 2\sigma_t z_t \\ &= -\alpha_t^2\varepsilon_t\phi_t \end{aligned}$$

Assuming for the moment that P_t is full rank, one obtains the dynamics

$$\dot{\hat{\theta}}_t = -\varepsilon_t\alpha_t^2 P_t^{-1}\phi_t \tag{22}$$

for $\hat{\theta}_t$. In fact, P_t^{-1} can also be generated directly as the solution of an ODE analogous to (18). This saves computing the inverse continually to obtain the latest estimate of $\hat{\theta}_t$. The details are omitted since they are not relevant to the present development. It is instructive for a moment to think of the role of α_t and P_t^{-1} in (22). The matrix P_t^{-1} is often referred to as the gain matrix for classical least squares and indeed if $P_t^{-1} = I$ (the identity) then (22) resembles a gradient identification algorithm with gain α_t^2. In more generality then the product $\alpha_t^2 P_t^{-1}$ can be thought of as a gain for the $\hat{\theta}_t$ dynamics. If the true parameters are varying quickly (or B_t is large) then one would expect to require that $\alpha_t^2 P_t^{-1}$ is also large so that $\hat{\theta}_t$ can effectively track θ_t.

3. STABILITY OF THE IDENTIFICATION METHOD

In this section the stability properties of the least squares algorithm with forgetting function given by (8) are analysed using the classical Lyapunov function commonly used for the analysis of on-line identification algorithms for linear time-invariant systems.

An important issue in identification of time-varying systems is the validity of the model chosen to represent the dynamic behaviour of a time-varying system. Linear parametric models are particularly suitable for applications in adaptive control design since such models provide a structure which provides good prediction properties and control design capabilities in the same framework. This is despite the fact that data filtering, an integral part of the system identification procedure, causes inherent inaccuracies in parametric models due to the swapping terms associated with the interaction of the data filters and the time-varying parameters. Indeed, when the plant parameters are time-varying, then even if the parameter estimates are correct one must expect a prediction error generated by the swapping terms (the error η_t) in the model (1), whereas if the prediction error is zero one must expect a parametric error which compensates the swapping terms in the model. In practice, one wishes to trade off these two extremes.

In the case of linear time-invariant (LTI) systems (for which $\theta_t := \theta$ is constant) the classical approach to studying the stability of an on-line identification algorithm is to consider the time-varying Lyapunov function $V_t = (\hat{\theta}_t - \theta)^{\mathrm{T}} P_t (\hat{\theta}_t - \theta)$, where $\hat{\theta}_t$ is the least-squares parametric estimate given by solving (17) and P_t is the information matrix given by (15). Let n be the dimension of the regressor vector ϕ_t and let I_n denote the $n \times n$ identity matrix. If there exists a positive real constant $\delta > 0$ such that $P_t > \delta I_n$ (persistence of excitation) and $\dot{V}_t$ is strictly negative for all time $t \geqslant 0$, it follows (from Lyapunov's theorem) that $\hat{\theta}_t \to \theta$ asymptotically. Indeed, for LTI systems expanding the error E_t (6) one obtains

$$
\begin{aligned}
E_t &= \int_0^t r(t,\tau)^2 (\hat{\theta}_t - \theta)^{\mathrm{T}} \phi_\tau \phi_\tau^{\mathrm{T}} (\hat{\theta}_t - \theta)\,\mathrm{d}\tau - 2\int_0^t r(t,\tau)^2 \eta_\tau \phi_\tau^{\mathrm{T}} (\hat{\theta}_t - \theta)\,\mathrm{d}\tau + \int_0^t r(t,\tau)^2 \eta_\tau^2\,\mathrm{d}\tau \\
&= V_t - 2\int_0^t r(t,\tau)^2 \eta_\tau \varepsilon_\tau\,\mathrm{d}\tau + \int_0^t r(t,\tau)^2 \eta_\tau^2\,\mathrm{d}\tau
\end{aligned} \tag{23}
$$

For LTI systems the error term η_t is an exponentially decaying signal depending only on the initial conditions and stability margin of the data filter and independent of the rate of convergence of V_t or E_t. Consequently, the integral terms in the above expression decay exponentially and $E_t \to 0$ asymptotically if and only if $V_t \to 0$. Thus, for LTI systems it is reasonable to design the identification algorithm based on minimizing E_t (LS algorithms) and analyse such algorithms using the Lyapunov function V_t (which ensures convergence of the parameters to the system parameter).

In the case of linear time-varying (LTV) systems the asymptotic connection between the classical Lyapunov function V_t and the least-squares cost E_t is somewhat more tenuous. Consider, the direct generalization of the classical Lyapunov function to the case where θ_t are the time-varying parameters of the true system

$$V_t(\hat{\theta}_t) = (\hat{\theta}_t - \theta_t)^{\mathrm{T}} P_t (\hat{\theta}_t - \theta_t) \tag{24}$$

It is still the case that if $P_t > \delta I_n$ for some positive constant $\delta > 0$ and $\dot{V}_t$ is negative for all time $t \geqslant 0$ then $\hat{\theta}_t \to \theta_t$ asymptotically. However, it is no longer true that this implies $E_t \to 0$. Consider the expansion of E_t,

$$E_t = \int_0^t r(t,\tau)^2 (\hat{\theta}_t - \theta_\tau)^{\mathrm{T}} \phi_\tau \phi_\tau^{\mathrm{T}} (\hat{\theta}_t - \theta_\tau)\,\mathrm{d}\tau - 2\int_0^t r(t,\tau)^2 \eta_\tau \varepsilon_\tau\,\mathrm{d}\tau + \int_0^t r(t,\tau)^2 \eta_\tau^2\,\mathrm{d}\tau \tag{25}$$

This equation has certain fundamental differences from the LTI case. Firstly, the error terms associated with the integral of η_t will not die away since η_t is not a decaying signal. Secondly, the time dependence of θ_τ prevents one from obtaining the Lyapunov function from the first integral term. Equation (25) is a quantitative expression of the qualitative observation that it is not possible to minimize (at the same time) both the parametric error and prediction error for a parametric model of a time-varying system.

Nevertheless, by choosing σ_t and α_t suitably certain properties of the identification algorithm may be deduced. Inspecting (25) it can be seen that when σ_t is chosen sufficiently large, the rate of decay of the forgetting function relative to the time variation of θ_t is such that the data contributing to the least-squares cost is taken over a time interval during which θ_t is effectively constant. In this case, an approximate connection between the Lyapunov function V_t and the error E_t is regained. Explicitly, as $\sigma_t \to \infty$ the first integral in (25) will approach V_t and the error terms will tend to a term which depends directly on the value of η_t at the present time.

Lemma 3.1

Let P_t be defined by (15) with ϕ_τ from (1) and the forgetting function $r(t,\tau)$ given by (8). Let $B_t, C_t > 0$ denote (possibly time-varying) bounds on the parameter variation and model error (cf. Section 2, Assumptions I and II). Then a sufficient condition to ensure $\dot{V}_t$ is negative at time t is

$$\alpha_t^2 (C_t^2 - \varepsilon_t^2) + \frac{\|P_t\|_2 B_t^2}{2\sigma_t} < 0 \tag{26}$$

Proof. Directly computing the derivative of V_t yields

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} V_t(\hat{\theta}_t) &= 2(\hat{\theta}_t - \theta_t)^{\mathrm{T}} P_t (\dot{\hat{\theta}}_t - \dot{\theta}_t) + (\hat{\theta}_t - \theta_t)^{\mathrm{T}} \dot{P}_t (\hat{\theta}_t - \theta_t) \\
&= 2(\hat{\theta}_t - \theta_t)^{\mathrm{T}} (-\alpha_t^2 \varepsilon_t \phi_t) - 2(\hat{\theta}_t - \theta_t)^{\mathrm{T}} P_t \dot{\theta}_t + (\hat{\theta}_t - \theta_t)^{\mathrm{T}} (\alpha_t^2 \phi_t \phi_t^{\mathrm{T}} - 2\sigma_t P_t)(\hat{\theta}_t - \theta_t) \\
&= -2\alpha_t^2 \varepsilon_t (\varepsilon_t + \eta_t) - 2(\hat{\theta}_t - \theta_t)^{\mathrm{T}} P_t \dot{\theta}_t + \alpha_t^2 (\varepsilon_t + \eta_t)^2 - 2\sigma_t V_t(\hat{\theta}_t) \\
&= -\alpha_t^2 \varepsilon_t^2 + \alpha_t^2 \eta_t^2 - 2\sigma_t V_t(\hat{\theta}_t) - 2(\hat{\theta}_t - \theta_t)^{\mathrm{T}} P_t \dot{\theta}_t
\end{aligned}$$

There are two 'good' terms $-\alpha_t^2\varepsilon_t^2$ and $-2\sigma_t V_t(\hat{\theta}_t)$ and two bad destabilizing terms $\alpha_t^2\eta_t^2$ and $-2(\hat{\theta}_t-\theta_t)^{\mathrm{T}}P_t\dot{\theta}_t$ arising from the time variation of the parameters.

The exact value of V_t is unknown since it depends on the unknown true parameters. However, the dependence can be factored out of consideration by completing the square and dominating the 'bad' $-2(\hat{\theta}_t-\theta_t)^{\mathrm{T}}P_t\dot{\theta}_t$ by the 'good' $-2\sigma_t V_t(\hat{\theta}_t)$ term. Let $P_t^{1/2}$ be the unique‖ square root of P_t. Then direct calculations yield

$$\begin{aligned}(\hat{\theta}_t-\theta_t)^{\mathrm{T}}P_t\dot{\theta}_t &= (P_t^{1/2}(\hat{\theta}_t-\theta_t))^{\mathrm{T}}(P_t^{1/2}\dot{\theta}_t)\\ &\leqslant |P_t^{1/2}(\hat{\theta}_t-\theta_t)||P_t^{1/2}\dot{\theta}_t|\\ &\leqslant V_t(\hat{\theta}_t)^{1/2}\|P_t^{1/2}\|_2 B_t \end{aligned}\tag{27}$$

Here, in the third line Hölder's inequality is combined with the bound on the dynamics of θ_t (Assumption 1) and the properties of the 2-norm. Substituting into the expression for $\dot{V}_t$ obtained yields

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}V_t(\hat{\theta}_t) &\leqslant \alpha_t^2(C_t^2-\varepsilon_t^2)-(2\sigma_t V_t(\hat{\theta}_t)-2V_t(\hat{\theta}_t)^{1/2}\|P_t^{1/2}\|_2B_t)\\ &= \alpha_t^2(C_t^2-\varepsilon_t^2)+\frac{\|P_t^{1/2}\|_2^2B_t^2}{2\sigma_t}-2\sigma_t\left(V_t^{1/2}-\frac{\|P_t^{1/2}\|_2B_t}{2\sigma_t}\right)^2\end{aligned}\tag{28}$$

Thus, if one ignores the term in the bracket (since it is negative anyway) and since $\|P^{1/2}\|_2^2=\|P_t\|_2$, then a sufficient condition to ensure that $\dot{V}_t$ is negative is that

$$\frac{\mathrm{d}}{\mathrm{d}t}V_t(\hat{\theta}_t)\leqslant \alpha_t^2(C_t^2-\varepsilon_t^2)+\frac{\|P_t\|_2B_t^2}{2\sigma_t}<0 \qquad \square$$

Remark 3.2

(i) Note that if one were to take $\alpha_t=1$, then for $V_t\neq 0$ one may always choose $\sigma_t \gg 0$ sufficiently large such that $\dot{V}_t<0$ is negative. Such a choice is related to the effective decrease in norm l_t (11) of the forgetting function, rather than an intrinsic property of the identification algorithm. Indeed, choosing α_t and σ_t in this manner forces $P_t\to 0$. This characteristic of the analysis motivates the choice of a constraint, either directly on P_t (i.e. keeping the trace constant), or on the effective energy l_t of $r(t,\tau)$ (13).

(ii) If B_t is large in (26) and α_t and σ_t are chosen to satisfy (13) and (26) then clearly either $\|P_t\|_2$ is small (resulting in a large value for $\|P_t^{-1}\|_2$), or equally well $\alpha_t^2=2\sigma_t$ is large. Either of these cases acts to increase the effective gain of the $\hat{\theta}_t$ dynamics in (22) as expected.

‖The matrix $P_t^{1/2}$ is obtained from the spectral decomposition of $P_t=U_t^{\mathrm{T}}\Lambda_t U_t$ (where $U_t^{\mathrm{T}}U_t=I_n$ are orthogonal and $\lambda_t=\mathrm{diag}(\lambda_i(P_t))$ are the eigenvalues of P_t) by $P_t^{1/2}=U_t^{\mathrm{T}}\,\mathrm{diag}(\sqrt{\lambda_i})U_t$. Thus, $P_t^{1/2}$ is unique and is defined for semi-definite P_t.

(iii) Discarding the 'good' squared term in (28) is likely to lead to a slightly unrealistic anlaysis of the asymptotic behaviour of the identification algorithm since for $V_t \approx 0$ the discarded term would have effectively cancelled the 'bad' term $||P_t||_2 B_t^2/2\sigma_t$. However, (26) as written contains no unknown terms and can be used in a control design analysis.

Despite the approximations used in obtaining (26), this equation still provides considerable insight into the relationship between the gains α_t, σ_t, $||P_t||_2$ and the bounds B_t, C_t. For example, if $|\varepsilon_t| > |C_t|$ then $(C_t^2 - \varepsilon_t^2) < 0$ and by choosing α_t sufficiently large one may ensure that $\dot{V}_t < 0$. Moreover, it is clear that choosing σ_t sufficiently large will have the same effect. Indeed, if $|\varepsilon_t| > C_t$ then choosing $\alpha_t^2 = 2\sigma_t$ to satisfy (13) and substituting into (26) yields

$$\sigma_t^2 > \frac{||P_t||_2 B_t^2}{4(\varepsilon_t^2 - C_t^2)} \tag{29}$$

a sufficient condition for $\dot{V}_t < 0$. As $\varepsilon_t \to C_t$ then it is impossible to choose σ_t uniformly bounded to ensure (29) holds. This is directly related to the inability of the parametric model to fully represent the time-varying dynamics of the system. However, (29) can be used to generate an algorithm which should result in bounded identification performance. That is, for a given constant $\Delta > 0$ choosing

$$\sigma_t = \max\{\hat{\sigma}_t, \sigma_{\mathrm{b}}\}$$

where $\sigma_{\mathrm{b}} > 0$ is a given positive constant (ensuring the time-scaling does not degenerate) and $\hat{\sigma}_t$ is given by

$$\hat{\sigma}_t := \begin{cases} \dfrac{||P_t||_2 B^2}{4(\varepsilon_t^2 - C_t^2)} & \text{for } |\varepsilon_t| > \Delta + C_t \\ \dfrac{||P_t||_2 B^2}{4(\Delta^2 - 2\Delta C_t)} & \text{for } |\varepsilon_t| \leqslant \Delta + C_t \end{cases}$$

The choice of $\hat{\sigma}_t$ ensures that $\dot{V}_t$ is negative for all $\varepsilon_t^2 > \Delta + C_t^2$. If the model (1) is a good representation of the process then forcing V_t small should be equivalent to a small error ε_t. Thus, choosing σ_t according to the above algorithm should act to reduce V_t, and by association ε_t, until the performance bound $\varepsilon_t^2 \approx \Delta + C_t^2$ is reached.

The algorithm informally presented above is unsatisfying since it relies on a non-explicit relationship between ε_t and V_t. Rather than attempt to derive a more practical algorithm based on the above approach, Section 4 introduces a different Lyapunov function which can be used to obtain more quantitative information on the performance of the on-line identification algorithm.

4. FURTHER STABILITY ANALYSIS

In the previous section, the proposed on-line identification algorithm was analysed using a classical Lyapunov function (24) for on-line identification algorithms. In this section a slightly more general Lyapunov function is considered which may yield more useful results for LTV systems. The new Lyapunov function is used to choose a forgetting factor and normalizing factor and then to prove a theorem giving a convergence and stability results for the proposed identification algorithm.

The dual objectives of the proposed identification algorithm of minimizing prediction error as well as the parametric error leads one to consider the following Lyapunov function:

$$W_t(\hat{\theta}_t) = V_t(\hat{\theta}_t) + E_t(\hat{\theta}_t) \tag{30}$$

The function W_t combines a measure of the parametric error, V_t, along with the prediction error E_t. A further advantage of considering W_t is that the memory provided by the integral $E_t(\hat{\theta}_t)$ should help reduce sensitivity to quick changes in output or plant parameters.

Before continuing with an analysis of W_t, it is helpful to first consider the least-squares cost E_t directly. As a tool for the analysis of the identification algorithm, this cost has the disadvantage that it does not directly measure the convergence of the parameters $\hat{\theta}_t \to \theta_t$. However, its value is known exactly (unlike V_t) and it does measure the relative performance of the prediction error. The time derivative of E_t is simply

$$\dot{E}_t = \alpha_t^2 \varepsilon_t^2 - 2\sigma_t E_t$$

The best way of thinking of E_t is as a first order (time-varying) filtered version of ε_t^2

$$E_t = \frac{\alpha_t^2}{\mathrm{d}/\mathrm{d}t + 2\sigma_t} \varepsilon_t^2$$

Choosing α_t and σ_t to satisfy (13) (so that any change in E_t is not due to a change in the underlying norm of $r(t, \tau)$) leads to

$$\dot{E}_t = 2\sigma_t(\varepsilon_t^2 - E_t)$$

Given that $\sigma_t > 0$, it follows that E_t is decreasing if and only if $\varepsilon_t^2 < E_t$. Clearly, this equation contains no useful information for the choice of σ_t.

Consider the Lyapunov function $W_t = V_t + E_t$. The time derivative of W_t is

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} W_t &= \dot{V}_t + \dot{E}_t \\ &= -\sigma_t^2 \varepsilon_t^2 + \alpha_t^2 \eta_t^2 - 2\sigma_t V_t(\hat{\theta}_t) - 2(\hat{\theta}_t - \theta_t)^{\mathrm{T}} P \dot{\theta}_t + \alpha_t^2 \varepsilon_t^2 - 2\sigma_t E_t \\ &= \alpha_t^2 \eta_t^2 - 2\sigma_t W_t - 2(\hat{\theta}_t - \theta_t)^{\mathrm{T}} P_t \dot{\theta}_t \end{aligned}$$

In particular, there is a cancellation of the terms $\pm \alpha_t^2 \varepsilon_t^2$ present due to the contributions from $\dot{V}_t$ and $\dot{E}_t$. Using (27) and the bound $|\eta_t| \leqslant C_t$ yields

$$\frac{\mathrm{d}}{\mathrm{d}t} W_t \leqslant \alpha_t^2 C_t^2 - 2\sigma_t W_t + 2V_t^{1/2} \|P_t^{1/2}\|_2 B_t$$

Since $W_t \geqslant V_t$ (indeed the difference is $E_t > 0$) then $V_t^{1/2} \leqslant W_t^{1/2}$ and hence

$$\frac{\mathrm{d}}{\mathrm{d}t} W_t \leqslant -2\sigma_t W_t + 2\|P_t^{1/2}\|_2 B_t W_t^{1/2} + \alpha_t^2 C_t^2 \tag{31}$$

Equation (31) is of practical interest for several reasons. Though a number of approximations and bounds are used in its derivation it is clear that in the worst-case analysis this bound is tighter than (26) since no 'good' terms have been discarded. As a result, (31) contains unknown terms depending on the unknown value of V_t. These terms, however, are expressed as functions of W_t and all other terms in the expression are either general bounds or the parameters α_t and σ_t. Thus, it is possible to obtain a clear result relating the bounds C_t and B_t on the time-varying system parameters to the performance of the identification scheme.

Lemma 4.1

Let P_t be defined by (15) with ϕ_τ from (1) and the forgetting function $r(t, \tau)$ given by (8). Let B_t, $C_t > 0$ denote (possibly time-varying) bounds on the parameter variation and model error (cf. Section 2, Assumptions I and II). Let $\alpha_t^2 = 2\sigma_t > 0$ be chosen to satisfy (13). If

$$W_t > C_t^2$$

then there exists $\alpha_t > 0$, $\sigma_t > 0$ such that

$$\frac{\mathrm{d}}{\mathrm{d}t} W_t < 0$$

Proof. Given the assumptions of the lemma hold at time $t \geqslant 0$ then there exists $\Delta_t > 0$ such that

$$W_t = C_t^2 + \Delta_t$$

Substituting this into (31) to replace the W_t terms yields

$$\frac{\mathrm{d}}{\mathrm{d}t} W_t \leqslant -2\sigma_t C_t^2 - 2\sigma_t \Delta_t + 2\|P_t^{1/2}\|_2 B_t (C_t^2 + \Delta_t)^{1/2} + \alpha_t^2 C_t^2$$

Substituting $2\sigma_t = \alpha_t^2$ from (13) gives

$$\frac{\mathrm{d}}{\mathrm{d}t} W_t \leqslant -2\sigma_t \Delta_t + 2\|P_t^{1/2}\|_2 B_t (C_t^2 + \Delta_t)^{1/2}$$

Thus, it is clear that for all $\Delta_t > 0$, choosing

$$\sigma_t > \frac{\|P_t^{1/2}\|_2 B_t (C_t^2 + \Delta_t)^{1/2}}{\Delta_t}$$

ensures that $(\mathrm{d}/\mathrm{d}t) W_t < 0$.

It is interesting to note that Lemma 4.1 conforms with ones intuition in the sense that the performance limits on the identification algorithm are generated by the time-varying bound C_t on the model error while the bound on the parameter variation contributes only to modify the required value of σ_t, and consequently the rate of adaptation of $\hat{\theta}_t$.

In practice, it is of interest to know about the convergence properties (given poor initialization) of the proposed identification algorithm and its subsequent tracking properties. In practice, for online identification processes it is necessary to begin with an estimate of all the unknown plant and data filter parameters as well as the parameters of the identification scheme. Alternatively, should there be a jump discontinuity in parameter estimates and input/output signals during the time-evolution of the plant then one wishes to know how quickly the identified model will converge to the new (time-varying) parameters. Thus in the following theorem, arbitrary initial conditions for $P_0 > 0$ (symmetric positive-definite), ϕ_0, $\hat{\theta}_0 \in \mathbb{R}^n$ are chosen and both the transient and tracking performance of the proposed identification algorithm are analysed in terms of the evolution of the error bounds B_t and C_t.

Theorem 4.2

Let (1) represent the dynamics of a continuous-time, time-varying system and let $B_t, C_t > 0$ denote (time-varying) bounds on the parameter variation and model error (cf. Section 2, Assumptions 1 and 2) and let $\sigma_b > 0$ be positive constant under-bounding the rate function σ_t. Let $P_0 > 0$ (symmetric positive-definite), ϕ_0, $\hat{\theta} \in \mathbb{R}^n$ be given initial conditions at time $t = 0$. Let $\Delta > 0$ be a fixed constant parameter and choose

$$\frac{1}{2}\alpha_t^2 = \sigma_t = \max\left\{\frac{\|P_t^{1/2}\|_2 B_t (C_t^2 + \Delta)^{1/2}}{\Delta}, \sigma_b\right\} \tag{32}$$

Then (i) there exists T_0 such that for all $t \in (0, T_0)$

$$\dot{W}_t < 0$$

and (ii) for all $t \geqslant T_0$

$$W_t \leqslant C_t^2 + 2\Delta$$

Proof. First consider the case where at time $t = 0$, $W_0 > C_0^2 + 2\Delta$. Since the evolution of W_t is continuous then there exists an interval $(0, T_0)$ (with T_0 possibly infinite) such that $W_t > C_t^2 + 2\Delta$ for all $t \in (0, T_0)$. For $t \in (0, T_0)$ define $K_t > 0$ such that

$$W_t = C_t^2 + \Delta + K_t \tag{33}$$

By inspection $K_t \geqslant \Delta > 0$. Substituting the expression for W_t into (31) and recalling that $\alpha^2 = 2\sigma_t$ yields

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}W_t &\leqslant -2\sigma_t C_t^2 - 2\sigma_t\Delta_t - 2\sigma K_t + 2\|P_t^{1/2}\|_2 B_t (C_t^2 + \Delta_t + K_t)^{1/2} + \alpha_t^2 C_t^2 \\ &\leqslant -2\sigma_t\Delta_t + 2\|P_t^{1/2}\|_2 B_t (C_t^2 + \Delta_t)^{1/2} - 2\sigma K_t\end{aligned}$$

Now σ_t is certainly equal to or larger than the first value in the maximum operation in (32). Substituting into the above expression and cancelling terms yields

$$\frac{\mathrm{d}}{\mathrm{d}t}W_t \leqslant -2\sigma K_t \leqslant -2\Delta$$

for $t \in (0, T_0)$. In particular, since $\Delta > 0$ is constant then it follows W_t decreases to $C_t^2 + 2\Delta_t$ in finite time. If at time $t = 0$, $W_0 \leqslant C_0^2 + 2\Delta$ then choose $T_0 = 0$. Thus, it has been shown that $T_0 < \infty$ is finite and the first part (i) of the theorem is proved.

Part (ii) of the theorem follows from an analogous argument to Part (i). If W_t lies in the interval

$$W_t \in (C_t^2 + \Delta, C_t^2 + 2\Delta)$$

then once again W_t can be written as in (33), though in this case $K_t \geqslant 0$ is not necessarily strictly positive. Nevertheless, this is sufficient to ensure $\dot{W}_t \leqslant 0$ and consequently that

$$W_t \leqslant C_t^2 + 2\Delta$$

remains bounded by the upper bound of this interval for all following times. □

Remark 4.3

An upper bound for the finite time T_0 may be derived directly from an upper bound on the initial value W_0 of the Lyapunov function W_t,

$$T_0 \leqslant \frac{W_0}{2\Delta}$$

APPENDIX: ERROR SIGNALS FOR PARAMETRIC MODELS OF LTV SYSTEMS

In this appendix, a brief development of the derivation of the parametric model of a linear system is given with particular attention given to the question of estimating the error associated with the time variation of the parameters. The derivation of a parametric model for LTI systems is standard and may be found in any text on adaptive control, for example, Reference [21]. The method used here to estimate the error is based on the development in Middleton *et al.* [7] (cf. also References [22, 23, 25]).

Consider a single-input single-output (SISO) linear time-varying (LTV) system given by the solution of the ODE:

$$A(s)y = B(s)u \tag{34}$$

where $s = \mathrm{d}/\mathrm{d}t$ denotes the indeterminate of the system polynomials and

$$A(s) = s^n + a_{n-1}(t)s^{n-1} + \cdots + a_0(t)$$

$$B(s) = b_m(t)s^m + b_{m-1}(t)s^{m-1} + \cdots + b_0(t)$$

with $n > m$ and $\{a_{n-1}(t), \ldots, a_0(t), b_m(t), \ldots, b_0(t)\}$ time-varying unknown parameters. To obtain the classical parametric model for a linear system one begins by filtering the data y and u using an nth-order Hurwitz filter H,

$$H(s)y_f = y \quad H(s)u_f = u$$

where $H(s) = s^n + f_{n-1}s^{n-1} + \cdots + f_0$ for known constants $\{f_{n-1}, \ldots, f_0\}$. The parametrized model is based on applying the system (34) directly to the filtered signals y_f and u_f

$$A(s)y_f - B(s)u_f =: \eta_t \tag{35}$$

where η_t is a time-varying signal that measures the error between the actual behaviour of the left-hand side of this equation and that of the true system behaviour $A(s)y - B(s)u = 0$ (Mareels *et al.* [26] is a good introduction to behavioural analysis of systems for adaptive control). Define a regressor signal

$$\phi_t = (y_f^{(n-1)}, \ldots, y_f, u_f^{(n-1)}, \ldots, u_f) \in \mathbb{R}^{2n}$$

which is composed entirely of known signals drawn from the internal states of the data filters and a parameter vector

$$\theta_t = (f_{n-1} - a_{n-1}(t), \ \ldots, f_0 - a_0(t), 0, \ \ldots, 0, b_m(t), \ \ldots, b_0(t)) \in \mathbb{R}^{2n}$$

which is a time-varying unknown signal. The parametric model is now (cf. (1))

$$y_t := \theta_t^{\mathrm{T}} \phi_t + \eta_t$$

Properly understanding and estimating the error signal η_t is an important step in analysing LTV systems in a parametric framework. If the time-varying parameters θ_t vary rapidly and without any structure then the 'linear' form of (34) does not yield much insight into the behaviour of the signals y, u. Consequently, the parametric error representation (1) (based on LTI insight) should not be expected to carry much validity and the error η_t is likely to be large relative to the signals y_t and ϕ_t. In this case, a basic over-bound on the error η_t can be obtained from (35)

$$|\eta| \leqslant \kappa_t^{\mathrm{s}} |\phi_t|$$

where the superscript s denotes the fact that the bound $\kappa_t^{\mathrm{s}} > 0$ is derived from a static analysis of the error equation and

$$\kappa_t^{\mathrm{s}} \geqslant (1 + a_{n-1}^2(t) + \cdots + a_0^2(t) + b_m^2(t) + \cdots + b_0^2(t))^{1/2} \tag{36}$$

is an on-line *a priori* estimate of the norm of the vector of parameters $\{1, a_{n-1}(t), \ldots, a_0(t), b_m(t), \ldots, b_0(t)\}$. In such a situation, working with the parametric model (1), the best that can be hoped is to identify parameters $\hat{\theta}_t$ which adequately predict y_t but may bear little or no resemblance to the parameters in θ_t. Since the Lyapunov functions used in Sections (3) and (4) both contain a term penalizing $|\theta_t - \hat{\theta}_t|^2$ then it is reasonable that the bound on η_t (cf. Assumption II, Section 2) is a fundamental limit to the performance of the proposed identification algorithm.

If the parameters $\{a_{n-1}(t), \ldots, a_0(t), b_m(t), \ldots, b_0(t)\}$ are only slowly varying in time-relative to the (frozen-time) dynamics of the linear-system structure of (34) and those of the data filter $H(s)$ then it is expected that the parametric model (1) has validity. To exploit this structure for bounding the error η_t one applies the Hurwitz filter to (35) to obtain

$$\begin{aligned} H(s)\eta_t &= H(s)(A(s)y_f - B(s)u_f) \\ &= [H(s)A(s) - A(s)H(s)]\,y_f - [H(s)B(s) - B(s)H(s)]\,u_f \\ &\quad + A(s)(H(s)y_f) - B(s)(H(s)u_f) \\ &= [H(s)A(s) - A(s)H(s)]y_f - [H(s)B(s) - B(s)H(s)]u_f \end{aligned}$$

If $A(s)$ and $B(s)$ were linear time-invariant (LTI) polynomials then the 'swapping' terms $[H(s)A(s) - A(s)H(s)]$ and $[H(s)B(s) - B(s)H(s)]$ are zero and it follows that $H(s)\eta_t = 0$. Thus, for LTI systems η_t is an exponentially decaying signal associated with the initial condition of the data filter. In general, the swapping terms are non-zero due to the time-derivatives of time-varying parameters in the polynomials $A(s)$ and $B(s)$. Explicitly, one has [7, Equation (2.14)] (cf. also Lemma A.4 [22, p. 780])

$$H(s)]\eta_t = -\sum_{i=1}^{n}(-1)^{(i)}H_i(s)(A^{(i)}(s)y_f - B^{(i)}(s)u_f) \tag{37}$$

where

$$H_i(s) = \frac{\partial^i}{\partial s^i}H(s)$$

$$A^{(i)}(s) = a_{n-1}^{(i)}(t)s^{n-1} + \cdots + a_0^{(i)}(t)$$

$$B^{(i)}(s) = b_m^{(i)}(t)s^m + \cdots + b_0^{(i)}(t)$$

Equation (37) may be rewritten as

$$H(s)\eta_t = [H_1(s) \quad -H_2(s) \quad H_3(s) \quad \cdots \quad (-1)^{n-1}H_n(s)]\begin{pmatrix} \phi_t^{\mathrm{T}}\dot{\theta}_t \\ \phi_t^{\mathrm{T}}\theta_t^{(2)} \\ \vdots \\ \phi_t^{\mathrm{T}}\theta_t^{(n)} \end{pmatrix}$$

This equation is a multi-input single-output filter for the error signal η_t in terms of a set of inputs $[\phi_t^{\mathrm{T}}\dot{\theta}_t, \phi_t^{\mathrm{T}}\theta_t^{(2)}, \ldots, \phi_t^{\mathrm{T}}\theta_t^{(n)}\}$ which depend on the regressor vector ϕ_t, time derivatives of the unknown parameter vector θ_t, *and not on the actual values of the parameters themselves*. The error filter is stable (as a consequence of the stability of H) and proper (since the polynomials H_i are of degree $n - i$ for $i = 1, \ldots, n$). At time t, let

$$D_t \geqslant \max\{|\dot{\theta}_t|, |\theta_t^{(2)}|, \ldots, |\theta_t^{(n)}|\}$$

be an *a priori* on-line estimate of the 'rate' of time variation of the parameters. If the parameter vector θ_t is slowly time varying then it is expected that D_t will be small relative to the other signals in the system. Based on the above development one obtains a second bound on the size of the error η [7]

$$|\eta_t| \leqslant \kappa_t^{\mathrm{d}} := \left[\chi_0 \exp(-t\lambda_f) + \chi_1 \sup_{0 \leqslant r \leqslant t} (D_\tau |\phi_\tau| \mathrm{e}^{-\lambda_f(t-\tau)}) \right] \tag{38}$$

where χ_0 is the norm of the initial condition in the error filter, $\lambda_f > 0$ is the stability margin of the Hurwitz data filter $H(s)$ and χ_1 is a constant associated with the realization of the error filter (37). The superscript d on the bound κ_t^{d} denotes the fact that this bound is derived from a dynamic analysis of the error equation.

Equations (36) and (38) lead to an estimate of the constant C_t

$$C_t := \max \{\kappa_t^{\mathrm{s}} |\phi_t|, \kappa_t^{\mathrm{d}}\} \tag{39}$$

that may be used in Assumption II, Section 2. The bound C_t depends on *a priori* estimates κ_t^{s} and D_t on the time-variation of the parameter vector θ_t. The dynamic evolution of this bound, in comparison to the direct bound $|\dot{\theta}| \leqslant B_t$ (cf. Assumption I, Section 2), is important in practice since a short sharp transient in θ_t will generate errors that die off slowly in the data filter.

REFERENCES

1. Åström KJ, Eykhoff P. System identification a survey. *Automatica* 1971; **20**(1):21–38.
2. Tsypkin YZ. *Adaptation and Learning in Automatic Systems*. Academic Press: New York, USA, 1971.
3. Narendra KS, Kudva P. Stable adaptive schemes for system identification and control: Part I. *IEEE Transactions on Systems, Man and Cybernetics* 1974; **4**:541–551.
4. Narendra KS, Kudva P. Stable adaptive schemes for system identification and control: Part II. *IEEE Transactions on Systems, Man and Cybernetics* 1974; **4**:552-560.
5. Morse AS. Global stability of parameter adaptive control systems. *IEEE Transactions on Automatic Control* 1980; **25**:433–439.
6. Goodwin GC, Ramadge PJ, Caines PE. Discrete-time multivariable adaptive control. *IEEE Transactions on Automatic Control* 1980; **25**:456–461.
7. Middleton RH, Goodwin GC. Adaptive control of time-varying linear systems. *IEEE Transactions on Automatic Control* 1998; **33**(2):150–155.
8. Tsakalis K, Ioannou P. A new model reference controller structure. *IEEE Transactions on Automatic Control* 1989; **34**(10):1038–1046.
9. Giri F, M'Saad M, Dugard L, Dion JM. Pole placement direct adaptive control for time-varying ill-modeled plants. *Proceedings of the Conference on Decision and Control*, Los Angeles, CA, USA, 1987.
10. Anderson BDO, Bitmead RR, Johnson CR, Kokotovic PV, Kosut R, Mareels IMY, Praly L, Riedle B. *Stability of Adaptive Systems*: *Passivity and Averaging Analysis.* MIT Press: Cambridge MA, USA, 1986.
11. de Mathelin M, Lozano R. Robust adaptive identification of slowly time-varying parameters with bounded disturbances. *Proceedings of the European Control Conference*, *ECC97*, Bruxelles, Belgium, 1997.
12. Kreisselmeier G. Adaptive observers with exponential rate of convergence. *IEEE Transactions of Automatic Control* 1977; **22**(1):2–pt8.
13. Fortescue T, Kershenbaum L, Ydstie B. Implementation of self-tuning regulators with variable forgetting factors. *Automatica* 1981; **17**(6):831–835.
14. Cordero A, Mayne D. Deterministic convergence of a self-tuning regulator with variable forgetting factor. *IEE Proceedings* 1981; **128-D**(1):19–23.
15. Landau ID, Lozano R, M'Saad M. *Adaptive Control*. *Communications and Control Engineering*. Springer: London, Great Britain, 1998.

16. Lozano R, Dimogianopoulos D, Mahony R. Identification of linear time-varying systems using a modified least squares algorithm. *Proceedings of the IFAC Congress on Adaptive Control and Signal Processing*, Edinburgh, United Kingdom, 1998. Session 2, Friday afternoon.
17. Bittanti S, Bolzern P, Campi M. Convergence and exponential convergence of identification algorithms with directional forgetting factor. *Automatica* 1990; **26**(5):929–932.
18. Parkum JE. Recursive forgetting algorithms. *International Journal of Control* 1992; **55**(1):109–128.
19. Kulhavy R, Zarrop MB. On a general concept of forgetting. *International Journal of Control* 1993; **58**(4):905–924.
20. Cao L, Schwartz H. A novel recursive algorithm for directional forgetting. *Proceedings of the American Control Conference*, 1999.
21. Narendra KS, Annaswamy AM. *Stable Adaptive Systems*. Prentice-Hall: Englewood Cliffs, NJ, U.S.A., 1989.
22. Ioannou PA, Sun J. *Robust Adaptive Control*. Prentice-Hall PTR: Englewood Cliffs, NJ, USA, 1996.
23. Tsakalis K, Ioannou P. *Linear Time-Varying Systems*: *Control and Adaptation*. Prentice-Hall: Englewood Cliff, NJ, USA, 1993.
24. Goodwin GC, Sin KC. *Adaptive Filtering Prediction and Control*. Prentice-Hall: Englewood Cliffs, NJ, USA, 1984.
25. Dimogianopoulos D. Commande Adaptative des Systèmes Linéaires à Paramètres Constants ou Lentement Variant dans le Temps. *Ph.D. Thesis*, Heudiasyc, UMR 6599, Centre de Recherche de Royallieu, Université de Technologie de Compiègne, BP 20529, 60205 Compiegne Cedex, France, 1999 (in French).
26. Mareels I, Polderman J. *Adaptive Systems*: *An Introduction*, Systems and Control: Foundations and Applications. Birkhäuser: Boston MA, USA, 1996.

