Proceedings of the 11th
KACC, October 1996

NONLINEAR SELF-TUNING CONTROL INCORPORATING CAUTIOUS ESTIMATION

°D J G James and K J Burnham

Control Theory and Applications Centre, Coventry University, Priory Street, Coventry, CV1 5FB, UK
Tel: +44 1203 838564; Fax: +44 1203 838585; Email: g.james@coventry.ac.uk and ctac@coventry.ac.uk

Abstract The paper highlights the need for cautious least squares estimation when dealing with industrial applications of bilinear self-tuning control and indicates in qualitative terms the benefits of the approach over linear self-tuning control schemes. The cautious least squares algorithm is described and the use of cautious self-tuning in the context of both commissioning and implementation discussed.

Keywords Bilinear Systems, Self-tuning control, Cautious estimation

1. INTRODUCTION

Controllers that are based on conventional three term proportional, integral, derivative (PID) approaches are often found to be adequate for many industrial systems. However, with increasing demands for improved plant efficiency and overall system performance such schemes are frequently required to be continually re-tuned for each operating point. This is a time consuming task which neither guarantees optimality or repeatability. Consequently, alternative, schemes, termed self-tuning controllers (STC), have emerged which offer both facilities for automatic tuning of controller gains and the flexibility to realise a wide variety of control law objectives, from simple PID to multivariable generalised predictive control. In principle STC is a conceptually simple and straightforward approach which may be characterised by two coupled sub-algorithms; one for on-line parameter estimation and the other for control implementation. Fundamental to all forms of STC is the need to update a mathematical model representation of the plant and the integrity of this common element is a major contributory factor in determining the overall effectiveness of a particular STC scheme. A good summary of the developments in self-tuning control is given in the texts [12,19].

When attempting to apply STC to non-linear industrial systems it is common to find, in all but the simplest of cases, that standard techniques based on linear model structures produce performances that are inferior to conventional PID schemes. In an attempt to improve overall performance non-linear model structures have been considered, with particular attention on bilinear model structures; such models being representative of a wide range of industrial processes and plant [2,15].

2. CAUTIOUS LEAST SQUARES APPROACH

A single-input single-output (SISO) discrete time bilinear model, which has been used for applications such as high temperature furnaces, fermentation processes, waste water treatment and non-linear valve characteristics, is adopted as a basis for introducing the cautious least squares methodology. It takes the form of the non-linear auto-regressive moving average extended (NARMAX) model representation

$$A(q^{-1})y(t) = q^{-k}B(q^{-1})u(t) + \sum_{i=0}^{n_h}\sum_{j=1}^{m} \bar{y}(t-i-k)u(t-i-j-k+1)\eta_{ij} + v(t) \qquad (1)$$

where the polynomials $A(q^{-1})$ and $B(q^{-1})$ are defined by the general polynomial

$$\mathcal{L}(q^{-1}) = \ell_0 + \ell_1 q^{-1} + \ell_2 q^{-2} + \ldots + \ell_{n_\ell} q^{-n_\ell}$$

with $a_0=1$, $b_0\neq 0$ and $v(t)=e(t)+d(t)$; $u(t)$, $y(t)$, $e(t)$, $\bar{y}(t)$, and $d(t)$ being the input, output, white noise output disturbance, noise free output and local offset sequences respectively, $k\geq 1$ represents the system dead time expressed as an integer multiple of the sampling interval and q^{-1} is the backward shift operator defined as $q^{-i}y(t) \triangleq y(t-i)$.

In bilinear STC algorithms the coefficients of both the linear and bilinear terms are recursively estimated in order to update the coefficients in the NARMAX model (1) which is used in the controller update stage. In order to estimate the η_{ij} it is necessary to generate an estimate of the noise free output $\bar{y}(t)$. This is achieved using a steady-state Kalman filter scheme via use of the equivalent state space representation of (1)

$$\mathbf{x}(t+1) = \mathbf{P}\mathbf{x}(t) + \mathbf{Q}u(t) + \mathbf{R}v(t) + \sum_{i=1}^{m} u(t-i+1)\mathbf{N}_i\,\mathbf{x}(t) \qquad (2a)$$

$$y(t) = \mathbf{H}\mathbf{x}(t) + v(t) \qquad (2b)$$

where $\mathbf{x}(t) \in \mathbf{R}^N$, $N = n_j + k$ with $n_j = \max(n_a,n_b)$ and the matrices $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{H}$ and $\mathbf{N}_i$, $i = 1,2,\ldots,m$, are initially given in the implicit delay observable canonical form [4].

Rearranging the state equation 2(a), absorbing the bilinear terms, leads to a time-step quasi-linear state equation which is used to generate the estimated state vector $\hat{\mathbf{x}}(t)$. Eliminating the noise term, by substituting the output equation (2b) into the quasi-linear state equation, leads to the iterated steady-state observer (ISO)

$$\hat{\mathbf{x}}(t) = \left[\mathbf{I} - q^{-1}\ \mathbf{P}^{*}(\tilde{\mathbf{u}}(t)) - \mathbf{RH}\right]^{-1} \times \left[\ \mathbf{Q}^{*}(\hat{\mathbf{x}}(t-1))u(t-1) + \mathbf{R}y(t-1)\ \right] \qquad (3)$$

which is essentially an extension of the steady-state Kalman filter proposed in [18], where

$$\tilde{u}(t) = [\ u(t-1)\ u(t-2)\ ....\ u(t-m+1)]^T$$

and

$$P^*(\tilde{u}(t)) = \begin{cases} P\ ,\ m = 1 \\ P + \sum_{i=2}^{m} u(t-i+1)\ N_i\ ,\ m \geq 2 \end{cases}$$

with $\quad Q^*(x(t)) = Q + N_i\ x(t).$

The estimate of the noise free output $\bar{y}(t)$ is then given by $\hat{x}_n(t)$ the nth element of the estimated state vector $\hat{x}(t)$.

Rearranging (1) leads to

$$y(t) = z^T(t)\Theta(t) + \zeta(t) \qquad (4)$$

where $z \in \mathbb{R}^p$ is the observation vector

$$z^T(t) = [\ -y(t-1)\ ...\ -y(t-n_a);\ \ u(t-k)\ ...\ u(t-k-n_b);$$

$$\hat{x}_n(t-k)u(t-k-m+1)\ ...\ \hat{x}_n(t-k-n_b)u(t-k-n_b-m+1);\ \ 1]$$

and $\Theta \in \mathbb{R}^p$ is the parameter vector

$$\Theta^T(t) = \ [a_1...a_{n_a}\ ;\ b_0...b_{n_b}\ ;\ \eta_{01}...\eta_{n_b 1}\ ;...;\ \eta_{0m}...\eta_{n_b m}\ ;\ \delta]$$

with δ being an estimate of the offset term, and $\zeta(t)$ is a sequence of residuals (comprising errors due to both measurement and estimation) which becomes equal to the noise sequence e(t) upon convergence. The standard recursive least squares (RLS) algorithm, based on minimising the cost function

$$J_1\ (\Theta) = (y - Z\ \Theta)^T\ \Lambda\ (y - Z\ \Theta) \qquad (5)$$

for generating the parameter estimates $\hat{\Theta}(t)$ of the bilinear system is identical to RLS algorithm for a linear system, with the observation vector being appropriately extended in order to accommodate the additional multiplicative terms. In (5)

$$y(t) = [y(1)\ y(2)\ ...\ y(t)]^T,\ \ Z(t) = [z(1),\ z(2),\ ...\ ,\ z(t)]^T$$

and $\Lambda = \text{diag}\ (\lambda^t\ \lambda^{t-1}\ ...\ \lambda^2,\ \lambda)$, where $\lambda \leq 1$ is a specified forgetting factor allowing the user to place greater emphasis on more recent observations. The value of the forgetting factor can also be varied or regulated with time and this will give rise to increased adaptivity of the STC scheme. The resulting algorithm is given by

$$\hat{\Theta}(t) = \hat{\Theta}(t-1) + \varphi(t)\ [\ y(t) - z^T(t)\hat{\Theta}(t-1)\ ] \qquad (6a)$$

where $\quad \varphi(t) = \Phi(t-1)\ z(t)\ [\ 1 + z^T(t)\ \Phi(t-1)\ z(t)\ ]^{-1} \qquad (6b)$

$$\Phi(t) = \Big[[\ I - \varphi(t)z^T(t)\ ]\ \Phi(t-1)\ /\ \lambda(t)\Big] + \Psi \qquad (6c)$$

with I, $\varphi(t)$, $\Phi(t)$, $\lambda(t)$ and Ψ being the identity matrix, gain vector, error covariance matrix $[Z^T\ \Lambda\ Z]^{-1}$, variable forgetting factor and process noise covariance matrix respectively.

Since Φ stores information regarding parameter variation the introduction of Ψ in (6c) enables additional 'engineering knowledge' to be incorporated within the covariance matrix update equation. This being similar to the covariance update of a Kalman filter configured for parameter estimation, where the diagonal and off-diagonal elements of Ψ reflect the expected variation of individual parameter values and the expected cross-correlations between the parameter values respectively [16]. These elements are best determined from practical experience with the plant; however, a good 'rule of thumb' for tracking purposes suggest that the diagonal elements should range from 0.001 to 0.1 for the case of slow to rapid tracking respectively. The off-diagonal elements, which correspond to the sympathetic behaviour of the parameters, are set according to any knowledge of likely correlation, with values in the range ± 0.001 to ± 0.01 indicating weak to strong correlation. Such an approach does require experience with the plant and judicious tailoring and tuning of the elements to obtain the 'best' performance. Consequently, an Adaptive Kalman Filter (AKF) has been developed [14] which is able to automatically adjust the diagonal elements of Ψ on-line as the estimation algorithm evolves. The AKF is, however, unable to adjust the off-diagonal elements; these being preset manually as previously described. Another feature of the RLS algorithm (6) is that it provides its own error analysis with the variance of the error of the individual parameter estimates being indicated by the corresponding diagonal elements of the matrix Φ, which by definition are required to be positive quantities. The extension of the RLS algorithm to incorporate the bilinear coefficients η_{ij} can lead to greater correlation in the parameter estimates and giving rise potentially to biased estimates. Whilst this may not be a problem in implementing STC, it can lead to errors in model based predictions when used in conjunction with predictive STCs. In order to overcome this potential problem cautious least squares (CLS) has been introduced [3,5].

Essentially, CLS enables the designer to influence the estimation algorithm through a practical knowledge of the system. Caution is incorporated into the estimation algorithm via a simple modification of the standard RLS cost function J_1 in (5), giving the modified cost function

$$J_c(\Theta) = J_1 + J_2\ ;\ J_2 = (\Theta - \Theta_s)^T(\Theta - \Theta_s) \qquad (7)$$

in which the additional or 'cautious term' attempts to minimize the deviation of the estimated parameter vector $\hat{\Theta}(t)$, generated from the standard RLS algorithm (5), from some pre-specified 'safe set' of parameter values Θ_s; this safe set being identified off-line from input/output time series analysis or from a practical knowledge of the system. The parameter vector $\tilde{\Theta}(t)$ which minimises the modified cost function $J_c(\Theta)$ of (7) is called the cautious parameter vector. It is given by

$$\tilde{\Theta}(t) = [\Phi^{-1}(t) + I(t)]^{-1}\ [\Phi^{-1}\ \hat{\Theta}(t) + \Theta_s(t)] \qquad (8)$$

Since a recursive solution of (8) is not immediate, the approach is realised as a tandem operation of two separate procedures. First $\hat{\Theta}(t)$ is obtained using the RLS algorithm (6) to minimize J_1 only, followed by the determination of $\tilde{\Theta}$ from CLS by minimizing J_2 only, which involves a further p iterations of the algorithm. For clarity, it is convenient to assume that whilst time is 'frozen' between successive iterations of the RLS algorithm the discrete time index t is replaced by the index j, where j = 1,2, ... , p. The sequential CLS algorithm then takes the form

$$\Theta_j = \Theta_{j-1} + \varphi_j [e^T_j(\Theta_s - \Theta_{j-1})] \quad (9a)$$

$$\varphi_j = \Phi_{j-1} e_j [1 + e^T_j \Phi_{j-1} e_j]^{-1} \quad (9b)$$

$$\Phi_j = [I - \varphi_j e^T_j] \Phi_{j-1} \quad (9c)$$

where the e_j are the orthogonal unit vectors defined by

$$e^T_j = [\delta_{1j} \delta_{2j} \delta_{3j} \dots \delta_{pj}] \quad (9d)$$

in which δ_{ij} is the Kronecker delta function.

The additional CLS algorithm is initiated at each time step, following each iteration of RLS, with $\Theta_0 = \hat{\Theta}(t)$ and $\Phi_0 = \Phi(t)$, where $\hat{\Theta}(t)$ and $\Phi(t)$ are respectively the estimated parameter vector and covariance matrix generated from RLS. Following the p additional iterations of CLS, the cautious parameter estimate and covariance matrix become $\tilde{\Theta}(t) = \Theta_p$ and $\tilde{\Phi}(t) = \Phi_p$ respectively, with the cautious parameter vector $\tilde{\Theta}(t)$ being used for control purposes within the overall STC. At the next time step $\tilde{\Theta}(t)$ and $\tilde{\Phi}(t)$ are fed back to the RLS algorithm such that

$$\hat{\Theta}(t-1) = \tilde{\Theta}(t) \text{ and } \Phi(t-1) = \tilde{\Phi}(t) \text{ respectively.}$$

Note that the 'artificial data' in the form of the orthogonal unit vectors (9d) provides an artificial excitation signal within the estimation algorithm; thus to some extent alleviating the potential problems of covariance blow-up [1] during steady state periods of operation, without the need to disturb/perturb the plant.

3. PRACTICAL IMPLEMENTATION

Within the Research Centre at Coventry STC has been applied to a wide range of industrial processes and plant, including rotary hydraulic systems [10], engine testing dynamometers [13], gas engines [17] and high temperature furnace applications [7,9].

Experiences with the implementation of STC has revealed that linear STC often struggles to match the performance of a well tuned industry standard PID scheme. One of the major problems of a conventional STC lies in the inevitable plant/model mismatch which occurs when use is made of linear model structures. Indeed it is recognition of this fundamental shortfall of the linear approach that prompted the study into bilinear forms of model based STC. An unfortunate consequence of a linear STC is that the parameters of the assumed linear model necessarily vary widely as the model attempts to replicate the plant. By adopting a bilinear model structure, it is found that variations in the model parameters are significantly reduced, thereby effectively removing the need for enhanced estimation techniques, such as use of forgetting factors, etc. A distinct advantage of the reduced variation in the parameters is that the estimation algorithm can be constrained to limit the amount of parameter variation allowed. This leads to an overall robust STC scheme in which model uncertainties are accommodated within the allowable parameter variation.

Since it exhibits most of the nonlinearities that are commonly experienced with industrial plant, the high temperature furnace is considered here to describe some of the problems that may be encountered when applying STC in practice. For example, such a plant is predominantly first order having input dependent steady state gain and dynamic response characteristics. In particular, with increased input, in terms of gas valve position, the system gain decreases and the speed of response become faster; such a characteristic being typical of a bilinear system. Since the nominal time constant and steady state gain are known, it is possible to obtain a nominal safe set for the model parameters Θ_s. When use is made of a linear model structure

$$y(t) = -a_1 y(t-1) + b_0 u(t-1)$$

it is found that the model parameters can vary significantly. However, when use is made of a bilinear model structure

$$y(t) = -a_1 y(t-1) + b_0 u(t-1) + \eta_1 u(t-1) y(t-1)$$

then it is found that the model parameter variation is considerably reduced.

The introduction of CLS allows constraints to be placed on the estimated model parameter values, thus leading to an overall practical realisable self-tuning algorithm. Whilst in principle CLS may be applied sequentially at each time step, this may lead to superfluous computation and alternative methods of implementation have been devised. One method, which exploits the diagonal form of the covariance matrix upon reset is to combine CLS with covariance matrix reset. This has the advantage of cautioning all parameter estimates simultaneously. However, it may be undesirable to make use of resetting techniques under certain situations, and CLS may be applied cyclically. In this way each individual element of the parameter vector is cautioned every (p−1) iterations of the RLS algorithm.

It has been found from experience [11], that it is better to apply caution as an event driven procedure rather than on a regular basis and in this respect it is useful to define a safe region about the safe set, such that

$$\Theta \in \left\{ \Theta_S \pm w \right\}$$

where $\left\{ \Theta_s \pm w \right\}$ denotes the safe region

$$\left\{ \Theta : \Theta_{s_i} - w_i < \Theta_i < \Theta_{s_i} + w_i \, , \; i = 1,2, \dots , p \right\}$$

and $w = [\Delta\Theta_1 \, \Delta\Theta_2 \dots \Delta\Theta_p]^T$, with $\Delta\Theta_i$ being the specified variation about the safe set parameter value Θ_{s_i}.

Following each successive iteration of RLS, the additional CLS algorithm is activated, by applying caution to individual elements should they reside outside this safe region; that is

if $\hat{\Theta} \in \left\{ \Theta_s \pm w \right\}$
then CLS is not applied

whilst if $\hat{\Theta} \notin \left\{ \Theta_s \pm w \right\}$
then CLS is applied to the violating elements.

Using such an approach, there is a minimum of additional computation and CLS is active only when necessary.

In addition to overcoming the problem of bias, that may arise in the presence of weakly correlated entries in the observation vector, other advantages of CLS include:

- prevention of the occurrence of covariance matrix blow up due to the estimator 'falling-to-sleep'; this being achieved by effectively keeping the algorithm alert without disturbing the plant
- commissioning of the self-tuning algorithm; this being achieved by constraining the estimated parameter values to belong to a region which defines a 'sensible' and stable system
- fault detection; this being achieved by monitoring the number of occurrences that a given parameter value requires to be cautioned. Successive calls for caution could well be indicative of an incipient fault condition.

4. CONCLUSIONS

The paper has reviewed some of the fundamental features of self-tuning control and, in particular, has discussed issues pertinent to its successful practical implementation. The deficiencies in practice of linear model based self-tuning control are considered. In an attempt to overcome these deficiencies the usage of bilinear model structures, within the self-tuning algorithm, have been reported and the advantages of incorporating cautious least squares within the estimation algorithm discussed. The resulting self-tuning controller, which has been applied successfully to a number of practical industrial applications, is believed to offer a realistic approach in practice and represents a significant step forward in promoting model based control of nonlinear systems.

REFERENCES

[1] K. J. Astrom, "Self-tuning Regulators, Design Principles and Applications," in K. S. Narendra and R. V. Monopoli, (Eds.), *Applications of Adaptive Control*, Academic Press, 1980.

[2] C. Bruni, G. Di-Pillo and G. Koch, "Bilinear Systems: An Appealing Class of 'Nearly Linear' Systems in Theory and Applications," *IEEE Trans Aut. Control*, Vol 19, pp. 334-348, 1974.

[3] K. J. Burnham and D. J. G. James, "Use of Cautious Estimation in the Self-tuning Control of Bilinear Systems," *Proc. RAI/IPAR*, Vol. 1, Toulouse, France, pp. 419-432, 1986.

[4] K. J. Burnham, D. J. G. James and D. N. Shields, "Self-tuning Control of Bilinear Systems," *J. Optimal Control Applications and Methods*, Vol. 8, pp. 147-157, 1987.

[5] K. J. Burnham and D. J. G. James, "Implementation of Self-tuning Controllers on Industrial Plant," *Proc. IMACS-IFAC Symp. Modelling and Control of Technological Systems*, Vol. 1, Lille, France, pp. 50-56, 1991.

[6] K. J. Burnham, K. J. Disdell, D. J. G. James and C. A. Smith, "Developments in Industrial Applications of Self-tuning Control," *Control Eng. Practice*, Vol. 3, No. 9, pp. 1265-1276, 1995.

[7] K. J. Disdell, "Bilinear Self-tuning Control of Industrial Systems," *PhD Thesis*, Coventry University, U.K., 1995.

[8] A. Dunoyer, L. Balmer, K. J. Burnham and D. J. G. James, "On the Characteristics of Practical Bilinear Model Structures," *J. Systems Science*, Vol. 22, No. 2, 1996.

[9] S. G. Goodhart, "Self-tuning Control of Industrial Systems," *PhD Thesis*, Coventry University, U.K., 1991.

[10] S. G. Goodhart, K. J. Burnham and D. J. G. James, "Self-tuning Control of a Non-linear Industrial Test Rig," *J. Systems Science*, Vol. 19, No. 1, pp. 5-19, 1993.

[11] S. G. Goodhart, K. J. Burnham and D. J. G. James, "Bilinear Self-tuning Control of a High Temperature Heat Treatment Plant," *IEE Proc. D Control Theory and Appl.*, Vol. 141, No. 1, pp. 12-18, 1994.

[12] C. J. Harris and S. A. Billings (Eds.), "Self-tuning and Adaptive Control," *IEE Control Engineering Series*, Vol. 15, (2nd Edn.), Peter Peregrinus, 1988.

[13] P. J. King, "Adaptive Control Applied to an Engine Test Cell," *PhD Thesis*, Coventry University, U.K., 1992.

[14] P. J. King, K. J. Burnham and D. J. G. James, "An Adaptive Kalman Filter for On-line Parameter Estimation," *J. Systems Science*, Vol. 20, No. 1, pp. 61-75, 1994.

[15] R. R. Mohler, "Bilinear Control Processes," *Academic Press*, New York, 1973.

[16] A. Randall, K. J. Burnham and D. J. G. James, "A Study of Kalman Filtering Techniques for Joint State and Parameter Estimation in Self-tuning Control," *J. Systems Science*, Vol. 17, No. 3, pp. 31-43, 1991.

[17] C. A. Smith, "Self-tuning Gas Engine Speed Governor," *PhD Thesis*, Coventry University, U.K., 1995.

[18] K. Warwick, "Self-tuning Regulators - A State-Space Approach," *Int. J. Control*, Vol. 33, No. 5, pp. 839-858, 1981.

[19] P. E. Wellstead and M. B. Zarrop, "Self-tuning Systems: Control and Signal Processing," *John Wiley and Sons*, Chichester, 1991.