Electronic Journal of Applied Statistical Analysis
EJASA, Electron. j. app. stat. anal. 1(2008), 24 – 32
ISSN 2070-5948, DOI 10.1285/i20705948v1n1p24
http://siba2.unile.it/ese/ejasa
http://faculty.yu.edu.jo/alnasser/EJASA.htm

BAYESIAN APPROACH FOR ROBUST PARAMETER TRACKING

Anton Abdulbasah Kamil
School of Distance Education, University Sains Malaysia, 11800 USM, Penang, Malaysia.
Email: anton@usm.my

Received 18 June 2008; accepted 5 August 2008
Available online 18 August 2008

Abstract. *In this paper we study the problem of tracking of time-varying parameters of a dynamical system. The problems also facing at the finite number of expected parameter changes and finite number of possible measurement model. We consider a stochastic model of parameter development with some form of obsolete information forgetting. It will be shown that it is possible to track rapidly time-varying plant parameters using extension of Bayesian viewpoint (continuous and discrete parameters) with requiring the prior information that can improve tracking for abrupt changes.*

Keywords. *Time-varying parameters, dynamical system, Bayes theorem, ARX model, tree pruning.*

1. Introduction

System identification deals with choosing mathematical models from a known model set to characterize the input–output behavior of an unknown system from finite noisy data. Noise, finite length of data, and time variation are some of the issues that limits the choice of a complex model set Saligrama R. Venkatesh and Munther A. Dahleh, 2001). As it is defined, system identification is a process for obtaining a mathematical model of a dynamic system based on a set of measured data from the system (Ljung, 1987) It involves taking a time history of input(s) and measured output(s) and fitting the parameters of a model structure such that its output error is minimized. The success of this technique is dependent on the

initial choice of the model structure and the amount and quality of data used to "train" the model.

Although there are many results available for system identification of time-varying parameters of a dynamical system. One of the basic difficulties is to track and to find the finite number of models of expected parameter changes and finite number of possible measurement models. The principal difficulty arises because of an interplay between noise and unmodeled dynamics. In the absence of noise the problem, in most situations, reduces to a function approximation problem for which well-known solutions already exist.

System identification is an important step to control an object when its model is unknown. When the system is parameterized, then the task of system identification is to estimate the unknown parameters contained in the system. One may first collect data with fixed sample size, and then derive the estimates by minimizing some performance index based on the collected data. The minimization may be carried out iteratively in order to use information contained in the data as completely as possible. This type of estimation is usually called as the block algorithm in contrast to the recursive algorithm. The block algorithm sometimes is not satisfactory, because having obtained the estimate based on the data with sample size *N*, one has to compute the estimate from beginning if some new data arrive. This makes computation rather time-consuming. Some solutions that can we proposed are; (i) for fast time-varying systems, it was suggested that the parameter tracking requires some form of obsolete information forgetting, (ii) forgetting induced by a stochastic model of parameter development, (iii) prior information formally described in terms of covariance matrices of parameter increments, (iv) information contained in future data, (v) several alternative hypotheses available-probability distribution over the tree of possible histories.

The problem of dimensionality also must be noted, especially for curse of dimensionality – tree pruning is a must, decision based approach and approximation based approach. In this study, the general results will applied to ARX (auto regressive with exogenous inputs) model parameters.

The following sections describe the plant to be tracked and the parameter trackers. The suggested models such as Bayesian approach, alternative parameter development and tree pruning are discussed in this paper. Section 5 concludes the paper.

2. Bayesian approach to identification

We start with the model structure and parameters. The knowledge of the process based on a given data:

$$D^{t-1} = \{u(1), y(1), ..., u(t-1), y(t-1)\} \tag{1}$$

is described by predictive c.p.d.f (conditional predictive density function),

$$p\left(y(t) \mid D^{t-1}, u(t)\right) \text{ for } t = 1, 2, 3,$$

where the predictive c.p.d.f. itself is,

$$p\left(y(t) \mid D^{t-1}, u(t)\right) = \int p\left(y(t) \mid D^{t-1}, u(t), \theta(t)\right) p\left(\theta(t) \mid D^{t-1}\right) d\theta(t) \tag{2}$$

We call equation (2) as model structure plus knowledge of the unknown parameters.

The data-update step (information contained in a new pair of data $\{u(t), y(t)\}$) is given by:

$$p(\theta(t) \mid D^t, u(t)) \propto p(y(t) \mid D^{t-1}, u(t), \theta(t)) p(\theta(t) \mid D^{t-1})$$

and the time-update step (information contained in the model) is given by:

$$p(\theta(t+1) \mid D^t) = \int p(\theta(t+1) \mid D^t, \theta(t)) p(\theta(t) \mid D^t) d\theta(t) \quad (3)$$

We call equation (3) as parameter development model.

Example: ARX model

The ARX model, shown in Figure 1, is the simplest model incorporating the stimulus signal. The estimation of the ARX model is the most efficient of the polynomial estimation methods because it is the result of solving linear regression equations in analytic form. Moreover, the solution is unique. In other words, the solution always satisfies the global minimum of the loss function. The ARX model therefore is preferable, especially when the model order is high. The disadvantage of the ARX model is that disturbances are part of the system dynamics. The transfer function of the deterministic part of the system and the transfer function of the stochastic part of the system have the same set of poles. This coupling can be unrealistic. The system dynamics and stochastic dynamics of the system do not share the same set of poles all the time. However, you can reduce this disadvantage if you have a good signal-to-noise ratio. When the disturbance $e(n)$ of the system is not white noise, the coupling between the deterministic and stochastic dynamics can bias the estimation of the ARX model. Set the model order higher than the actual model order to minimize the equation error, especially when the signal-to-noise ratio is low. However, increasing the model order can change some dynamic characteristics of the model, such as the stability of the model.

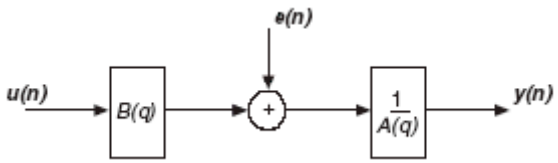

Figure 1. ARX Model Structure

ARX model:

$$\begin{aligned} y(t) &= \sum_{i=1}^{n} a_i(t) y(t-i) + \sum_{i=0}^{n} b_i(t) u(t-i) + e(t) \\ &= z^T(t)\theta(t) + e(t) \end{aligned}$$

with parameter vector:

$$\theta(t) = \left[a_1(t), ..., a_n(t), b_0(t), ..., b_n(t) \right]^T$$

and regressor vector (finite-memory subset of $\{D^{t-1}, u(t)\}$),

$$z(t) = \left[y(t-1), ..., y(t-n), u(t), ..., u(t-n) \right]^T$$

$e(t) \sim N(0, r)$ -white noise with known variance, independent of $\{D^{t-1}, u(t)\}$, known variance. The expected parameter changes will follow the random walk,

$$\theta(t+1) = \theta(t) + v(t)$$

where $v(t) \sim N(0, Q(t))$, is a white noise and independent of the past data.

The ARX model defines c.p.d.f as,

$$p\left(y(t) \mid D^{t-1}, u(t), \theta(t)\right) = N\left(z^T(t)\theta(t), r\right)$$

And the parameter development model defines c.p.d.f. as,

$$p\left(\theta(t+1) \mid D^t, \theta(t)\right) = N\left(\theta(t), Q(t)\right)$$

from the information above, we can form the prior distribution normal as

$$p\left(\theta(t) \mid D^{t-1}\right) = N\left(\hat{\theta}(t \mid t-1), C(t \mid t-1)\right)$$

The normality is reproduced and sufficient to operate on the sufficient statistics. Then, the data-update step-formulas became,

$$\hat{\theta}(t \mid t) = \hat{\theta}(t \mid t-1) + \frac{C(t \mid t-1)z(t)}{r + z^T(t)C(t \mid t-1)z(t)}\varepsilon(t \mid t-1)$$

$$C(t \mid t) = C(t \mid t-1) - \frac{C(t \mid t-1)z(t)z^T(t)C(t \mid t-1)}{r + z^T(t)C(t \mid t-1)z(t)}$$

$$\varepsilon(t \mid t-1) = y(t) - z^T(t)\hat{\theta}(t \mid t-1)$$

and the time-update step formulas given by,

$$\begin{aligned} \hat{\theta}(t+1 \mid t) &= \hat{\theta}(t \mid t) \\ C(t+1 \mid t) &= C(t \mid t) + Q(t) \end{aligned} \tag{4}$$

We call (4) as linear forgetting.

3. Alternative Parameter Development/Measurement Models

In this section we start with the possible parameter development histories. Set of alternative parameter development models with even c_i, $i = 0, ..., h$

$$p\left(\theta\left(t+1\right)|D^{t},\theta\left(t\right),c_{i}\left(t\right)\right)$$

That will tell us the information about parameter change at time *t* contained in "future" data. The update prior probability distribution is given by,

$$p\left(c_{i}\left(t\right)|D^{t}\right)$$

and to the *d*-step lag posterior distribution,

$$p\left(c_{i}\left(t\right)|D^{t+d}\right)$$

For this purpose, we require the tree of all possible "parameter development histories",

$$\begin{aligned}&p\left(c_{i_0}\left(t\right),c_{i_1}\left(t+1\right),...,c_{i_{k-1}}\left(t+k-1\right)|D^{t+k}\right)\propto\\&p\left(y\left(t+k\right)|D^{t+k-1},u\left(t+k\right),c_{i_0}\left(t\right),...,c_{i_{k-1}}\left(t+k-1\right)\right)\\&\quad\times p\left(c_{i_0}\left(t\right),c_{i_1}\left(t+1\right),...,c_{i_{k-1}}\left(t+k-1\right)|D^{t+k-1}\right).\end{aligned}$$

Time-update step becomes,

$$\begin{aligned}&p\left(c_{i_0}\left(t\right),c_{i_1}\left(t+1\right),...,c_{i_k}\left(t+k\right)|D^{t+k}\right)=\\&p\left(c_{i_k}\left(t+k\right)|D^{t+k}\right)p\left(c_{i_0}\left(t\right),c_{i_1}\left(t+1\right),...,c_{i_{k-1}}\left(t+k-1\right)|D^{t+k}\right).\end{aligned}\tag{5}$$

We call equation (5) as prior distribution independent on the previous changes. For posterior distribution we have,

$$p\left(c_{i_0}\left(t\right)|D^{t+d}\right)=\sum_{\{i_1,...,i_d\}}p\left(c_{i_0}\left(t\right),c_{i_1}\left(t+1\right),...,c_{i_d}\left(t+d\right)|D^{t+d}\right).$$

This posterior distribution is the branches starting from the root of the tree and can be explained as,

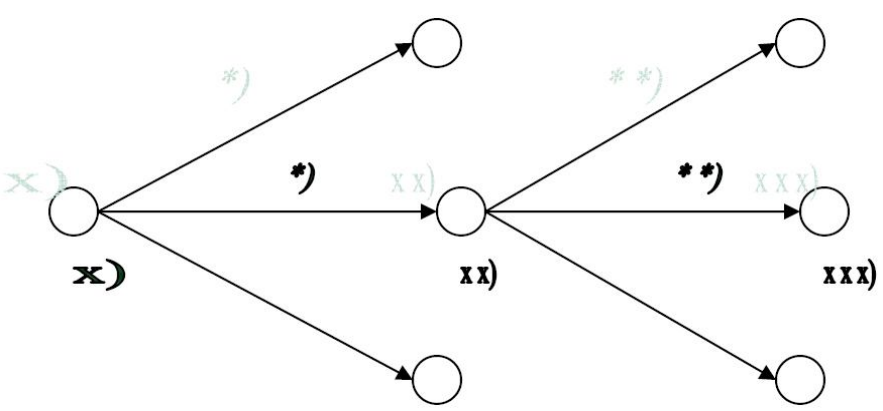

Figure 2. Tree of alternative hypotheses for depth 2.

Notes:

*) = $c_{i_0}(t)$,

**) = $c_{i_1}(t+1)$

x) = $p\left(\theta(t) \mid D^{t-1}\right)$;

xx) = $p\left(\theta(t+1) \mid D^{t}, c_{i_0}(t)\right)$;

xxx) = $p\left(\theta(t+2) \mid D^{t+1}, c_{i_0}(t), c_{i_1}(t+1)\right)$

Example: tree for ARX model

The parameter development-covariance matrices of parameter changes for ARX model is,

$$p\left(\theta(t+1) \mid D^{t}, \theta(t), c_i\right) = N\left(\theta(t), Q^{(i)}(t)\right)$$

and the time-update step for *i*-th hypothesis is

$$C\left(t+1 \mid t, c_i(t)\right) = C(t \mid t) + Q^{(i)}(t)$$

Evaluation of the posterior probability,

$$p\left(y(t+k) \mid D^{t+k-1}, u(t+k), c_{i_0}(t), ..., c_{i_{k-1}}(t+k-1)\right) =$$
$$N\left(z^T(t+k)\hat{\theta}\left(t+k \mid t+k-1, c_{i_0}, ..., c_{i_{k-1}}\right) : r + z^T(t+k) C\left(t+k ? t+k-1, c_{i_0}, ..., c_{i_{k-1}}\right) z^T(t+k)\right)$$

The set of alternative measurement model, with events c^j, $j = 0, ..., k$

$$p\left(y(t) \mid D^{t-1}, u(t), \theta(t), c^j(t)\right)$$

Then we can build up the structure tree using data/time update step and for robustness, we must concern with model of outliers, offset, frozen sensor, etc. For example outliers for ARX model can be written as,

$$p\left(y(t) \mid D^{t-1}, u(t), \theta(t), c^j(t)\right) = N\left(z^T(t)\theta(t), r^{(j)}\right)$$

and the data-update step conditioned by *j*-th hypothesis

$$\hat{\theta}(t\,|\,t)=\hat{\theta}(t\,|\,t-1)+\frac{C(t\,|\,t-1)z(t)}{r^{(j)}+z^{T}(t)C(t\,|\,t-1)z(t)}\varepsilon(t\,|\,t-1)$$
$$C(t\,|\,t)=C(t\,|\,t-1)-\frac{C(t\,|\,t-1)z(t)z^{T}(t)C(t\,|\,t-1)}{r^{(j)}+z^{T}(t)C(t\,|\,t-1)z(t)}$$

for large *r* this will reduce Kalman gain.

4. Tree Pruning

There are two approaches how to keep the tree depth within a given limit: (i) decision based: select the branch corresponding to "optimal" change, cut off other branches, shifts the root to the next code on the selected branch. (ii) approximation-based: replace the mixture in the first floor by a single c.p.d.f., use this approximation as the root of a new tree.

The decision based approach for tree pruning, using standard sequential statistical decision problem. We define loss function as $L(c_j,d_i)$, so the expected posterior loss (risk) for $d_j(t)$ based on data D^{t+d},

$$q\left(p_c(t\,|\,t+d),d_j(t)\right)=\sum_{i=0}^{h}L\left(c_i(t),d_j(t)\right)p\left(c_i(t)\,|\,D^{t+d}\right)$$

where $p_c(t\,|\,t+d)$ is notation for the marginal distribution. Bayes decision function easily evaluated, optimal decision can be write as

$$d^{*}(t)=\arg\min_{d_j(t)}\ q\left(p_c(t\,|\,t+d),d_j(t)\right)$$

and a symmetrical loss function is,

$$L(c_i,d_j)=\begin{matrix}0\ for & i=j,\\ 1\ for & i\neq j.\end{matrix}$$

This we call MAP (maximum a posterior probability) decision-making algorithm.

Loss function for parameter tracking typically non symmetrical: or c_0 no change. This make reasonable loss functions as

$$L(c_0,d_0)=L(c_1,d_1)=0$$
$$L(c_0,d_1)<L(c_1,d_0)$$

This will increased sensitivity to noise multiply by disable tracking capability. For more complex decisions: we should build another floor of the tree with variable depth tree optimal stopping rule. Control design for approximation is necessary that are: "decision time" *t*, "real time" *t+d*, and using the formula for control design information about parameters,

$$p\left(\theta\left(t+d\right)\mid D^{t+d}\right)=\sum_{\{i_0,\ldots,i_{d-1}\}} p\left(\theta\left(t+d\right)\mid D^{t+d},c_{i_0}\left(t\right),\ldots,c_{i_{k-1}}\left(t+d-1\right)\right)$$
$$\times p\left(c_{i_0}\left(t\right),\ldots,c_{i_{k-1}}\left(t+d-1\right)\mid D^{t+d}\right)$$

The last step we can do smoothing with using: filter with forward and backward run, linear forgetting-large Q separates "past from "future", and the last is piecewise constant parameters.

5. Conclusion

The Novel approach to time-varying parameter tracking can be done by using an extension of a Bayesian viewpoint (continuous and discrete parameters), prior information, set of alternative models of parameter changes/measurements, and improved tracking for abrupt changes.
There are still open problems for algorithm analysis and the prior probability distribution over hypotheses (we can extend using Markov model).

Acknowledgements

The work funded by the Short Term Research Project, Universiti Sains Malaysia, Grant 304/PJJAH/638072.

References

A. Zheng and M. Morari, 2000, Robust control of linear time-varying systems with constraints, *International Journal of Robust and Nonlinear Control* 10, 1063–1078.

Ljung, L., 1987, *System Identification: Theory For The User*, Prentice Hall, Inc., New Jersey,

N.J. Gordon., D.J. Salmond., & A.F.M. Smith., 1993, Novel approach to nonlinear-non-Gaussian Bayesian state estimation, *IEE Proceeding-F* 140 (2), 107-113.

P.J. Huber, 1981, *Robust Statistics*. New York: John Wiley and sons.

Saligrama R. Venkatesh and Munther A. Dahleh, 2001, On System Identification of Complex Systems from Finite Data, *IEEE TRANSACTIONS ON AUTOMATIC CONTROL*, 46, (2), 235-257.

Rousseeuw and A.M. Leroy. 1987, *Robust regression and outlier detection*. New York: John Wiley and sons.

Y. Bar-Shalom and T. Fortmann, 1988, *Tracking and Data Association*, Boston: Academic Press, ISBN 0-12-079760-7.

