Time delayed process model parameter estimation: a classification of techniques

Aidan O’Dwyer
Dublin Institute of Technology, aidan.odwyer@dit.ie

Follow this and additional works at: http://arrow.dit.ie/engscheleart

Part of the Controls and Control Theory Commons

Recommended Citation
TIME DELAYED PROCESS MODEL PARAMETER ESTIMATION: A CLASSIFICATION OF TECHNIQUES

A.O’Dwyer

School of Control Systems and Electrical Engineering, Dublin Institute of Technology, Kevin St., Dublin 8, Ireland.

Keywords: Delay, parameter estimation.

Abstract: An extensive though scattered literature exists on the estimation of the model parameters of time delayed processes. However, it is possible to identify themes that are common to many of the proposed techniques. The intention of this paper is to provide a framework against which the literature may be viewed.

1. Introduction

A time delay may be defined as the time interval between the start of an event at one point in a system and its resulting action at another point in the system. Delays are also known as transport lags, dead times or time lags; they arise in physical, chemical, biological and economic systems, as well as in the process of measurement and computation. For brevity reasons, this paper will consider only those applications where the delay is estimated in the presence of other process parameters.

The purpose of the identification determines the type of process model required. Newell and Lee [1] suggest that the model complexity that may be reasonably identified from experimental data depends on the data quality available and the analysis technique used. The authors suggest that a cautious approach is to identify a first order lag plus delay (FOLPD) model from the experimental data and that an optimistic approach is to identify a second order system plus delay (SOSPD) model from the data. Appropriate modelling methods for real processes are also considered by other authors [2, 3]. A broad conclusion from these and other papers is that even if the process has no physical delay, it is possible to model such a (possibly high order) process by a low order time delayed model; the delay estimated may be a combination of an actual delay and contributions due to high order dynamic terms in the process transfer function. It is also reasonable that either a FOLPD or SOSPD model should be estimated, as either of these approximate process models is sufficiently accurate for many applications. However, if a priori information on the process is available (such as the process order), the estimation of the full order time delayed model may be indicated.

Estimation methods for time delayed processes may be broadly classified into time domain and frequency domain techniques; these techniques may be either off-line or on-line, with on-line estimation requiring recursive estimation in a closed loop environment. Time domain estimation methods will be treated first. A number of off-line estimation techniques are documented, for single input, single output (SISO) and multi-input, multi-output (MIMO) model structures, in open loop and closed loop. A discussion of multiple model estimation techniques will then be carried out. A number of on-line estimation techniques will subsequently be treated, followed by a discussion of gradient methods for parameter estimation; the latter methods may be implemented in either open loop or closed loop, and in either an off-line or on-line manner. Frequency domain estimation techniques may be classified in a similar manner to time domain estimation methods. The use of the frequency domain, as a means of estimating the time delayed model parameters, has a certain intuitive appeal, since the delay contributes to just the phase term of the frequency response. Other possibilities for estimation are subsequently detailed. In each section, conclusions as to the applicability of various classes of methods will be drawn, as appropriate. General conclusions from the literature review will subsequently be drawn. For space reasons, not all relevant references can be included in the paper; such references will be available from the author at the conference.

2. Time domain methods for parameter and delay estimation

2.1 Off-line estimation methods

2.1.1 Experimental open loop methods

One of the first such methods was described by Ziegler and Nichols [4], in which the time constant and time delay of a FOLPD model are obtained by constructing a tangent to the experimental open loop step response at its point of inflection. The tangent intersection with the time axis at the step origin provides a time delay estimate; the time constant is estimated by calculating the tangent intersection with the steady state output value divided by the model gain. Similar tangent methods may also be used to determine SOSPD model parameters [5-7]. The major disadvantage of all these methods is the difficulty of determining the point of inflection in practice.

Some methods that eliminate this disadvantage use two or more points on the process step response to estimate FOLPD model parameters [8, 9] or use two,
three or more points on the process step or pulse response to estimate SOSP model parameters [10-14]. An alternative experimental method involves calculating appropriate model parameters from the area under the step response output curve [15-17]. Other methods are also of interest [18].

Experimental open loop tests have the advantage of simplicity. However, the parameters identified may vary with process operating conditions and the step change size and direction. In addition, the process must be sufficiently disturbed by the change, to obtain reasonably accurate dynamic information, with the possibility that the process may be forced outside the region of linear behaviour. There is also a reluctance among plant management to permit such disturbances to be introduced for parameter estimation purposes. The process time scale must also be known in advance in order to determine when the transient response has been completed.

2.1.2 Experimental closed loop methods

These methods typically involve the analytical calculation of the model parameters from unity feedback, proportionally controlled, closed loop experimental step response output measurements. The delay is often approximated by a rational polynomial in the continuous time domain [19-23], though this is not absolutely necessary [24]. Other authors calculate the ultimate gain and frequency of a unity feedback, proportionally controlled, closed loop system from the experimental step response, and subsequently determine the time delayed model parameters [25-28]. A combination of the methods may also be used to determine the best time delayed model [29]. Identification strategies in a unity feedback, PI, PID or dead-time compensated, closed loop system may also be used [3, 29-34].

Refinements to the published algorithms are possible; however, the robustness of many of the estimation methods to noise on the process response is questionable. This comment does not apply to the characteristic areas method [15], in which the area under the closed loop step response output curve is used to calculate the model parameters.

2.1.3 Multiple model estimation methods

These methods are based on estimating a number of different process models, for different delay and (often) model order values. The model parameters chosen minimise a cost function that depends on the difference between the process and model outputs. The model order, parameters and delay index (which is the integer value of the delay divided by the sample time) may be estimated [35-38]. Some authors concentrate on estimating the delay and process parameters only [39-46].

The attraction of multiple model estimation methods is that the grid searching used will facilitate the estimation of parameters corresponding to the global minimum of the cost function, even in the presence of local minima, provided enough models are estimated. However, the methods are computationally intensive.

2.2 On-line estimation methods

The delay may be approximated by a rational polynomial in the continuous time domain and the resulting model parameters estimated recursively, from which the delay may be deduced [44, 47, 48].

Alternatively, the method of overparameterisation may be used, which involves subsuming the delay term into an extended z domain numerator polynomial. The parameters are estimated recursively, and the delay is calculated based on the numerator parameters identified; for a noise free system, all numerator parameters whose indices are smaller than the delay index should be identified as zero. Only delay values that are integer multiples of the sample period are directly estimated by the method. The delay portion that is a fraction of the sample period may be calculated from the numerator parameters identified [49]; however, the robustness of the calculation method in the presence of noise is questionable. An overparameterisation method example is described by Kurz and Goedecke [50], who define a robust method for estimating the SISO model parameters that is equivalent to determining the best match between the impulse response of the overparameterised model and the impulse response of a non-overparameterised model with a pure delay; however, the method is computationally intensive. Other methods offer various trade-offs between robustness and computational load [44, 51-55]; the most promising method is defined by Teng and Sirisena [54], because of its relative computational simplicity. A recursive method to estimate the parameters, order and delay index for both a stochastic and deterministic system, using an overparameterised method to estimate the delay, is also described [56]. Some authors identify time delayed MIMO process models using the method of overparameterisation [57, 58].

The method of overparameterisation is a natural extension of methods used in delay-free identification applications. However, the computational burden of the identification algorithm increases with the square of the number of estimated parameters, the persistent excitation condition is more difficult to satisfy for overparameterised models and the high order numerator polynomial increases the likelihood of common factors in the numerator and denominator polynomials in the estimation model, rendering identification more difficult. A high-order correlation approach is an alternative to the overparameterisation method [59].

2.3 Gradient methods of parameter and delay estimation

Gradient methods of parameter estimation are based on updating the parameter vector (which includes
the delay) by a vector that depends on information about the cost function to be minimised. The gradient algorithms normally involve expanding the cost function as a second order Taylor's expansion around the estimated parameter vector. Typical gradient algorithms are the Newton-Raphson, the Gauss-Newton and the steepest descent algorithms, which differ in their updating vectors. The choice of gradient algorithm for an application depends on the desired speed of tracking and the computational resources available. It is important that the error surface in the direction of the delay (and indeed the other parameters) should be unimodal if a gradient algorithm is to be used successfully. However, the error surface is often non-unimodal. In these circumstances, strategies for locating global minima may involve multiple optimisation runs, each initiated at a different starting point with the starting points selected by sampling from a uniform distribution. The global minimum is then the local minimum with the lowest cost function value among all the local minima identified.

Gradient algorithms based on the Newton-Raphson method have been defined; Liu [60], for example, describes a parameter updating scheme for a general order time delayed model based on the algorithm. The Gauss-Newton algorithm has been used to estimate FOLPD model parameters, in a Smith predictor structure [61]. A number of modifications of the approach have also been considered [62, 63]. The Gauss-Newton algorithm has also been used in an open loop application [64, 65] to estimate FOLPD model parameters. Other such approaches are also described [66]. The straightforward nature of the steepest descent algorithm has motivated its application to the estimation of process parameters; Elmaguer et al. [67], for example, estimate the delay using the algorithm and estimate the non-delay parameters recursively. Other gradient algorithms have also been used for parameter estimation [68-71]; Gawthrop et al. [68], for example, update the delay based on the partial derivative of the error squared with respect to the delay. The most popular gradient algorithm is the Gauss-Newton algorithm, as it combines good tracking speed and moderate computational intensity.

3. Frequency domain methods for parameter and delay estimation

Typically, the process frequency response must be estimated before model parameters are estimated. Methods for estimating the process frequency response include correlation analysis, spectral analysis and methods based on the ratio of Fourier transforms [63].

The process frequency response may be used to graphically estimate FOLPD and SOSPD model parameters [72, 73] and the parameters of higher order delayed models [74]. The disadvantages of the method are the tediousness of the procedure and the introduction of errors in fitting model parameters using a trial and error approach; in addition, the identification of more general transfer function models is difficult using the method. The process frequency response may also be used to analytically estimate FOLPD and SOSPD model parameters [5, 75, 76] and the parameters of higher order delayed models [76, 77].

Alternatively, the model parameters may be estimated by minimising the squared error between the process and model frequency responses. For an arbitrary order time delayed model, many of the techniques available require a continuous time delay approximation, using an appropriate rational polynomial; the delay itself is not identified [78]. However, Dos Santos and De Carvalho [79] explicitly estimate the parameters of a general order time delayed model by determining the model order and the pole and zero value estimates iteratively from the delay, with the delay estimate calculated based on a least squares procedure from the phase plot. An alternative multiple model estimation method involves selecting the delay iteratively and determining the remaining model parameters in a least squares sense [73]. Other least squares methods have also been proposed [76, 80]. It is also possible to fit a low order delayed model to the process response, in a least squares sense [73, 81-85].

The time delayed model parameters may also be determined from the identification of one or more points on the process frequency response obtained when a relay is switched into the closed loop compensated system [17, 55, 86-104]. Indeed, further work in this area is possible, as it is more common to use such relay techniques for PI/PID autotuning rather than for model parameter estimation.

4. Other methods of process parameter and delay estimation

The identification of time delayed processes using neural networks is a subject of recent research. Bhat and McAvoy [105], for instance, propose a method to strip a back propagation neural network to its essential weights and nodes; the stripping algorithm is capable of identifying the delay and order of a FOLPD process (in the discrete time domain). More recent contributions have also been made [106, 107].

Process order estimation strategies may also be used to estimate the process delay (in the discrete time domain), since the delay appears as an increase in the numerator transfer function model order. Delay estimation using these strategies would depend on a priori knowledge of the order of the non-delay part of the process.

It is also possible to estimate the time delayed process parameters using the delta operator rather than the z (or shift) operator. Keviczky and Banyas [108], in an analogue of a method defined by these authors in the z domain [53], identify the delay index using overparameterisation in the delta domain. There is further scope to estimate the delay and other model parameters in the delta domain, using techniques similar
to those used in the z domain.

Finally, the use of genetic algorithms for process identification is beginning to attract interest. Genetic algorithms search from a population of points, use information about the cost function (rather than its derivative or other auxiliary knowledge used by gradient algorithms) and have a random component, quantified as a mutation rate, that helps drive the model parameters towards values corresponding to the global minimum of a possibly non-unimodal cost function; such cost functions often arise in the identification of delayed processes. Genetic algorithms are considered to be one extreme solution to the exploitation-exploration trade-off; the algorithms trade-off large computation time, and poor accuracy of the global minimum, with reliability in calculating the global minimum. Yang et al. [109], for example, use a genetic algorithm to estimate the denominator parameters and delay of a reduced order process model, while using the less computationally intensive least squares algorithm to subsequently determine the numerator model parameters (which is a linear problem).

5. Conclusions

This paper has considered a wide variety of methods for time delayed model parameter estimation, in both the continuous time and discrete time domains. It is clear that gradient techniques, both in the frequency and time domains, have the potential to rapidly estimate the model parameters [63]. The use of other methods, such as multiple model estimation methods or genetic algorithms, in combination with gradient methods, may be one way of determining the global minimum of the cost function with more certainty.

It remains true to declare that the choice of identification method (and indeed compensation method) for a process with delay depends on the application. There is still a lot of interest in the identification of FOLPD and/or SOSPD process models, using, for example, experimental closed loop methods or by analysing the process output when a relay is switched into the closed loop compensated system in place of the controller. This is due to the low computational intensity involved in identifying such models, to concerns about how complex a model may reasonably be identified from experimental data and to the subsequent use of PI or PID controllers for compensation purposes. There is scope to apply some of the identification methods in question to the estimation of the parameters of delayed MIMO process models.

The identification of higher order time delayed models is still conditioned on a priori information on the process; few applications exist in which the parameters of such higher order models are identified in a black box manner from process input and output data. In addition, few unified approaches to the estimation problem have emerged.

6. References


61. MARSHALL, J.E.: ‘Control of timedelay systems’ (IEE Control Engineering Series, Peter Peregrinus Ltd., 1979).