Adaptive Control Strategies for Industrial Use

Lecture Notes in Control and Information Sciences
Edited by M. Thoma and A. Wyner

137

S. L. Shah, G. Dumont (Eds.)

Adaptive Control Strategies for Industrial Use

Proceedings of a Workshop
Kananaskis, Canada, 1988

Springer-Verlag
Berlin Heidelberg New York
London Paris Tokyo Hong Kong

Editors
Dr. Sirish L. Shah
Dept. of Chemical Engineering
University of Alberta
Edmonton
Canada T6G 2G6

Dr. Guy Dumont
Dept. of Electrical Engineering
University of British Columbia
2356 Main Hall
Vancouver
Canada V6T 1W5

ISBN 3-540-51869-x Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-51869-x Springer-Verlag New York Berlin Heidelberg

Printing: Mercedes-Druck, Berlin
Binding: B. Helm, Berlin
2161/3020-543210 Printed on acid-free paper.

PREFACE

This book contains proceedings of the workshop on Adaptive Control Strategies for Industrial Use, held at Lodge Kananaskis, Alberta, Canada during June of 1988. Over 70 participants, 30 from industry and the remaining from academia, from 10 countries came together. 20 of the 26 papers presented at this workshop are published in this volume.

The objective of this workshop was to bring together engineers from industry and scientists from universities to focus attention on new developments and practical enhancements for using adaptive control in industry. The workshop was held over a two and one-half day period and provided a forum for a tutorial introduction, through survey-type plenary sessions, to the state of the art in adaptive process control. Attention was also focussed through technical sessions, vendor demonstrations and panel discussions on the process control needs of industry and the mechanisms for transfer of current adaptive control technology between academia and industry.

Two panel discussions were held during the workshop. The first panel discussion was titled "Process Control Needs of the Industrial Community". Panelists from the petroleum, petrochemical, mining and pulp and paper sectors of industry participated in the discussion. The general consensus was that while PID control is suitable for a majority of loops, adaptive control is appropriate for many difficult loops in industry. In addition, lack of suitably trained personnel was identified as the cause of slow technology transfer. The potential benefits of adaptive feedforward control were stressed in this as well as the second panel discussion. The second panel discussion titled "Is a General-Purpose Adaptive Controller Feasible?" was held with several leading academics as panelists. The answer to this question was a unanimous: "No" in the sense that a universal adaptive controller was not only impossible but also undesirable. The discussion then focussed attention on "What important features should the next generation of adaptive controllers have?" The answer to this question led to a long discussion with the main conclusion that a necessary requirement for a practical adaptive controller is an intelligent supervisory system which would oversee the integrated performance and tuning of the estimator, the controller and the appropriate signal conditioning and filtering. The workshop ended on this positive note.

There are a number of people to whom we owe many thanks for making this a successful meeting. Firstly, we would like to acknowledge the administrative and secretarial help we received through the Department of Chemical Engineering, University of Alberta, and the Pulp and Paper Centre, University of British Columbia, as well as the Conference Grants Committee at the University of Alberta for partial financial support of this meeting. Finally, it is a pleasure to acknowledge the help of our graduate students, who looked after the program preparation, registration details, and audio-visual requirements at the workshop.

May, 1989 Sirish L. Shah and Guy Dumont

TABLE OF CONTENTS

RECENT DEVELOPMENTS IN ADAPTIVE CONTROL

IMPLEMENTATION ISSUES IN ADAPTIVE CONTROL

APPLICATIONS OF ADAPTIVE CONTROL

AI AND INDUSTRIAL ADAPTIVE CONTROLLERS

Recent Developments in Adaptive Control

SELF-TUNING MULTISTEP OPTIMISATION CONTROLLERS

David W. Clarke
Department of Engineering Science, Oxford University
Oxford OX1 3PJ, England

Abstract

The widespread demand for increased product quality and the growing use of high-throughput energy-intensive plant means that alternatives to multi-loop PID controllers must be investigated. One significant approach, compatible with current powerful microcomputers, is multistep or long-range predictive control. This uses a process model and an optimization algorithm to determine the best set of future controls for achieving a desired closed-loop perfomance. The design of these controllers depends on the choices of model, optimization method, and performance index, but they have been applied successfully to a wide range of processes, including multi input/output, constrained, dead-time, and multi-modal plant.

The paper discusses the basic philosophy of long-range predictive control, the criteria for selecting an appropriate model structure, the choice of cost-function and its unconstrained and constrained optimization. In practice the process model must be determined experimentally, leading to a self-tuned or adaptive design, and modifications to the standard recursive least-squares parameter estimator are described. Of particular importance is the correct conditioning and filtering of data, especially when handling unmodelled dynamics. The initialisation and industrial application of self-tuned predictive control is outlined.

1 Introduction

Consider a spray-drying tower. A slurry is forced through a nozzle near the top of the tower so that droplets fall through a counter-current flow of heated air to settle as dried powder at the base. A 'quench' damper modulates the flow of heated air entering the tower and an exhaust damper controls the rate of air extraction at the top. The air flow through the tower is to be regulated: too high a flow entrains the particles whereas if it is too low the drying action is insufficient. Tower pressure must be set below atmospheric mainly for safety reasons: the operators might wish to open an inspection hatch. There is significant one-way interaction between the variables, as tower pressure is affected by both flow dampers, and quite strong nonlinearity: equally-spaced increments in the exhaust damper cause changes in the pressure with gains varying by 5 to 1. The relatively fast yet simple dynamics indicate a sample interval of less than 2 seconds.

Preheated oil, mixed with recycled gas, is fed into a hydrotreater reactor where the sulphur and nitrogen are converted to hydrogen sulphide and ammonia, and unsaturated hydrocarbons are saturated. The exothermic reactions are controlled by the addition of cold quench gas between each of the reactor beds using valves which must not be more than 60% open during normal operation. The control objective, for this highly interactive process with complex dynamics, is to maintain 'weighted average bed temperature' at a set-point whilst minimizing energy costs and ensuring that variables stay within prespecified limits. Measured disturbances, usable as feedforward signals, are the variations in feed flow-rate and recycled gas temperature. With a time to steady-state of 90 minutes and a sample interval of 3 minutes there are 30 samples over the plant's rise-time.

A high-speed compliant link is found to have detectable flexure modes ranging from 18Hz up to over 1kHz. It is controlled by a direct-drive DC motor and the tip (end-effector) position is sensed by a light untorqued rigid link. The angular positions of hub and tip are transmitted via shaft encoders to multiple microcomputers for feedback control. Variations of end-mass, such as when picking up a load, modify the modal frequencies. The required path for the tip is predetermined so that *future* values of reference are known. The link is designed for fast slewing which inevitably involves torque saturation of the motor and requires a sample rate of at least 60Hz.

The above are examples of typical high-performance control problems for which classical approaches are unsuitable without a great deal of effort in design and tuning. Yet they are all cases for which long-range predictive control (LRPC) has been successfully applied to real plant: MIMO Generalized Predictive Control (GPC) using a DEC LSI11 for the spray-dryer (Lambert,E., 1987); constrained Dynamic Matrix Control (DMC) using an IBM PC-AT for the reactor (Cutler and Hawkins, 1987); SISO GPC using twin Motorola 68020s for the flexible link (Lambert,M., 1987). This paper develops the basic ideas of LRPC, discusses what choices the designer has in achieving different performance objectives, and shows how the methods can be used in practice. It concentrates on methods such as GPC which can have a self-tuned or adaptive mode and for which there are theorems to demonstrate the stabilization of unstable, nonminimum-phase plant.

A long-range predictive controller is a combination of the following basic components:

- A *model* $\mathcal{M}(\theta)$ with which future plant ouputs $y(t+j)$ can be predicted at time t based on assumptions about present and future controls. A good choice of model structure is crucial for an effective LRPC design.
- Knowledge of the time-behaviour of the future set-point $w(t+j)$. If known in detail, as in robotics or some batch process applications, these are called *preprogrammed* set-points; otherwise a simple choice is to make the future set-point have the known current value.
- A cost- or objective-function $J(\mathbf{e}, \mathbf{u})$ where $\mathbf{e}$ is the vector of future system-

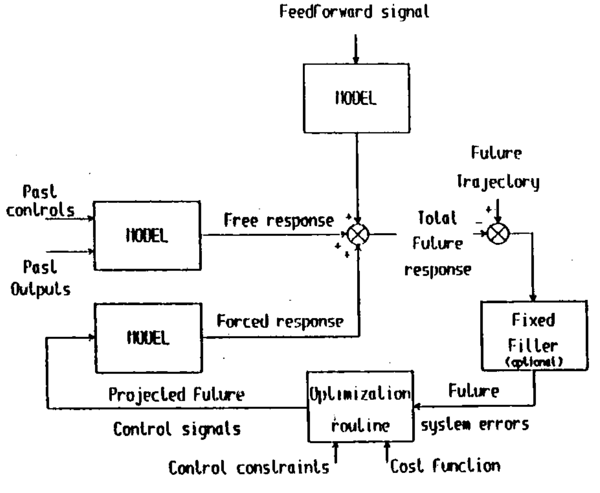

Figure 1: The structure of an LRPC loop

errors and **u** is the vector of appropriate controls. Important choices here include the *range* of future errors to include in the cost (e.g. one possibility is over the last half of the rise-time) and in the *number* of future control actions to consider as contributing to the future response.

- An optimization routine which minimises J, either *unconditionally* (assuming there are no constraints) or subject to known constraints such as control amplitude limits. The unconditional minimum can be derived analytically and hence involves the minimum number of computations.
- (Optionally) various filters and transfer-functions added to the design to broaden the range of performance objectives (e.g. closed-loop pole-placement) or to improve its robustness against practical inevitabilities such as unmodelled dynamics and disturbances.

The general structure of a long-range predictive controller is shown in Fig.1. Clearly there are a large range of possibilities for each of the components described above, so that very many distinct designs can be produced. Practical considerations, however, restrict the freedom of choice: these points are discussed in the following sections.

One of the most important questions is whether to choose full-value, offset, or incremental signals as a basis of the design. Suppose internal calculations take place using variables $\tilde{u}$, $\tilde{y}$ representing the plant's input and output, then we can select:

Full-value : $\tilde{u}(t) \equiv u(t)$, so that if $\tilde{u}$ is the result of the control computations its value is directly transmitted to the plant. Similarly $\tilde{y}(t) \equiv y(t)$ is used for the measured variable.

Offset : $\tilde{u}(t) \equiv u(t) - u_0$, where u_0 is some predetermined mean-value which is added to the algorithm output to obtain the transmitted control. Similarly $\tilde{y}(t) \equiv y(t) - y_0$; here the mean-value can be obtained either *ab initio* from a steady-state reading or recursively by low-pass filtering.

Incremental : $\tilde{u}(t) \equiv u(t) - u(t-1) \equiv \Delta u(t)$, where Δ is the backward-differencing operator. Hence a design with $\tilde{u}(t)$ produces a control $u(t) = u(t-1) + \tilde{u}(t)$. This is seen automatically to append an integrator in the forward-path – clearly desirable in order to have zero offset against constant load-disturbances. The internal variable $\tilde{y}(t)$ is also given by the differenced data $y(t) - y(t-1)$, being zero when the measured variable is constant.

There are many reasons why the incremental form is to be preferred. In self-tuning or adaptive control the estimated model is generally a local-linearization about the current operating point, but if the offset approach is used with constant $[u_0, y_0]$ these might not correspond to the correct values if the plant or its operating point change. In the control calculation of LRPC future values of $\tilde{u}$ are considered: one reasonable assumption is that there is a *control horizon* beyond which the control becomes constant and hence the corresponding increments are zeros. This means that there are significantly fewer variables involved in the optimization, leading to faster computations. This will be explored in more detail later.

There is one valid objection to the simple use of incremental data is that high frequencies (where there will be noise and effects of unmodelled dynamics) are emphasised. This is overcome by appropriate filtering (see later); indeed the offset approach using a computed u_0 obtained by the low-pass filter:

$$\hat{u}_0(t) = \beta \hat{u}_0(t-1) + (1-\beta)u(t),$$

gives an internal variable $\tilde{u}$ with overall *high-pass* filtering:

$$\tilde{u}(t) = \frac{\beta\Delta}{1-\beta q^{-1}} u(t),$$

where q^{-1} is the backward-shift operator. This filter blocks low frequencies (from the Δ operator) and has unit gain at high frequencies. Hence an incremental formulation in conjunction with suitable filtering has similar behaviour to the offset approach.

2 Choice of process model and output prediction

The purpose of a model is to predict the output response $\tilde{y}(t+j)$ based on past known inputs and measured outputs $\{\tilde{u}(t-i);\ \tilde{y}(t-i)\}$, and depending on the additional effect of current and future controls. (Obviously if there are constraints on the plant's control actuation such as motor-torque limits these should be taken into account as accurately as possible so that $\tilde{u}$ reflects the *actual* value affecting the plant rather than the possibly clipped demand of the LRPC algorithm). A model can be thought of as having two aspects: its *structure* $\mathcal{M}(.)$ and its actual *parameter* set $\{\theta\}$.

The derivation and computer implementation of an LRPC algorithm depends on the assumed structure $\mathcal{M}$. A 'good' design gives satisfactory answers to the following questions:

- Can $\mathcal{M}$ represent a very general class of plants? For example, can it deal with dead-time, unstable, lightly-damped, high-order systems simply by changes in parameter values? If so, then an LRPC code would not need *ad hoc* modifications when applied to different problem areas.
- Is the number of parameters minimal with $\mathcal{M}$ still giving adequate predictions? For then the computational burden could be minimised.
- Can prior knowledge be easily incorporated? This is particularly relevent to adaptive algorithms: for example if in robot manipulation the only unknown is the load mass a simple algorithm could be used to determine its value, and bounds on likely loads could be prespecified as an error-check.
- Is there a realistic assumption about load-disturbances? Some algorithms attempt to model these directly (which might be a fruitless exercise); most approaches make only implicit (and often unacknowledged) assumptions here. In practice the *minimal* assumption is that there is a constant load-disturbance (corresponding say to steady-state heat-loss or to constant load-torque) implying that even with a zero control signal the measured output would be non-zero. This is a further reason for insisting on offset or incremental models.

For any given application, associated with the structure are the particular parameters which need to be determined. This can be achieved by detailed mathematical modelling or simulation, but more often direct experiments on the plant are required. If the plant dynamics are reasonably time and set-point independent, a prior exercise can provide once-and-for-all parameters. If they vary with the set-point a series of related experiments could provide sets of parameters suitable for 'gain-scheduling'. In general, however, variations can (in principle: more difficult in practice) be handled by an adaptive algorithm which tracks changes as reflected in the plant's I/O behaviour. Hence acquiring good parameters involves answering the following questions:

- Do I need to perform specific open-loop experiments on the plant or is 'normal operating data' (such as with a closed-loop test) acceptable?

- Must I inject a predetermined test-signal such as a step, or can other signals be used?
- Should I/O data be prefiltered to accentuate the model fit over particular frequency ranges?
- What is the effect of noise, nonlinearities, and unmodelled dynamics on the quality of the model and the subsequent closed-loop LRPC behaviour?
- Will my estimator be able to track time-variations?

Details of parameter estimation are given later; here we simply note that parameter estimation is simpler if $\mathcal{M}$ involves a minimal number of parameters.

The simplest general-purpose model is the impulse response or weighting sequence from which the output $y(t)$ is derived by the convolution sum:

$$y(t) = \sum_{i=1}^{\infty} h_i u(t-i).$$

The only assumption here is superposition: the plant is linear with arbitrary dynamics having parameters $\{h_i\}$ being points on its unit-pulse response. In principle, however, there are an infinite number of parameters, so for implementations there must be *truncation* after some point N, assuming $[h_i = 0, i > N]$. This model can be written in operator form as:

$$y(t) = H(q^{-1})u(t),$$

where $H(q^{-1})$ is the FIR polynomial:

$$H(q^{-1}) = h_1 q^{-1} + h_2 q^{-2} + \cdots + h_N q^{-N}.$$

The problem with FIR models is that they require a very large number of parameters to represent stiff dynamic systems accurately. The sample interval h must be smaller than the smallest time-constant of interest and the model 'length' must be such that Nh exceeds the plant's settling-time. A typical choice of settling-time for overdamped dynamics is 5 times the largest time-constant: hence with only a 1:10 range of time-constants at least 50 parameters may be necessary.

A closely related plant representation is its step-response. Instead of taking the input to be a series of pulses it is considered as a set of 'moves' or increments. Superposition then provides the output:

$$y(t) = s_1 \Delta u(t-1) + s_2 \Delta u(t-2) + \cdots + s_i \Delta u(t-i) + \cdots,$$

where the parameters $\{s_i\}$ are points on the unit-step response. Again s_i must be truncated at the point N where the response has settled, and previous moves are assumed to provide an 'initial condition' y_0, giving:

$$y(t) = y_0 + \sum_{i=1}^{N} s_i \Delta u(t-i).$$

Similar considerations about the numbers of parameters apply; indeed $\{h\}$ and $\{s\}$ are related by the iterations:

$$s_0 = 0; s_i = s_{i-1} + h_i, \; i = 1, 2, \ldots N;$$

$$h_i = s_i - s_{i-1} = \Delta s_i, \; i = 1, 2, \ldots N.$$

Put simply, the step-response is the integral of the pulse-response.

An alternative with a long history in self-tuning control is the difference-equation:

$$\begin{aligned} y(t) + a_1 y(t-1) &+ a_2 y(t-2) + \cdots a_{na} y(t - na) = \\ b_1 u(t-1) &+ b_2 u(t-2) + \cdots b_{nb} u(t - nb). \end{aligned}$$

This is often called the DARMA (Deterministic AutoRegressive and Moving-Average) model, having the operator form:

$$A(q^{-1}) y(t) = B(q^{-1}) u(t),$$

where A and B are polynomials of degree na and nb in the backward-shift operator. All the above model forms can be related by:

$$\Delta S(q^{-1}) = H(q^{-1}) = A^{-1}(q^{-1}) B(q^{-1}),$$

though note here that an n'th order DARMA model can give an exact representation of a stiff n'th order plant as it does not need truncation. In particular it can emulate *unstable* processes which do not admit pulse- or step-response models. It can also handle deadtime by appropriate changes to the order of the $B(q^{-1})$ polynomial: k samples of deadtime increases nb by k and the leading k parameters become zero. Hence in adaptive applications where the deadtime might vary one possibility is to use a relatively high order B and accept that some leading or trailing coefficients might become insignificant. It is important here to ensure that the controller design (such as the choice of horizons) is insensitive to these changes: LRPC, unlike some approaches such as minimum variance, is acceptably robust.

A SISO state-space model, giving access to extensive theoretical and algorithmic results, is of the form:

$$\begin{aligned} \mathbf{x}(t+1) &= \mathbf{A}\mathbf{x}(t) + \mathbf{b}u(t) \\ y(t) &= \mathbf{c}^T \mathbf{x}(t). \end{aligned}$$

It is interesting to note that if a unit-pulse sequence {1,0,0,...} is injected into this model the response is given by:

$$\{h\} = \{0, \mathbf{c}^T\mathbf{b}, \mathbf{c}^T\mathbf{A}\mathbf{b}, \ldots\} \Longrightarrow h_i \equiv \mathbf{c}^T\mathbf{A}^{i-1}\mathbf{b}.$$

It is possible to convert directly from a DARMA model into state-space using an observable canonical form with $n = \max(na, nb)$ and:

$$\mathbf{A} = \begin{bmatrix} -a_1 & 1 & 0 & 0 & \cdots & 0 \\ -a_2 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & & \\ -a_n & 0 & 0 & 0 & \cdots & 1 \end{bmatrix},$$

$$\begin{aligned} \mathbf{b} &= [b_1, b_2, \ldots, b_n]^T, \\ \mathbf{c} &= [1, 0, 0, \ldots, 0]^T. \end{aligned}$$

At some stage in an LRPC design some estimate of the states (which in general are not directly accessible) is required, for which an observer of the form:

$$\hat{\mathbf{x}}(t) = \mathbf{A}\hat{\mathbf{x}}(t-1) + \mathbf{b}u(t-1) + \mathbf{k}(y(t) - \mathbf{c}^T\hat{\mathbf{x}}(t-1))$$

can be used. It is possible, however, to define a *nonminimal* state comprising simply of past inputs and outputs:

$$\mathbf{x}(t) \equiv [y(t), y(t-1), \ldots, y(t-na); u(t-1), u(t-2), \ldots, u(t-nb)]^T.$$

In this case all elements of the state are accessible; moreover a state-feedback of the form $u(t) = \mathbf{k}^T\mathbf{x}(t)$ gives a controller which can be immediately interpreted as a transfer-function.

There are many advocates for each of the model structures. A tentative (and possibly prejudiced) assessment is:

Impulse-response : Easy to formulate the corresponding LRPC designs. Standard PRBS/ crosscorrelation is usable for parameter estimation. Needs many parameters for a good fit with stiff or lightly-damped dynamics, so adaptation might be poor. Cannot be applied to unstable plant (unless stabilized by inner-loop feedback). Some highly complex process dynamics might be handled well: consider for example a plant with parallel paths having significantly different dynamics. Significant truncation problem: how big should N be?

Step-response : As above: initial parameter estimation (reaction curve) even easier *provided* that there are no load-disturbances during the test. Its incremental model formulation is more 'natural'.

DARMA : Minimal parameterization. Can deal with dead-time, lightly-damped and unstable dynamics. No truncation problem. Need to choose two (na, nb) rather than one (N) model orders (in practice choose na equal to the number of 'difficult' poles and nb large enough to deal with the expected range of dead-time). Must use an algorithm (Section 4) for parameter estimation, though there are simple results for obtaining second-order dead-time models from reaction curves.