IDENTIFICATION OF TIME VARYING SYSTEMS

Peter Young
Centre for Research on Environmental Systems and Statistics, Lancaster University, U.K., and Centre for Resource and Environmental Studies, Australian National University, Canberra, Australia

Keywords: Time variable parameters; recursive estimation; Kalman filter; fixed interval smoothing; transfer function models; maximum likelihood; instrumental variables; state-dependent parameters.

Contents

Summary

The article outlines the development of time variable parameter (TVP) estimation as an approach to modeling nonstationary dynamic systems. It then describes one of the latest methods for estimating time variable parameters in dynamic auto-regressive, exogenous variables (DARX) and dynamic transfer function (DTF) models. In the case of DARX models, the estimation methodology is based on standard recursive filtering and fixed interval smoothing algorithms.

In the DTF case, however, it is a combination of special recursive instrumental variable filtering and fixed interval smoothing algorithms. The practical utility of the various estimation algorithms is demonstrated by two examples. The first is a simulation example that illustrates well the advantages of DARX estimation when compared with earlier TVP estimation algorithms based on exponential forgetting. The second is a practical example of DTF estimation based on a well known set of gas furnace data.

1. Introduction

In 1960, Rudolf Kalman published his seminal paper on recursive state variable estimation and what has since come to be known, almost universally, as the Kalman filter (KF). This produced a revolution in estimation theory and practice, first in the control and systems literature, and subsequently in most other areas of engineering, science and the social sciences. Amongst many other things, it stimulated a growing interest in how recursive estimation methods could be exploited to model time varying or 'nonstationary' systems. The original motivation for this *Time Variable Parameter* (TVP) estimation research was the modeling of nonstationary dynamic processes and the use of such recursive algorithms in adaptive control system design. However, a later motivation has been its use as a tool in adaptive forecasting and signal processing.

Amongst the different approaches to TVP estimation, the following three are deserving of most attention:

1. The Extended (or Re-linearized) Kalman Filter (EKF)

   Here, the stochastic state space model of the dynamic system is extended to include simple stochastic models for the TVPs (e.g. the simple random walk model: see Section 3).The resulting model is nonlinear because the original system state variables are multiplied by the state variables arising from the adjoined TVPs. As a result, the state estimates of this nonlinear model are then updated by a KF-type algorithm with the equations linearized in some manner at each recursive update, based on the current recursive parameter estimates.

2. Shaping the Memory of the Algorithm

   Here, the predecessor of the KF, the recursive algorithm developed by K. F. Gauss in the nineteenth century for estimating the constant parameters in linear regression models, is modified to include a 'forgetting factor' or 'weighting kernel' that shapes the memory of the estimator and so allows for TVP estimation. This is much the most popular approach to TVP estimation but its performance is rather limited when compared with approach 3, below.

3. Modeling the Parameter Variations

   Here, the roles of the state equations and observation equations are reversed. The model of the system now appears in the observation equation and the state equations are used to model the TVPs appearing in this model, again using simple stochastic models such as the random walk. This is the most sophisticated and flexible approach and represents the current state-of-the-art in TVP estimation.

Until comparatively recently, the main emphasis in all three of these approaches has been the 'on-line' or 'real-time' estimation of the TVPs. As a result, most algorithms have been of the 'filtering' type, where the estimate $\hat{\mathbf{a}}_{k|k}$ of the TVP vector $\mathbf{a}_k$, at any sampling instant k, is a function of all the data up to and including this k^{th} instant. Surprisingly, given its ultimate power, the extension of these methods to the 'off-line' analysis situation was not considered very much at first, despite the fact that a mechanism of such 'smoothing' estimation was available in the form of various fixed interval smoothing (FIS) algorithms and publications on this subject. Here, the FIS estimate $\hat{\mathbf{a}}_{k|N}$ of $\mathbf{a}_k$ is based on *all* of the data available over a 'fixed interval' of *N* samples, usually the full sample length of the time series data. Later research placed this approach in an optimal context based on maximum likelihood estimation of the associated 'hyperparameters' (see Section 3. below).

2. Simple Limited Memory Algorithms

Before illustrating the value of a unified, statistical approach to TVP estimation based on modeling the parameter variations, it is instructive to take a brief look at the less sophisticated, deterministic algorithms based on explicitly restricting the memory of the recursive estimation algorithm. For simplicity, let us consider a single input, single output system (although the extension to multi-input systems is straightforward). In the case of a TVP or *Dynamic Transfer Function* (DTF) representation (The term 'dynamic' is used here for historical reasons, primarily because the parameters are defined as evolving in a stochastic, dynamic manner.), the model takes the following form:

$$y_k = \frac{B(z^{-1},k)}{A(z^{-1},k)} u_{k\text{-}\delta} + \xi_k \qquad\qquad t = 1,\ldots,N \qquad (1)$$

where u_k and y_k are, respectively, the input and output variables measured at the k^{th} sampling instant; z^{-1} is the backward shift operator, i.e,. $z^{-r} y_k = y_{k\text{-}r}$; and $A(z^{-1},k)$ and $B(z^{-1},k)$ are *time variable coefficient polynomials* in z^{-1} of the following form:

$$\begin{aligned} A(z^{-1},k) &= 1 + a_{1,k} z^{-1} + a_{2,k} z^{-2} + \ldots + a_{n,k} z^{-n} \\ B(z^{-1},k) &= b_{0,k} + b_{1,k} z^{-1} + b_{2,k} z^{-2} + \ldots + b_{m,k} z^{-m} \end{aligned} \qquad (2)$$

The term δ is a pure time delay, measured in sampling intervals, which is introduced to allow for any temporal delay that may occur between the incidence of a change in u_k and its first effect on y_k . Finally, ξ_k represents uncertainly in the relationship arising from a combination of measurement noise, the effects of other unmeasured inputs and modeling error. Normally, ξ_k is assumed to be independent of u_k and is modelled as an AutoRegressive (AR) or AutoRegressive-Moving Average (ARMA) stochastic process, although this restriction can be avoided by the use of instrumental variable methods, as discussed below.

In the more restricted case of the *Dynamic Auto-Regressive, Exogenous variables* (DARX) model ξ_k
is defined as

$$\xi_k = \frac{1}{A(z^{-1},k)} e_k ,$$

where e_k is assumed to be zero mean white noise. This has the advantage that equation (1) can be written in the following alternative vector equation or 'regression' form:

$$y_k = \mathbf{z}_k^T \mathbf{p}_k + e_k \tag{3}$$

where now,

$$\begin{aligned}
\mathbf{z}_k^T &= \left[-y_{k-1} - y_{k-2} \ldots - y_{k-n} \; u_{k-\delta} \; \ldots \; u_{k-\delta-m} \right] \\
\mathbf{p}_k &= \left[a_{1,k} \; a_{2,k} \ldots a_{n,k} \; b_{0,k} \; b_{1,k} \ldots b_{m,k} \right]^T \\
&= \left[p_{1,k} \; p_{2,k} \; \ldots \; p_{n+m+1,k} \right]^T
\end{aligned} \tag{4}$$

If we wish to limit the memory of the estimation algorithm, it is necessary to specify the nature of the memory process. Two main memory functions have been suggested: R*ectangular Weighting-into-the-Past* (RWP); and *Exponential-Weighting-into-the-Past* (EWP). The latter approach is the most popular and can be introduced into the least squares problem formulation by considering an EWP least squares cost function $\mathcal{J}_2^{EWP}$ of the form,

$$\mathcal{J}_2^{EWP} = \sum_{i=1}^{k} \{ y_i - \mathbf{z}_i^T \hat{\mathbf{p}}_i \}^2 \lambda(k - i) \tag{5}$$

where $0 < \lambda < 1.0$ is a constant related to the time constant T_e of the exponential weighting by the expression $\lambda(k) = \exp\,(-k\Delta t / T_e)$, and Δt is the sampling interval in time units appropriate to the application. Of course, with $\lambda = 1.0$, $\mathcal{J}_2^{EWP}$ becomes the usual, constant parameter, least squares cost function $\mathcal{J}_2$.

The recursive algorithm derived by the minimization of the EWP cost function (5) takes the following form:

$$\begin{aligned}
\hat{\mathbf{p}}_k &= \hat{\mathbf{p}}_{k-1} + \mathbf{P}_{k-1} \mathbf{z}_k \left[\lambda + \mathbf{z}_k^T \mathbf{P}_{k-1} \mathbf{z}_k \right]^{-1} \{ y_k - \mathbf{z}_k^T \hat{\mathbf{p}}_{k-1} \} \\
\mathbf{P}_k &= \tfrac{1}{\lambda} \{ \mathbf{P}_{k-1} - \mathbf{P}_{k-1} \mathbf{z}_k \left[\lambda + \mathbf{z}_k^T \mathbf{P}_{k-1} \mathbf{z}_k \right]^{-1} \mathbf{z}_k^T \mathbf{P}_{k-1} \}
\end{aligned} \tag{6}$$

This estimation algorithm is one form of the recursive EWP least squares algorithm, although other forms are possible. These can all be considered in terms of the EWP coefficient λ, or 'forgetting factor' as it is often called. Amongst the possibilities are *constant trace algorithms*, the use of *adaptive forgetting factors*, including *start-up forgetting factors;* and *Directional Forgetting* (DF). In the latter DF algorithm, the $\mathbf{P}_k$ matrix update takes the form:

$$\mathbf{P}_k = \{ \mathbf{P}_{k-1} - \mathbf{P}_{k-1} \mathbf{z}_k \left[r_{k-1}^{-1} + \mathbf{z}_k^T \mathbf{P}_{k-1} \mathbf{z}_k \right]^{-1} \mathbf{z}_k^T \mathbf{P}_{k-1} \} \tag{7a}$$

where a typical choice for r_k is,

$$r_k = \lambda^* + \frac{1-\lambda^*}{\mathbf{z}_{k+1}^T \mathbf{P}_k \mathbf{z}_{k+1}} \qquad (7b)$$

in which λ^* plays a similar role to λ in the EWP algorithm (6).

-
-
-

Bibliography

Beck, M. B., Stigter, J. D., and Lloyd Smith, D. (2002), Elasto-Plastic Deformation of the Structure. In *Environmental Foresight and Models*: *A Manifesto*, (ed. M. B. Beck), 323-350. Oxford: Elsevier. [A good recent discussion of the EKF approach].

Benveniste, A., Métivier, M and Priouret, P. (1990) *Adaptive Algorithms and Stochastic Approximation*, Berlin: Springer-Verlag. [A recent book on TVP estimation].

Bryson, A.E. and Ho, Y.C. (1969). *Applied Optimal Control*, San Francisco: Blaisdel. [a seminal book on estimation and control that discusses fixed interval smoothing estimation].

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Jnl. Royal Stat. Soc., Series B* **39,** 1-38. [description of the EM algorithm].

Gauss, K.F. (1826). *Theoria combinationis observationum erroribus minimis obnoxiae*, Parts 1, 2 and supplement, *Werke* **4**, 1-108. [the original derivation of the recursive least squares algorithm: English translation and explanation in Appendix 2 of Young, 1984].

Jakeman, A.J. and Young, P.C. (1984). Recursive filtering and the inversion of ill-posed causal problems. *Utilitas Mathematica* **35**, 351-376. [First use of the GRW model in TVP estimation].

Jazwinski, A.H. (1970). *Stochastic Processes and Filtering Theory*, New York: Acad. Press. [An excellent text on nonlinear filtering and linearized filtering algorithms].

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. *ASME Transactions Journal Basic Engineering* **82D**, 35-45. [The seminal work on recursive estimation and filter theory in the Twentieth Century].

Kitagawa, G. and Gersch, W. (1996). *Smoothness Priors Analysis of Time Series*, New York: Springer-Verlag. [A good book on TVP and numerically intensive estimation based on Monte Carlo simulation analysis].

Kulhavy, R. (1987). Restricted exponential forgetting in real-time identification, *Automatica* **23**, 345-346. [first paper on the DF algorithm].

Lee, R. C. K, (1964). *Optimal Identification, Estimation and control.* Cambridge, Mass: M.I.T. Press. [An important early book on TVP estimation].

Ljung, L. and Gunnarsson, S. (1990) Adaption and tracking in system identification – a survey, *Automatica.* **26,** 7-21. [a review of TVP estimation]

Niedźwiecki, M. (2000) *Identification of Time Varying Processes*, Chichester: J. Wiley: [The most recent book on TVP estimation].

Norton, J.P. (1986). *An Introduction to Identification*, New York: Academic Press. [A useful and easily readable text that includes discussion of TVP estimation and FIS estimation].

Schweppe, F. (1965). Evaluation of likelihood function for Gausssian signals. *IEEE Transactions on Information Theory* **11**, 61-70. [The original paper on prediction error decomposition].

Söderstrom, T. and Stoica, P. (1983) *Instrumental Variable Methods for System Identification,* New York: Springer-Verlag. [A good text on IV estimation].

Wellstead, P.E. and Zarrop, M.B. (1991). *Self-Tuning Systems: Control and Signal Processing,* Chichester: Wiley, 140-159. [A readable book that includes discussion of TVP estimation: the simulation example is based on an example in this book].

Young, P.C. (1969). Applying parameter estimation to dynamic systems: Part I, Theory, *Control Engineering,* **16**, 10, 119-125; Part II Applications, **16**, 11, 118-124. [An early two-part paper by the author on TVP estimation];

Young, P.C.(1984). *Recursive Estimation and Time-Series Analysis,* Berlin: Springer-Verlag. [A comprehensive tutorial treatment of recursive and TVP estimation, including RWP, EWP and DARX recursive estimation algorithms that exploit IV and FIS methods of estimation].

Young, P.C.(1999). Nonstationary time series analysis and forecasting. *Progress in Environmental Science* **1**, 3-48. [The author's most recent paper on TVP estimation].

Young, P.C.and Pedregal, D.J. (1999). Recursive and en-bloc approaches to signal extraction. *Journal of Applied Statistics* **26**, 103-128. [A recent paper that shows the interrelationships between recursive FIS, Wiener-Kolmogorov-Whittle and 'regularization' methods of optimal smoothing].

Young, P.C., Pedregal, D.J. and Tych, W. (1999). Dynamic harmonic regression, *Journal of Forecasting* **18**, 369-394. [A recent paper that illustrates how TVP estimation can be used for adaptive forecasting].

Young, P.C., McKenna, P and Bruun, J. (2001). Identification of nonlinear stochastic systems by state dependent parameter estimation. *International Journal of Control*, **74**, 1837-1857. [The most recent paper on SDP estimation].

Biographical Sketch

Peter Young is Professor of Environmental Systems and Director of the Centre for Research on Environmental Systems and Statistics at Lancaster University, U.K. He is also Adjunct Professor of Environmental Systems at the Australian National University, Canberra. He obtained B.Tech. and M.Sc. degrees at Loughborough University before moving to Cambridge University, where he obtained his Doctoral degree in 1970. Following two years as a civilian scientist, working for the U.S. Navy in California, he was appointed University Lecturer in Engineering and a Fellow of Clare Hall, Cambridge University, in 1970. As a result of his novel research on environmental modeling, he was then invited to become Professorial Fellow and Head of the Systems Group in the new Centre for Resource and Environmental Studies at the Australian National University in 1975. Finally, he moved back to U.K. in 1981, where he was Head of the Department of Environmental Science for seven years, before assuming his current position. He is well known for his work on recursive estimation and his most recent research has been concerned with data-based modeling, forecasting, signal processing and control for nonstationary and nonlinear stochastic systems. The applications of this methodological research are wide ranging, from the environment, through ecology and engineering to business and macro-economics.