RECURSIVE ALGORITHMS

Han-Fu Chen
Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, P. R. China.

Keywords: least squares, extended least squares, forgetting factor, instrumental variable, stochastic approximation, stochastic gradient, convergence, consistency, rate of convergence, diminishing excitation, least mean squares.

Contents

Summary

Various recursive identification algorithms are presented for the discrete-time stochastic time-invariant systems. The convergence issue is addressed: Sufficient conditions for strong consistency of various recursive estimates are given; Convergence rates are also provided for LS and ELS estimates. The diminishing excitation method is introduced, by which the strong consistency of parameter estimates can be achieved for a feedback control system.

Tracking a time-varying parameter, i.e., identifying time-varying coefficients in discrete-time, stochastic linear systems is an important problem not only for system and control but also for signal processing. KF, LMS and RLS are the most commonly used recursive algorithms in the area. The basic properties are described for these algorithms. Finally, besides the historical issue and continuous-time systems briefly addressed in "Concluding Remarks", some open problems are pointed out as well.

1. Introduction

Many real-world phenomena can be approximately described by linear models which linearly relate system outputs with inputs by differential or difference equations. The approximation errors in some cases may be modeled as random variables. As a result, the real system is modeled by a linear stochastic difference (or differential) equation with unknown parameters to be determined. In this case, the topic of system identification is to estimate or define the unknown parameters contained in the model on the basis of the input-output data derived from the observation on the real system under consideration. If a batch of input-output data is collected from the system and used to estimate the model parameters, then such a procedure is called o -line identification. If the estimate for model parameters is updated by a recursive algorithm at each time when some new data becomes available, then such a procedure is called recursive identification.

The unknown parameters in the model may be the model-equation coefficients, orders, and time delay. Estimates for orders and time delay for a linear system are normally derived by minimizing some information criteria such as AIC, BIC, ΦIC and CIC etc, but they are hardly to be on-line and recursive. Therefore, in all recursive identification methods the system orders and time delay or their upper bounds are usually assumed to be known, and the main effort of system identification is devoted to estimate the system coefficients by which the system input and output are related.

To be precise, let us consider the discrete-time case. Let $u_k \in \mathbb{R}^l, y_k \in \mathbb{R}^m$ be system input and output, respectively. The real system may be modeled by

$$\begin{aligned} y_{k+1} = {} & A_1 y_k + A_2 y_{k-1} + \cdots + A_p y_{k-p+1} + B_1 u_k + B_2 u_{k-1} \\ & + \cdots + B_{qu_k-q+1} + \varepsilon_{k+1} \end{aligned}, \tag{1}$$

where $A_1, A_2, ..., A_p, B_1, B_2, ..., B_q$ are unknown coefficient matrices, ε_{k+1} characterizes the approximation error, and it may be modeled as an m-dimensional random vector. In this case, (p, q) are the system orders, or upper bounds for the true orders. Since the latest input that effects on y_{k+1} is u_k, the time delay for system (1) is one.

Notice that A_i and B_j, are $m\times m$ and $m\times l$ matrices, respectively, $i = 1, ..., p$, $j = 1, ..., q$. Let us combine all unknown coefficients to a big matrix θ:

$$\theta^T = [A_1, A_2, ..., A_p, B_1, B_2, ..., B_q] \tag{2}$$

and denote the input-output data by

$$\phi_k^T = [y_k^T, y_{k-1}^T, ..., y_{k-p+1}^T u_k^T u_{k-1}^T, ..., u_{k-q+1}^T], \tag{3}$$

where X^T denotes the transpose of matrix X (or vector X).

By Eqs. (2) and (3), the model (1) can be rewritten as

$$y_{k+1} = \theta^T \phi_k + \varepsilon_{k+1}. \tag{4}$$

Recursive algorithms are used to estimate the unknown coefficient matrix θ by using the input-output data, and the estimate is on-line updated while receiving the new input-output observation.

When θ is time-independent, the system (4) or (1) is called time invariant system. But θ may vary with time *k*, i.e., $\theta = \theta_k$, then the system

$$y_{k+1} = \theta_k^T \phi_k + \varepsilon_{k+1} \tag{5}$$

is called time-varying system.

The noise ϵ_{k+1} in Eq. (1) can also be modeled, for example, ϵ_{k+1} may be driven by a sequence of mutually independent random vectors $\{ w_k \}$ (or, more general, a martingale difference sequence), i.e.,

$$\varepsilon_{k+1} = w_{k+1} + C_1 w_k + \cdots + C_r w_{k-r+1}. \tag{6}$$

This kind of noise $\{\varepsilon_k\}$ is called MA (moving average) process, and system (1) with $\{ \varepsilon_{k+1} \}$ being defined by Eq. (6) is called ARMAX system, where AR (autoregression) implicates the part in Eq. (1) containing $\{ y_k \}$, and *X* the part with input terms. If $u_k \equiv 0$ and $\{\varepsilon_k\}$ is given by Eq. (6), then Eq. (1) is called ARMA process.

In what follows, the typical recursive estimation algorithms will be given for estimating the constant matrix θ in Eq. (4), and it will be shown how to make the estimates converge to the true coefficient matrices. The convergence rates will also be presented. For the time varying system (5), recursive estimation algorithms for tracking the time-varying coefficient θ_k will be addressed. In conclusion, some open problems will be pointed out.

2. Recursive Algorithm for Constant Coefficients.

2.8. Least Squares (LS)

For system (4) the parameter matrix θ is to be estimated based on $\{ y_{k+1}, \phi_k, k = 1, 2, ..., n \}$, n = 1, 2, ... A common and natural way is to estimate θ by minimizing the sum of squared errors, i.e., the estimate for θ is selected by minimizing the criterion

$$V_{n+1}(\hat{\theta}) = \sum_{k=0}^{n} \| y_{k+1} - \hat{\theta}^T \phi_k \|^2, n = 1, 2, ... \tag{7}$$

with respect to $\hat{\theta}$. The minimizer denoted by $\hat{\theta}_n$ is called the LS estimate.

At time *k*, the model approximation error ε_{k+1} is unknown. Therefore, $\theta^T\phi_k$ is the best-predicted value for the output at time *k* + 1, and hence $y_{k+1} - \hat{\theta}^T\phi_k$ can be viewed as a prediction error for the system output. To minimize $V_{n+1}(\hat{\theta})$ is to minimize the prediction errors. In other words, the estimate for θ is chosen such that the identified model best fits the observed data.

The criterion $V_{n+1}(\hat{\theta})$ is a quadratic form with respect to $\hat{\theta}$. Consequently, there exists a unique minimum for $V_{n+1}(\hat{\theta})$ which is achieved at

$$\hat{\theta}_{n+1} = \left(\sum_{i=0}^{n}\phi_i\theta_i^T\right)^{-1}\sum_{i=0}^{n}\phi_i y_{i+1}^T \tag{8}$$

provided the inverse exists. This is the well known least-squares (LS) estimate.
Note that the LS estimate $\hat{\theta}_{n+1}$ given by Eq. (8) depends on *n* + 1, the data size, and is in an o -line form. For the on-line identification, it is convenient to express $\{\hat{\theta}_k\}$ in the recursive way:

$$\hat{\theta}_{k+1} = \theta_k + a_k P_k\phi_k(y_{k+1}^T - \phi_k^T\theta_k) \tag{9}$$

$$P_{k+1} = P_k - a_k P_k\phi_k\phi_k^T P_k\,, \tag{10}$$

$$a_k = (1+\phi_k^T P_k\phi_k)^{-1}\,. \tag{11}$$

This is the recursive LS algorithm for identifying system (1) or (4) given initial values $\hat{\theta}_0$ and P_0.

The LS estimate may give a good estimate for θ only in the case where $\{\varepsilon_k\}$ is a sequence of uncorrelated random vectors. If $\{\varepsilon_k\}$ is a correlated sequence, the LS may not give a satisfactory estimate. This is why one has to apply the extended least squares estimate.

2.9. Extended Least Squares (ELS)

Now, assume ε_{k+1} is given by Eq. (6), i.e., the real system is modeled by an ARMAX process

$$\begin{aligned} y_{k+1} = {} & A_1 y_k + \cdots + A_p y_{k-p+1} + B_1 u_k + \\ & \cdots + B_q u_{k-q+1} + w_{k+1} + C_1 w_k + \cdots + C_r w_{k-r+1} \end{aligned} \tag{12}$$

and all matrix coefficients A_1, ..., A_p, B_1, ..., B_q, C_1, ..., C_r have to be identified. In this case, similar to Eq. (2) the unknown coefficients are still denoted by θ.

$$\theta^T = [A_1, A_2, ..., A_p, B_1, ..., B_q, C_1, ..., C_r]. \qquad (13)$$

Since $\{w_k\}$ is not observed, θ cannot be estimated by Eqs. (9)-(11), if ϕ_k is simply extended by adding $[w_k^T, ..., w_{k-r+1}^T]$. It is a natural way to replace e.g. w_k by its estimate $\hat{w}_k \triangleq y_k - \theta_k^T \phi_{k-1}$ and define

$$\phi_k^T = [y_k^T, ..., y_{k-p+1}^T, u_k^T, ..., u_{k-p+1}^T, \hat{w}_k^T, ..., \hat{w}_{k-r+1}^T] \qquad (14)$$

Then the recursive algorithm (9)-(11) with ϕ_k defined by Eq. (14) can still be used to estimate θ given by Eq. (13). This is the extended least-squares (ELS) estimate for θ given by Eq. (13).

-
-
-

Bibliography

Benveniste A., M. Métivier and P. Priouret. (1990). *Adaptive Algorithms and Stochastic Approximations*, 365pp. Berlin:Springer-Verlag. [This book presents stochastic approximation methods applied to tracking time-varying parameters. The analysis is based on the ODE method.]

Box E. P. and G. Jenkins.(1970). *Time Series Analysis; Forecasting and Control*, 553pp. San Francisco:Holder-Day.. [This book provides practical methods for identifying ARMA processes and demonstrates many application examples.]

Chen H. F. and Guo L. (1991) *Identification and Stochastic Adaptive Control*, 435pp. Boston:Birkhäuser. [This book presents detailed convergence analysis for LS, ELS and SG algorithms, gives their convergence rates, and shows how to use the diminishing excitation method to derive strongly consistent estimates for feedback control systems. The book also provides estimation methods for orders and time-delay with detailed analysis. Continuous-time systems are also addressed.]

Chen, H. F. (2002) *Stochastic Approximation and Its Applications*, 357pp. Boston: Kluwer Academic Publishers. [This book presents the stochastic approximation algorithm with expanding truncations and uses the trajector-subsequence (TS) method for analyzing the path-wise convergence of the algorithm. Applications to various problems arising from optimization, signal processing, and systems and control are demonstrated.]

Guo L. and Ljung L.(1995) Performance analysis of general tracking algorithms. *IEEE Trans. on Automatic Control*, **40**, (8) , 1388-1402. [This paper presents theoretical analysis for tracking time-varying parameters.]

Hannan E. J. and Deistler M. (1988). *The statistical Theory of Linear Systems*, 380pp. New York: Wiley. [This book concerns estimation of coefficients, orders and time-delay for ARMAX systems with input being independent of system noise, i.e., the feedback control is excluded from consideration.]

Kushner H. J. and Yin G. (1997). *Stochastic Approximation Algorithm and Applications*, New York: Springer-Verlag. [This book presents stochastic approximation methods and analyses the convergence of algorithms based on the ODE and weak convergence methods.]

Ljung L. and Söderström T. (1983). *Theory and Practice of Recursive Identification*, 529pp. Cambridge, Massachusetts: MIT Press. [This book provides various recursive algorithms for identifying system's parameters and many guides for their applications in practice.]

Solbrand G, Ahlen A., and. Ljung, L. Recursive. (1985). Methods for On-line Identification. *Int. J. of Control*, **41** (1), 177-191. [On off-line recursive computation of estimates.]

P. C. Young (1999). Nonstationary time series analysis and forecasting, *Progress in Environmental Science*, **1**, 3-48. [On off-line estimation and recursive fixed interval smoothing for time-varying systems.]

Biographical Sketch

Han-Fu Chen received the Diploma from the Department of Mathematics and Mechanics, Leningrad (St. Petersburg) University, Russia. He joined the Institute of Mathematics, Chinese Academy of Sciences in 1961. Since 1979 he has been with the Institute of Systems Science, which now belongs to the Academy of Mathematics and Systems Science, Chinese Academy of Sciences. He currently is a Professor of the Laboratory of Systems and Control of the Institute. He was the Director of the Institute of Systems Science for the period of 1995-1998. His research interests are mainly in stochastic systems including identification, adaptive control, recursive estimation, stochastic approximation and its applications to system, control, and signal processing. He has authored and coauthored more than 150 journal papers and 7 monographs. Han-Fu Chen is the Editor of the journals *Systems Science and Mathematical Sciences*, and *Control Theory and Applications* and is involved in the editorial boards of several international and domestic journals. He was elected to a Member of the Chinese Academy of Sciences in 1993 and to an IEEE Fellow in 1996. He serves as a member of the IFAC Council (2002-2005), served as the President of the Chinese Association of Automation (1993-2002) and on the Technical Board of IFAC (1996-2002). He was the IPC Chair of the 14th IFAC World Congress held in 1999.