Automatica, Vol. 25, No. 3, pp. 335–348, 1989
Printed in Great Britain.

0005-1098/89 $3.00 + 0.00
Pergamon Press plc

Survey Paper

Model Predictive Control: Theory and Practice—a Survey*

CARLOS E. GARCÍA,† DAVID M. PRETT† and MANFRED MORARI‡

The survey concludes that the flexible constraint-handling capabilities of Model Predictive Control make it most suitable for demanding multivariable process control problems.

Key Words—Computer control; predictive control; process control; quadratic programming; (constrained control).

Abstract—We refer to Model Predictive Control (MPC) as that family of controllers in which there is a direct use of an explicit and separately identifiable model. Control design methods based on the MPC concept have found wide acceptance in industrial applications and have been studied by academia. The reason for such popularity is the ability of MPC designs to yield high performance control systems capable of operating without expert intervention for long periods of time. In this paper the issues of importance that any control system should address are stated. MPC techniques are then reviewed in the light of these issues in order to point out their advantages in design and implementation. A number of design techniques emanating from MPC, namely Dynamic Matrix Control, Model Algorithmic Control, Inferential Control and Internal Model Control, are put in perspective with respect to each other and the relation to more traditional methods like Linear Quadratic Control is examined. The flexible constraint handling capabilities of MPC are shown to be a significant advantage in the context of the overall operating objectives of the process industries and the 1-, 2-, and ∞-norm formulations of the performance objective are discussed. The application of MPC to non-linear systems is examined and it is shown that its main attractions carry over. Finally, it is explained that though MPC is not inherently more or less robust than classical feedback, it can be adjusted more easily for robustness.

INTRODUCTION

THE PETRO-CHEMICAL industry is characterized as having very dynamic and unpredictable marketplace conditions. For instance, in the course of the last 15 years we have witnessed an enormous variation in crude and product prices. It is generally accepted that the most effective way to generate the most profit out of our plants while responding to marketplace variations with minimal capital investment is provided by the integration of all aspects of automation of the decision making process (García and Prett, 1986; Prett and García, 1988).

- *Measurements*. The gathering and monitoring of process measurements via instrumentation.
- *Control*. The manipulation of process degrees of freedom for the satisfaction of operating criteria. This typically involves two layers of implementation: the single loop control which is performed via analog controllers or rapid sampling digital controllers; and the control performed using real-time computers with relatively large CPU capabilities.
- *Optimization*. The manipulation of process degrees of freedom for the satisfaction of plant economic objectives. It is usually implemented at a rate such that the controlled plant is assumed to be at steady state. Therefore, the distinction between control and optimization is primarily a difference in implementation frequencies.
- *Logistics*. The allocation of raw materials and scheduling of operating plants for the maximization of profits and the realization of the company's program.

Each one of these automation layers plays a unique and complementary role in allowing a company to react rapidly to changes. Therefore, one layer cannot be effective without the others. In addition, the effectiveness of the whole approach is only possible when all manufacturing plants are integrated into the system.

Although maintaining a stable operation of the process was possibly the only objective of control systems in the past, this integration imposes more demanding requirements. In the petro-chemical industries control systems need to satisfy one or more of the following practical performance criteria.

- *Economic*. These can be associated with either maintaining process variables at the targets dictated by the optimization phase or dynamically minimizing an operating cost function.
- *Safety and environmental*. Some process variables must not violate specified bounds for reasons of personnel or equipment safety, or because of environmental regulations.
- *Equipment*. The control system must not drive

* Received 5 February 1987; revised 22 June 1988; received in final form 17 October 1988. The original version of this paper was presented at the IFAC Workshop on Model Based Control which was held in Atlanta, Georgia, U.S.A., during June 1988. The Published Proceedings of this IFAC meeting may be ordered from: Pergamon Press plc, Headington Hill Hall, Oxford OX3 0BW, U.K. This paper was recommended for publication in revised form by Editor K. J. Åström.

† Shell Development Company, P.O. Box 1380, Houston, TX 77110, U.S.A.

‡ Chemical Engineering, 206–41, California Institute of Technology, Pasadena, CA 91125, U.S.A. Author to whom correspondence should be addressed.

the process outside the physical limitations of the equipment.

• *Product quality*. Consumer specifications on products must be satisfied.

• *Human preference*. There exist excessive levels of variable oscillations or jaggedness that the operator will not tolerate. There can also be preferred modes of operation.

In addition, the implementation of such integrated systems is forcing our processes to operate over an ever wider range of conditions. As a result, we can state the control problem that any control system must solve as follows.

On-line update the manipulated variables to satisfy multiple, changing performance criteria in the face of changing plant characteristics.

The whole spectrum of process control methodologies in use today is faced with the solution of this problem. The difference between these methodologies lies in the particular assumptions and compromises made in the mathematical formulation of performance criteria and in the selection of a process representation. These are made primarily to simplify the mathematical problem so that its solution fits the existing hardware capabilities. The natural mathematical representation of many of these criteria is in the form of dynamic objective functions to be minimized and of dynamic inequality constraints. The usual mathematical representation for the process is a dynamic model with its associated uncertainties. The importance of uncertainties is increasingly being recognized by control theoreticians and thus are being included explicitly in the formulation of controllers. However, one of the most crucial compromises made in control is to ignore constraints in the formulation of the problem. As we explain below these simplifications can deny the control system of its achievable performance.

It is a fact that in practice the operating point of a plant that satisfies the overall economic goals of the process will lie at the intersection of constraints (Arkun, 1978; Prett and Gillette, 1979). Therefore, in order to be successful, any control system must anticipate constraint violations and correct for them in a systematic way: violations must not be allowed while keeping the operation close to these constraints. The usual practice in process control is to ignore the constraint issue at the design stage and then "handle" it in an *ad hoc* way during the implementation. Since each petro-chemical process (or unit) is unique we cannot exploit the population factor as in other industries (e.g. aerospace). That is, we cannot afford extreme expenses in designing an *ad hoc* control system that we know will not work in another process and therefore its cost cannot be spread over a large number of applications. Due to the increase in the number of applications of this type (resulting from the need to achieve integration), this implies an enormous burden both in the design and maintenance costs of these loops. In our experience, these costs more than offset the profitability of any *ad hoc* control system.

In conclusion, economics demand that the control systems must be designed with no *ad hoc* fixups and transparent specification of performance criteria such as constraints. Our experience has demonstrated that Model Predictive Control (MPC) techniques provide the only *methodology* to handle constraints in a systematic way during the design and implementation of the controller. Moreover, in its most general form MPC is not restricted in terms of the model, objective function and/or constraint functionality. For these reasons, it is the only methodology that currently can reflect most directly the many performance criteria of relevance to the process industries and is capable of utilizing any available process model. This is the primary reason for the success of these techniques in numerous applications in the chemical process industries.

In this paper the MPC methodology is reviewed and compared with other seemingly identical techniques. We particularly emphasize the unconstrained version of MPC since it is only in this form that it is possible to compare it with other schemes. Then the several existing forms of constrained MPC are reviewed, concluding with the non-linear MPC approaches. Although the issue of model uncertainties in MPC techniques is not dealt with in this paper, some comments on robustness of MPC are included.

HISTORICAL BACKGROUND

The current interest of the processing industry in MPC can be traced back to a set of papers which appeared in the late 1970s. In 1978 Richalet *et al.* described successful applications of "Model Predictive Heuristic Control" and in 1979 engineers from Shell (Cutler and Ramaker, 1979; Prett and Gillette, 1979) outlined "Dynamic Matrix Control" (DMC) and reported applications to a fluid catalytic cracker. In both algorithms an *explicit* dynamic model of the plant is used to predict the effect of future actions of the manipulated variables on the output (thus the name "Model Predictive Control"). The future moves of the manipulated variables are determined by optimization with the objective of minimizing the predicted error subject to operating constraints. The optimization is repeated at each sampling time based on updated information (measurements) from the plant.

Thus, in the context of MPC the control problem including the relative importance of the different objectives, the constraints, etc. is formulated as a dynamic optimization problem. While this by itself is hardly a new idea, it constitutes one of the first examples of large-scale dynamic optimization applied routinely in real time in the process industries.

The MPC concept has a long history. The connections between the closely related minimum time optimal control problem and Linear Programming were recognized first by Zadeh and Whalen (1962). Propoi (1963) proposed the moving horizon approach which is at the core of all MPC algorithms. It became known as "Open Loop Optimal Feedback". The extensive work on this problem during the 1970s was reviewed in the thesis by Gutman (1982). The connection between this work and MPC was discovered by Chang and Seborg (1983).

Since the rediscovery of MPC in 1978 and 1979, its popularity in the Chemical Process Industries has increased steadily. Mehra *et al.* (1982) reviewed a number of applications including a superheater, a steam generator, a wind tunnel, a utility boiler connected to a distillation column and a glass furnace. Shell has applied MPC to many systems, among them a fluid catalytic cracking unit (Prett and Gillette, 1979) and a highly non-linear batch reactor (García, 1984). Matsko (1985) summarized several successful implementations in the pulp and paper industries.

Several companies (Bailey, DMC, Profimatics, Setpoint) offer MPC software. Cutler and Hawkins (1987) report a complex industrial application to a hydrocracker reactor involving seven independent variables (five manipulated, two disturbance) and four dependent (controlled) variables including a number of constraints. Martin *et al.* (1986) cites seven completed applications and ten under design. They include: fluid catalytic cracker—including regenerator loading, reactor severity and differential pressure controls; hydrocracker (or hydrotreater) bed outlet temperature control and weight average bed temperature profile control; hydrocracker recycle surge drum level control; reformer weight average inlet temperature profile control; analyzer loop control. The latter has been described in more detail by Caldwell and Martin (1987). Setpoint (Grosdidier, 1987) has applied the MPC technology to: fixed and ebulating bed hydrocrackers; fluid catalytic crackers; distillation columns; absorber/stripper bottom C_2 composition control and other chemical and petroleum refining operations.

In academia MPC has been applied under controlled conditions to a simple mixing tank and a heat exchanger (Arkun *et al.*, 1986) as well as a coupled distillation column system for the separation of a ternary mixture (Levien, 1985; Levien and Morari, 1987). Parrish and Brosilow (1985) compared MPC with conventional control schemes on a heat exchanger and an industrial autoclave.

Most of the applications reported above are multivariable and involve constraints. It is exactly these types of problems which motivated the development of the MPC control techniques. Largely independently a second branch of MPC emerged, the main objective of which is *adaptive* control. Peterka's predictive controller (Peterka, 1984), Ydstie's extended-horizon design (Ydstie, 1984) and EPSAC developed by DeKeyser and van Cauwenberghe (1982), DeKeyser *et al.* (1985) are in this category as well as Clarke's generalized predictive control algorithm (Clarke *et al.*, 1987a, b). These developments are essentially limited to SISO systems with extension to the MIMO case conceptually straightforward but very involved when the details are considered. The constrained case is not considered in any detail in these papers. Because of the different underlying philosophy these algorithms are outside the focus of this paper. Nevertheless some cross references will be useful at times because these algorithms were largely developed for the non-adaptive case with the adaptation added in an *ad hoc* manner based on recursive least squares (or similar) parameter estimates. The stability and robustness of the *adaptive* scheme was generally not analyzed.

MODELS

All the derivations in this paper will be carried out for general MIMO systems. Occasionally, in the interest of providing special insight SISO systems will be discussed separately. The idea of MPC is not limited to a particular system description, but the computation and implementation depend on the model representation. Depending on the context we will readily switch between state space, transfer matrix and convolution type models. We will assume the system to be described in state space by

$$x(k) = Ax(k-1) + Bu(k-1) \tag{1}$$

$$y(k) = Cx(k). \tag{2}$$

For zero-initial conditions the equivalent transfer matrix representation is

$$y(z) = P(z)u(z) \tag{3}$$

where

$$P(z) \triangleq C(zI - A)^{-1}B. \tag{4}$$

Because most chemical engineering processes are open-loop stable our discussion will be limited to stable systems. The extension of the presented results to unstable systems is described elsewhere (Morari and Zafiriou, 1989). When A is stable the inverse in (4) can be expanded into a Neuman series

$$P(z) = \sum_{i=0}^{\infty} CA^iBz^{-i-1} \tag{5}$$

$$\triangleq \sum_{i=1}^{\infty} \bar{H}_i z^{-i} \tag{6}$$

where $\bar{H}_i$ are the impulse response coefficients, the magnitudes of which vanish as $i \to \infty$. Thus, in the time domain we have the truncated *impulse response* model

$$y(k) = \sum_{i=1}^{n} \bar{H}_i u(k-i) \tag{7}$$

and with the definitions

$$\bar{H}_i = H_i - H_{i-1} \tag{8}$$

$$H_i = \sum_{j=1}^{i} \bar{H}_j \tag{9}$$

the truncated *step response* model

$$y(k) = \sum_{i=1}^{n-1} H_i \Delta u(k-i) + H_n u(k-n) \tag{10}$$

where

$$\Delta u(k) = u(k) - u(k-1) \tag{11}$$

and H_i are the step response coefficients. Depending on the time delay structure of the system the leading step response coefficient matrices may be zero or have zero elements.

MPC ALGORITHM FORMULATIONS

The name "Model Predictive Control" arises from the manner in which the control law is computed (Fig. 1). At the present time k the behavior of the process over a horizon p is considered. Using a model the process response to changes in the manipulated variable is predicted. The moves of the manipulated

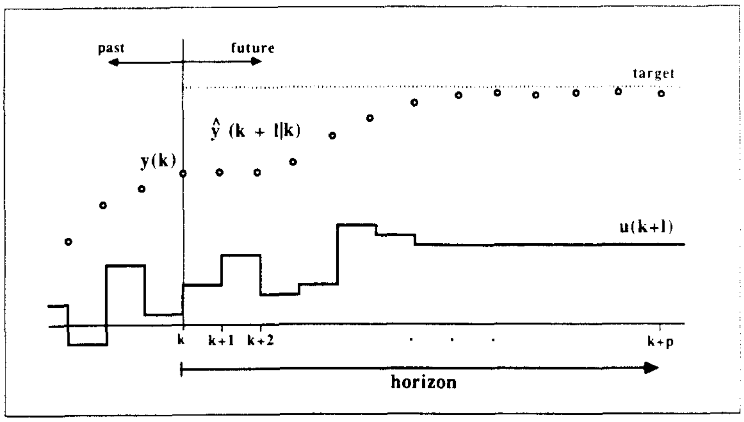

FIG. 1. The "moving horizon" approach of Model Predictive Control.

variables are selected such that the predicted response has certain desirable characteristics. Only the first computed change in the manipulated variable is implemented. At time $k+1$ the computation is repeated with the horizon moved by one time interval.

We will demonstrate how DMC and MAC are derived. All other MPC algorithms which have been proposed are very similar.

Dynamic Matrix Control (*DMC*)

The manipulated variables are selected to minimize a quadratic objective

$$\min_{\Delta u(k)\cdots\Delta u(k+m-1)} \sum_{l=1}^{p} \|\hat{y}(k+l \mid k) - r(k+l)\|^2_{\Gamma_l} + \|\Delta u(k+l-1)\|^2_{B_l} \quad (12)$$

$$\hat{y}(k+l \mid k) = \sum_{i=1}^{l} H_i \Delta u(k+l-i) + \sum_{i=l+1}^{n-1} H_i \Delta u(k+l-i) + H_n u(k+l-n) + \hat{d}(k+l \mid k) \quad (13)$$

$$\hat{d}(k+l \mid k) = \hat{d}(k \mid k) = y_m(k) - \sum_{i=1}^{n-1} H_i \Delta u(k-i) + H_n u(k+l-n) \quad (14)$$

$$\sum_{l=1}^{p} C^j_{y_l} \hat{y}(k+l \mid k) + C^j_{u_l} u(k+l-1) + c^j \le 0; \quad j = 1, n_c \quad (15)$$

$\hat{y}(k+l \mid k)$ = predicted value of y at time $k+l$ based on information available at time k

$\hat{d}(k+l \mid k)$ = predicted value of additive disturbances at process output at time $k+l$ based on information available at time k

$y_m(k)$ = measurement of y at time k

$\Delta u(k+l) = u(k+l) - u(k+l-1)$

$H_i, i = 1, n$ = model step response matrix coefficient

n = truncation order

n_c = number of constraints

p = horizon length (in general $p \gg n$)

m = number of manipulated variable moves in the future ($\Delta u(k+l) = 0 \ \forall l \ge m$; $m < p$)

$\|x\|^2_Q = x^T Q x$

Γ_l, B_l = weighting matrices

$C^j_{y_l}, C^j_{u_l}, c^j$ = constant matrices.

The prediction of output (13) involves three terms on the right-hand side. The first term includes the present and all future moves of the manipulated variables which are to be determined so as to solve (12). The second term includes only past values of the manipulated variables and is completely known at time k. The third term is the predicted disturbance $\hat{d}$ which is obtained from (7). $\hat{d}(k+l \mid k)$ is assumed constant for all future times ($l \ge 0$). At time k it is estimated as the difference between the measured output $y(k)$ and the output predicted from the model. In block diagram notation (14) corresponds to a model $\tilde{P}$ in parallel with the plant P (Fig. 2(a)) with the resulting feedback signal equal to $\hat{d}(k \mid k)$. Equations (12)–(15) define a Quadratic Program which is solved on-line at every time step. This "controller" is represented by block Q in Fig. 2(a).

Though computationally more involved than standard linear time invariant algorithms, the flexible constraint handling capabilities of MPC are very attractive for practical applications: a stuck valve can be simply specified by the operator on the console as an additional constraint for the optimization program. The algorithm will automatically adjust the actions of all the other manipulated variables to compensate for this failure situation as well as possible. In an unexpected emergency which a traditional fixed-logic scheme might find difficult to cope with, MPC will keep the process operating safely away from all constraints or allow the operator to shut it down in a smooth manner.

Model Algorithmic Control (*MAC*)

MAC is distinctive from DMC in three aspects.

(1) Instead of the step response model involving Δu, an impulse response model involving u is employed. If the input u is penalized in the quadratic objective, then the controller does not remove offset. This can be corrected by a static offset compensator (García and Morari, 1982). If the input u is not penalized then extremely awkward procedures are necessary to treat non-minimum phase systems (Mehra and Rouhani, 1980).

(2) The number of input moves m is not used for tuning ($m = p$).

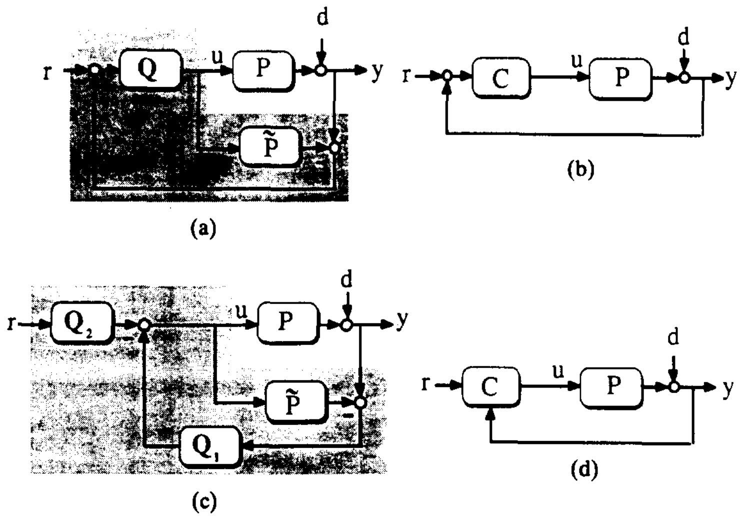

FIG. 2. (a) IMC structure. (b) Equivalent classic feedback structure. (c) Two-degrees-of-freedom IMC structure. (d) Equivalent two-degrees-of-freedom classic feedback structure.

(3) The disturbance estimate (14) is filtered. Let $y(k)$ be the measurement and $\hat{y}(k)$ the model prediction. Then the disturbance estimate is defined recursively as

$$\hat{d}(k+l \mid k) = \alpha\hat{d}(k+l-1 \mid k) + (1-\alpha)(y_m(k) - \hat{y}(k)) \quad (16)$$

with $\hat{d}(k \mid k) = 0$, $0 \le \alpha < 1$. Equation (16) adds a first-order exponential filter with an adjustable parameter α in the feedback path (Fig. 2(a)). For $r = 0$ this is equivalent to augmenting Q. α is a much more direct and convenient tuning parameter than the weights, horizon length, etc. in the general MPC formulation. α is directly related to closed-loop speed of response, bandwidth and robustness, but does not affect nominal ($P = \tilde{P}(z)$) stability. García and Morari (1982, 1985a,b) have analyzed the effect of this filter in detail.

Analysis

The MPC formulation (12)–(15) looks reasonable and attractive, and has been used extensively in industrial applications. However, a complete and general analysis of its properties (stability, robustness and performance) is not possible with the currently available tools. In general, the resulting control law is time varying and cannot be expressed in closed form. We would like to compare MPC with other design techniques and discuss alternate formulations and extensions. For this purpose we will concentrate first on the unconstrained case because only here a rigorous analysis is possible. Then we discuss the different ways by which constraints can be handled and their implications. Finally we will review some extensions to non-linear systems.

UNCONSTRAINED MPC

Without constraints problem (12)–(14) is a standard linear least squares problem which can be solved explicitly quite easily. With the moving horizon assumption a linear time invariant controller is found. García and Morari (1985b) have shown how to obtain the controller transfer function from the linear least squares solution.

Structure

García and Morari (1982) were the first to show that the structure depicted in Figs 2(a) and (c) is inherent in all MPC and other control schemes. It will be referred to as the Internal Model Control (IMC) structure in this paper. Here P is the plant, $\tilde{P}$ a model of the plant and Q (Q_1 and Q_2) the controller(s). y is the measured output, r the reference signal (setpoint), u the manipulated variable and d the effect of the unmeasured disturbances on the output. The total MPC system which has to be implemented consists of the model $\tilde{P}$ and the controller(s) Q (Q_1 and Q_2) and is indicated by the shaded box in Fig. 2. In this section and throughout most of the paper we will assume that the model $\tilde{P}$ is a perfect description of the plant ($P = \tilde{P}$). We will retain the tilde ($\tilde{\ }$), however, to emphasize the distinction between the real plant P and the model $\tilde{P}$ which is a part of the control system. The IMC structures in Figs 2(a) and (c) have largely the same characteristics. Initially we will concentrate our analysis on Fig. 2(a). Subsequently the advantages of employing two controller blocks Q_1 and Q_2 as in Fig. 2(c) will be addressed.

The following three facts are among the reasons why MPC is attractive.

Fact 1. The IMC structure in Fig. 2(a) and the classic control structure in Fig. 2(b) are equivalent in the sense that any pair of external inputs $\{r, d\}$ will give rise to the same internal signals $\{u, y\}$ if and only if Q and C are related by

$$Q = C(\mathrm{I} + \tilde{P}C)^{-1} \quad (17)$$

$$C = Q(\mathrm{I} - \tilde{P}Q)^{-1}. \quad (18)$$

Fact 2. If $P = \tilde{P}$ then the relation between any input and output in Fig. 2(a) is *affine*† in the controller Q. In particular

$$y = PQ(r - d) + d \tag{19}$$

$$e = y - r = (\mathrm{I} - PQ)(d - r). \tag{20}$$

Fact 3. If P is stable, then the MPC system in Fig. 2(a) is internally stable if and only if the classic control system with C defined by (18) is internally stable. In particular when $P = \tilde{P}$ the MPC system is stable if and only if Q is stable.

These facts have the following important implications.

Because of Fact 1 the performance of unconstrained MPC is not inherently better than that of classic control as one might be led to believe from the literature. Indeed, for any MPC there is an equivalent classic controller with identical performance.

Q can be considered an alternate parametrization of the classic feedback controller C, albeit one with very attractive properties: the set of all controllers C which gives rise to closed-loop stable systems is essentially impossible to characterize. On the contrary the set of all controllers Q with the same property is simply the set of all stable Q's (Fact 3). Furthermore, all important transfer functions (e.g. (19) and (20)) are affine in Q but non-linear functions of C (Fact 2). From a mathematical point of view it is much simpler to optimize an affine function of Q by searching over all stable Q's than it is to optimize a non-linear function of C subject to the complicated constraint of closed-loop stability. From an engineering viewpoint it is attractive to adjust a controller Q which is *directly* related to a setpoint and disturbance response (19) and (20) and where (in the absence of model uncertainty) closed-loop stability is automatically guaranteed as long as Q is stable. On the other hand even when C is a simple PID controller it is usually not obvious how closed-loop performance is affected by the three adjustable parameters and for what parameter values the closed-loop system is stable. As apparent from (19) Q plays the role of a *feedforward* controller. The design of feedforward controllers is generally much simpler than that of feedback controllers.

The main limitation of MPC in Fig. 2(a) is apparent from (20): both the disturbances d and the reference signals r affect the error e through the same transfer matrix $(\mathrm{I} - PQ)$. If r and d have different dynamic characteristics it is clearly impossible to select Q simultaneously for good setpoint tracking and disturbance rejection. For the "Two-degrees-of-freedom Structure" in Fig. 2(c) and the equivalent classic structure in Fig. 2(d) (for $P = \tilde{P}$) we find

$$e = (\mathrm{I} - \tilde{P}Q_1)d - (\mathrm{I} - \tilde{P}Q_2)r. \tag{21}$$

Here in the absence of model error the two controller blocks make it possible to design independently for good disturbance response and setpoint following. The equivalent classical feedback controller is described by

$$u = Q_1(\mathrm{I} - \tilde{P}Q_1)^{-1}Q_1^{-1}Q_2 r - Q_1(\mathrm{I} - \tilde{P}Q_1)^{-1}y$$
$$\triangleq C_1 r - C_2 y. \tag{22}$$

An excellent historical review of the origins of the structure in Fig. 2(a), which has as its special characteristic a model in parallel with the plant, is provided by Frank (1974). It appears to have been discovered by several people simultaneously in the late 1950s. Newton *et al.* (1957) used the structure to transform the closed-loop system into an open-loop one so that the results of Wiener could be applied to find the H_2-optimal controller Q. When the Smith Predictor (Smith, 1957) is written in the form shown in Fig. 3 where $\tilde{P}^*$ is the SISO process model without time delay it can be noticed that its structure also contains a process model in parallel with the plant. Independently Zirwas (1958) and Giloi (1959) suggested the predictor structure for the control of systems with time delay. Horowitz (1963) introduced a similar structure and called it "model feedback".

Frank (1974) first realized the general power of this structure, fully exploited it and extended the work by Newton *et al.* (1957) to handle persistent disturbances and setpoints. Youla *et al.* (1976a,b) extended the convenient "Q-parametrization" of the controller C to handle unstable plants. In 1981 Zames ushered in the era of H_∞-control utilizing for his developments the Q-parametrization. At present it is used in all robust controller design methodologies.

Unaware of all these developments the process industries both in France (Richalet *et al.*, 1978) and in the U.S. (Cutler and Ramaker, 1979; Prett and Gillette, 1979) exploited the advantages of the parallel model/plant arrangement. Brosilow (1979) utilized the Smith Predictor parametrization to develop a robust design procedure and García and Morari (1982, 1985a,b) unified all these concepts and referred to the structure in Fig. 2(a) as Internal Model Control (IMC) because the process *model* is explicitly an *internal* part of the controller.

The two-degrees-of-freedom structure is usually attributed to Horowitz (1963). It has been analyzed by many people since (Vidyasagar, 1985).

Tuning guidelines

As we will analyze in more detail below the problem (12)–(14) is very closely related to the standard Linear Quadratic Optimal Control problem, for which a wealth of powerful theoretical results is available. In particular, exact conditions on the tuning parameters are known which yield a stabilizing

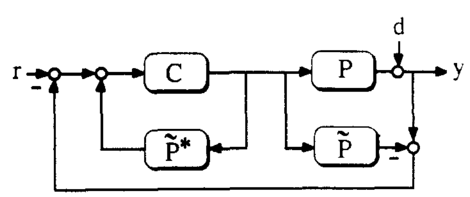

FIG. 3. Alternate representation of Smith Predictor controller.

† The relation between x and y is called *affine* when $y = A + Bx$.

feedback control law. Because of the finite horizon in (12) most conditions which guarantee that (12)–(14) will lead to a stabilizing controller are only sufficient. Thus, at this time the tuning of MPC has to proceed largely by trial and error with these sufficient conditions as guidelines.

For simplicity, the theorems below (García, 1982; García and Morari, 1982) are formulated for SISO systems without delays. Equivalent results for MIMO systems have been derived (García and Morari, 1985b) but are somewhat more complicated because specific non-singularity conditions have to be imposed.

Theorem 1. For $\Gamma_l \neq 0$, $B_l = 0$, selecting $m = p \leq n$ yields the model inverse control $Q(z) = z^{-1}P(z)^{-1}$.

This output deadbeat control law is only stable if all zeros of $P(z)$ are inside the unit circle. Even when it is stable it is generally very aggressive and can lead to intersample rippling if $P(z)$ has zeros close to $(-1, 0)$. This control law can be acceptable if the sampling time is relatively large.

Theorem 2. There exists a finite $B^* > 0$ such that for $B_l \geq B^*(l = 1, \ldots, m)$ the control law is stable for all $m \geq 1$, $p \geq 1$ and $\Gamma_l > 0$.

This theorem implies that penalizing control action can stabilize the system regardless of the other parameter choices.

Theorem 3. Assume $\Gamma_l = 1$, $B_l = 0$. Then for sufficiently small m and sufficiently large $p > n + m - 1$ the closed-loop system is stable.

Thus the horizon length p (relative to the number of manipulated variable moves m) plays a similar role as the input penalty parameter B_l.

Several other more specific results are available. Reid *et al.* (1979) derived the weighting matrices to yield a state deadbeat control law. Systems with a monotone discrete step response are analyzed by García and Morari (1982). See also the review by Clarke and Mohtadi (1987) for further stability conditions. It can also be easily shown that the control law (12)–(14) is of Type 1, i.e. no offset for step-like inputs (García and Morari, 1982).

Linear Quadratic Control (LQC) law computation

The objective of this section is to define a specific LQC problem and to relate it to the MPC problem (12)–(14). Let the process be described by

$$x(k+1) = Ax(k) + Bu(k) + \bar{w}_1(k) \tag{23}$$

$$d(k+1) = d(k) + \bar{w}_2(k) \tag{24}$$

$$y(k) = Cx(k) + d(k) + \bar{w}_3(k) \tag{25}$$

where $\bar{w}(k) = [\bar{w}_1(k), \bar{w}_2(k), \bar{w}_3(k)]^T$ forms a sequence of zero mean uncorrelated (in time) vector stochastic variables with the variance matrix $\bar{V}$. The disturbance d is a Wiener process. It can be pictured as a sequence of random steps the amplitudes of which are described by a normal distribution, and the time of occurrence follows a Poisson distribution (p. 56 of Chang (1961)). The deterministic equivalent of (24) is d = constant, as was assumed for MPC. It is only meaningful to define an optimal control problem for the system (23)–(25) if it is stabilizable and detectable.

$$\left\{\begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix}, \begin{bmatrix} B \\ 0 \end{bmatrix}\right\}$$

is clearly not stabilizable. By differencing (23)–(25) become (Prett and García, 1988)

$$\begin{bmatrix} \Delta x(k+1) \\ \bar{y}(k+1) \end{bmatrix} = \begin{bmatrix} A & 0 \\ CA & I \end{bmatrix}\begin{bmatrix} \Delta x(k) \\ \bar{y}(k) \end{bmatrix} + \begin{bmatrix} B \\ CB \end{bmatrix}\Delta u(k) + \begin{bmatrix} w_1(k) \\ w_2(k) \end{bmatrix} \tag{26}$$

$$y(k+1) = \bar{y}(k+1) + w_3(k+1) \tag{27}$$

where $w(k) = [w_1(k), w_2(k), w_3(k)]^T$ is again a sequence of zero mean vector stochastic variables the correlation functions of which can be derived from $\bar{V}$. $\bar{y}$ is an additional state. If we assume that $\bar{w}_1 = 0$ and $\bar{w}_3 = 0$ then $w(k)$ is again uncorrelated and the variance matrix of $\bar{w}(k)$ and $w(k)$ becomes

$$\bar{V} = V = \begin{bmatrix} 0 & 0 & 0 \\ 0 & V_0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{28}$$

We will also assume that the covariance matrix of the state estimate $(\Delta x, \bar{y})^T$ at the initial point is

$$V' = \begin{bmatrix} 0 & 0 \\ 0 & V_1 \end{bmatrix}. \tag{29}$$

Equations (28) and (29) imply that Δx is known perfectly all the time and that there is no measurement noise.

Controllability. System (26)

$$\left\{\begin{bmatrix} A & 0 \\ CA & I \end{bmatrix}, \begin{bmatrix} B \\ CB \end{bmatrix}\right\}$$

is controllable if and only if:

(1) $\{A, B\}$ is controllable;

(2) $\begin{bmatrix} A - I & B \\ CA & CB \end{bmatrix}$ has full row rank.

(The proof follows the arguments by Morari and Stephanopoulos (1980).) Condition 2 requires the number of manipulated variables to be at least as large as the number of controlled outputs. It can be interpreted further in the context of the steady-state form of (26)

$$\Delta y = [-CA(A - I)^{-1}B + CB]\Delta u. \tag{30}$$

If $\dim y = \dim u$ then according to Schurs' formula (Gantmacher, 1959) condition 2 becomes

$$\det\begin{bmatrix} A - I & B \\ CA & CB \end{bmatrix} = \det(A - I) \cdot \det(-CA(A - I)^{-1}B + CB) \neq 0. \tag{31}$$

Comparing (30) and (31) we note that controllability of (26) requires the steady-state gain matrix between

Δu and Δy to be nonsingular which is clearly a very reasonable requirement. In summary, conditions 1 and 2 are satisfied in any practical situation.

Detectability. System (26)

$$\left\{\begin{bmatrix} A & 0 \\ CA & I \end{bmatrix}, [0, I]\right\}$$

is detectable if and only if $\{A, CA\}$ is detectable. (This can be proved using the arguments by Morari and Stephanopoulos (1980).) Because A is stable, $\{A, CA\}$ is detectable.

Optimal control problem. Rewrite problem (12) for system (26)–(29) in a form similar to the one used on p. 539 of Kwakernaak and Sivan (1972)

$$\min_{\Delta u(k)\cdots\Delta u(k+p-1)} E\left\{\sum_{l=1}^{p} [y(k+l) - r(k+l)]^T R_3(l)[y(k+l) - r(k+l)] + \Delta u(k+l-1)^T R_2(l)\Delta u(k+l-1)\right\} \quad (32)$$

where $E\{\cdot\}$ is the expected value operator, $R_2 > 0$, $R_3 > 0$ and $R_2(l) = \infty$ for $l > m+1$.

The unique optimal solution is the feedback law

$$\Delta u(k) = F_1(k)\Delta\hat{x}(k) + F_2(k)[\hat{y}(k) - r(k)] \quad (33)$$

where the formula for $F(k)$ can be found on p. 494 of Kwakernaak and Sivan (1972) and $\Delta\hat{x}(k)$ and $\hat{y}(k)$ are the state estimates obtained from the optimal observer (p. 530 of Kwakernaak and Sivan (1972))

$$\Delta\hat{x}(k+1) = A\Delta\hat{x}(k) + B\Delta u(k) \quad (34)$$

$$\hat{y}(k+1) = CA\Delta\hat{x}(k) + y(k) + CB\Delta u(k). \quad (35)$$

Note that because of the specific noise assumptions the optimal observer is time invariant. According to (34) the states Δx are estimated in open-loop fashion from the process model. Equation (35) is identical with the corresponding part of (26) except that the measurement $y(k)$ appears instead of $\bar{y}(k)$: the estimate $\hat{y}(k+1)$ is the measurement $y(k)$ plus the effect of the manipulated variables as expressed through the model. In block diagram form the estimator (34) and (35) corresponds to a model in parallel with the plant as is characteristic of MPC.

Comparison of unconstrained MPC and LQC

(1) The control laws resulting from the two computational procedures are essentially equal. The only difference arises from the fact that the state space model and the truncated step response model are not identical (but can be made arbitrarily close to each other).

(2) The MPC computation requires the solution of a (possibly large) linear least squares problem. LQC involves the solution of some recursive matrix relations. The order of the system which determines the dimension of the matrices involved is generally much less than the truncation order of the step response model.

(3) For implementation it is desirable to use a time invariant control law. In the general LQC context time invariance is achieved by making the weights R_2 and R_3 constant and extending the horizon to infinity. Unconstrained MPC is time invariant because only the first control move obtained from the finite horizon problem is applied at each time step.

(4) The time invariant LQC is tuned by adjusting the constant weighting matrices R_2 and R_3. MPC can be tuned by adjusting R_2 and R_3 but a more frequently used tuning parameter is the number of moves m. The horizon p is usually selected such that $p \geq n$. There appears to be consensus in the literature that it is easier to pick a reasonable m a priori than R_2 and R_3.

(5) If the control problem is really of a stochastic nature, the implicit assumptions in MPC are quite restrictive (no measurement and state excitation noise) and might lead to performance which is inferior to what could be obtained by a full LQG approach. If—as is usually the case—the covariance matrices are *de facto* tuning parameters then the MPC performance should be no better or worse than LQG.

(6) It is well known that the LQG controller leads to a closed-loop stable system as the horizon is extended to infinity. A similar result can be proven for unconstrained MPC (García and Morari, 1982, 1985b).

(7) Both controllers reject sustained step-like disturbances, i.e. include implicitly integral action.

Internal Model Control (IMC)

The motivation behind the development of IMC was to combine the advantages of the different unconstrained MPC schemes and to avoid their disadvantages: easy on-line tuning via adjustment of physically meaningful parameters and without any concern about closed-loop stability; good performance without intersample rippling; ability to cope with inputs other than steps. IMC utilizes the controller structure shown in Fig. 2(a) (and (c)) and uses a filter for tuning in a similar manner as MAC. The design procedure which is described in detail by Morari and Zafiriou (1989) and Zafiriou and Morari (1986b) consists of two steps.

(1) First, $\hat{Q}$ is designed to yield a good response (usually in the least square sense) without regard for constraints on the manipulated variables and robustness.

(2) The IMC controller Q is found by augmenting $\hat{Q}$ with a low-pass filter $F(Q = \hat{Q}F)$ the parameters of which are adjusted either off-line or on-line to reduce the action of the manipulated variable and to improve the robustness (the controller $\hat{Q}$ is "detuned").

The controller $\hat{Q}_0(z)$ which minimizes $\sum_{k=0}^{\infty} e_k^2$ for a disturbance input d is derived by Zafiriou and Morari (1986b) (also see Morari and Zafiriou (1989))

$$\hat{Q}_0(z) = z(P_M(z)d_M(z))^{-1}\{(P_A(z)^{-1}d_M(z)z^{-1}\}_* \quad (36)$$

where the plant P is factored into an all pass P_A and a minimum phase part P_M. We have

$$P = P_A \cdot P_M \quad (37)$$

$$P_A(z) = z^{-N}\prod_{j=1}^{n}\frac{(1-\bar{\zeta}_j^{-1})(z-\zeta_j)}{(1-\zeta_j)(z-\bar{\zeta}_j^{-1})} \quad (38)$$

where $\zeta_j, j = 1, \ldots, n$, are the zeros of $P(z)$ outside the unit circle and the integer N is such that $z^N P(z)$ is semi-proper, i.e. its numerator and denominator polynomials have the same degree. The overbar denotes the complex conjugate. Similarly d is factored into d_A and d_M with $z^{N_d} d(z)$ semi-proper. Finally, the operator $\{\cdot\}_*$ in (36) implies that after a partial fraction expansion only the strictly proper stable (including poles at $z = 1$) terms are retained.

The H_2-optimal controller $\hat{Q}_0(z)$, however, may exhibit intersample rippling caused by poles of $\hat{Q}_0(z)$ close to $(-1, 0)$. A detailed study of the advantages and disadvantages and the theoretical reasons behind them, led Zafiriou and Morari (1985a) to a simple method for obtaining the IMC controller $\hat{Q}(z)$

$$\hat{Q}(z) = \hat{Q}_0(z) q_-(z) B(z) \tag{39}$$

where $q_-(z)B(z)$ replaces all the poles of $\hat{Q}_0(z)$ with negative real part with poles at the origin. The introduction of poles at the origin incorporates into the design some of the advantages of a deadbeat type response while at the same time known problems of deadbeat controllers, like overshoot, are avoided. Let $\pi_j, j = 1, \ldots, \rho$ be the poles of $\hat{Q}_0(z)$ with negative real part. Then

$$q_-(z) = z^{-\rho} \prod_{j=1}^{\rho} \frac{z - \pi_j}{1 - \pi_j} \tag{40}$$

$$B(z) = \sum_{j=0}^{m-1} b_j z^{-j} \tag{41}$$

where m is the "Type" of the input d and coefficients $b_j, j = 0, \ldots, m-1$ are computed so that $\hat{Q}(z)$ produces no steady-state offset (Zafiriou and Morari, 1986a,b). For offset-free tracking of steps and ramps we have:

Type 1

$$b_0 = 1; \tag{42}$$

Type 2

$$b_0 = 1 - b_1,\ b_1 = \sum_{j=1}^{\rho} \frac{\pi_j}{(1 - \pi_j)}. \tag{43}$$

Also the discrete filter $F(z)$ has to satisfy certain conditions for offset-free tracking. For steps and ramps we have:

Type 1

$$F(1) = 1; \tag{44}$$

Type 2

$$F'(1) = 0. \tag{45}$$

Commonly used filters are:

Type 1

$$F(z) = \frac{(1 - \alpha)z}{z - \alpha}; \tag{46}$$

Type 2

$$F(z) = \frac{(1 - \alpha)z}{z - \alpha} (\beta_0 + \beta_1 z^{-1} + \cdots + \beta_w z^{-w}) \tag{47}$$

with

$$\beta_0 = 1 - (\beta_1 + \cdots + \beta_w) \tag{48}$$

$$\beta_k = \frac{-6k\alpha}{(1 - \alpha)w(w + 1)(2w + 1)}, \quad k = 1, \ldots, w. \tag{49}$$

Any $w \geq 2$ will satisfy the no-offset property. However the higher the w, the closer (47) approximates (46).

Instead of the filter, a term penalizing input variations can be included in the quadratic objective function with a very similar effect (Morari and Scali, 1989). However, while the filter time constant α has direct physical significance (it is directly related to the closed-loop time constant/inverse bandwidth), the input action penalty term is rather artificial and the weight is difficult to select without trial and error using simulations. Also, any change in the penalty weight requires the optimal control problem to be resolved (linear least squares problem in DMC and MAC, Riccati equation or spectral factorization for LQC). A change in α, on the contrary, can be implemented directly without any computational effort. In their simplest form with filter (46) and (47) SISO IMC controllers are "one-knob controllers" with the knob corresponding to closed-loop bandwidth.

Conceptually, the two-step IMC design procedure extends to MIMO systems in a straightforward manner. In the first step the LQ optimal controller can be obtained from an MIMO version of (36). For rational plants the necessary all-pass factorization can be performed with standard software (Enns, 1984). The resulting controller does not necessarily lead to a decoupled response, which might be desired. Alternatively, Zafiriou and Morari (1985b, 1987) show how to design the controller $\hat{Q}$ such that the closed-loop transfer function has a specified structure.

There is much freedom for designing the filter. In the simplest case F is diagonal with one tuning parameter (the closed-loop time constant) for each output. Often this is sufficient for achieving satisfactory response and robustness characteristics. For ill-conditioned systems robust performance is sometimes difficult to obtain without off-diagonal adjustable filter elements. Zafiriou and Morari (1986a) have developed a gradient search procedure for the filter parameters in order to optimize robust performance. Analytic expressions for the gradient are derived. Though this method performed well on a few test examples it is potentially plagued by the nonconvexity and nondifferentiability of the objective function. The development of a more reliable technique is the subject of current research.

Comparison of IMC with DMC and MAC

(1) The structure inherent in all MPC schemes, referred to as IMC structure, corresponds to a very convenient way of parametrizing all stabilizing controllers for an open-loop stable plant. This makes the structure very useful for design: a stable MP controller implies a stable closed-loop system and vice versa.

(2) IMC reduces the number of adjustable parameters of DMC and MAC to a minimum and expands the role of the MAC filter. The controller can also be tailored to accommodate inputs different from steps in an optimal manner. For low-order models the

IMC design procedure generates PID controllers with one adjustable parameter.

(3) The advantage of unconstrained MPC (with the possible exception of IMC) over other LQ techniques is still to be demonstrated with more than case studies. Contrary to other techniques, however, the MPC ideas can be extended smoothly to generate non-linear time-varying controllers for linear systems with constraints.

CONSTRAINED MPC

As emphasized in the introductory sections of this paper, constraints are always present in any real life process control situation. Their importance has increased because supervisory optimizing control schemes frequently push the operating point toward the intersection of constraints (Arkun, 1978; Prett and Gillette, 1979). Most of today's control implementations handle constraints through split range controllers, overrides and more general min–max selectors with some logic. These schemes are difficult to design, debug, explain to the operating personnel and maintain. For an example of a mildly complicated scheme of this type the reader is referred to Bristol (1980). The main attraction of MPC is that the engineer/operator can enter the constraints in a direct manner and that the algorithm will automatically find the best solution satisfying *all* of them.

Structure

Recall that the IMC structure (Figs 2(a) and (c)) is effectively open loop when $P = \tilde{P}$ and that it is internally stable if and only if P and Q are stable. This result holds trivially even when P and Q are nonlinear or when the inputs to P are subject to saturation constraints as long as the *constrained* inputs are also fed to the model $\tilde{P}$ so that $P = \tilde{P}$ is preserved. While input saturation causes complex stability problems for the classic feedback structure (Figs 1(b) and (d)), it has no effect whatsoever on the stability of MPC. If input saturation is unlikely during normal operation and one is mainly concerned about stability under emergency conditions, the controller should simply be designed for the unconstrained system as discussed earlier in the paper and then *implemented* in the form suggested in Figs 2(a) and (c).

Because of the quasi open-loop structure, however, the unconstrained controller Q is unaware of the input constraints and the *performance* can suffer badly when the inputs saturate. Qualitatively, the inputs should be held longer at the constraints than what the controller in Figs 1(a) and (c) does, in order to compensate for the restricted control action. Specifically, Q should be designed with the constraints included explicitly in the MPC formulation. As discussed by Campo and Morari (1986) and summarized next there are a number of possibilities for defining an appropriate objective function. In general Q becomes nonlinear and time varying and no stability analysis procedure is available when the constraints are active.

2-*Norm objective function*

The objective function (12) employs the 2-norm spatially (for the output and input vectors at a particular time) and temporally (over the horizon length). Apparently, a QP of the type (12)–(15) has been included in MAC for quite a while (Mehra *et al.*, 1982) but the first detailed descriptions of solution procedures in the open literature were provided in the context of DMC by García and Morshedi (1984) and Ricker (1985).

The advantage of the 2-norm is that at least for the unconstrained case an explicit expression for the control law can be found and its stability, robustness and performance characteristics can be analyzed with standard tools. A disadvantage is that for MIMO systems with many future moves m and long horizon p storage requirements can be quite formidable.

1-*Norm objective function*

Instead of the 2-norm the spatial and temporal 1-norm can be used to express the objective

$$\min_{\Delta u(k)\cdots\Delta u(k+m-1)} \sum_{l=1}^{p}\sum_{i=1}^{r} w_{li}\,|\hat{y}_i(k+l\mid k) - r_i(k+l)| \qquad w_{li} \geq 0. \quad (50)$$

With the additional inequality constraints the resulting Linear Program (LP) is computationally simpler than the QP discussed previously. The long tradition of LP in optimal control was reviewed earlier in this paper. A 1-norm version of DMC was introduced by Morshedi *et al.* (1985).

A severe disadvantage of the 1-norm formulation is that even in the unconstrained case a simple fundamental analysis of stability and performance is not possible because an explicit closed-form expression for the (nonlinear) control law does not exist. Brosilow and co-workers (Brosilow *et al.*, 1984; Brosilow and Zhao, 1986) circumvent this problem in the following manner. They first design a controller for the unconstrained process by some standard procedure. Then they minimize the 1-norm of the error not between the process output and setpoint as suggested by (12) but between the predicted constrained output and the ideal unconstrained output. In this formulation a term for penalizing the control action is not necessary, which eliminates some tuning parameters. If the constraints are not active the optimal value for the objective function found by the LP is zero and the process output is equal to what would be obtained by the linear time invariant controller which was designed in the first step. Obviously when the constraints become active an analysis of the properties of the control algorithm becomes impossible. However, because the response is kept close to the response of the unconstrained system it is reasonable to expect that its properties (stability, robustness, etc.) remain preserved.

∞-*Norm objective function*

Campo and Morari (1986) adopt Brosilow's idea but use a different norm. The unconstrained controller is designed by the IMC procedure. Then the ∞-norm of the error between the predicted constrained and the ideal unconstrained process output is minimized

$$\min_{\Delta u(k)\cdots\Delta u(k+m-1)} \max_{\substack{l=1,P \\ i=1,r}} w_{li} \, |\hat{y}_i(k+l \mid k) - \hat{r}_i(k+l)|. \quad (51)$$

In practice, the ∞-norm is particularly meaningful when peak excursions from desired trajectories are to be avoided. The 1- and 2-norms tend to keep average deviations small but allow large peak deviations. Also, in general, the ∞-norm results in an LP which can be solved more efficiently than the 1-norm LP.

Summary

Because of our deep understanding of linear systems it is preferable to employ an algorithm which is linear and for which a closed-form expression can be derived and analyzed as long as the constraints are not active. QDMC and the formulation by Brosilow have these characteristics. The experience with different norms in conjunction with the Brosilow formulation has been too limited to date to draw any definite conclusions.

NON-LINEAR MPC

While we can deal with mild nonlinearities just by detuning linear controllers, it is likely that in the presence of strong nonlinearities, non-linear controllers offer distinct advantages.

Even though there any many important unresolved details the conceptual extension of the IMC structure to non-linear systems is straightforward. Nevertheless only Frank (1974) appears to have recognized its potential.

For non-linear systems the assumption of an unmeasured additive disturbance acting at the process output is usually artificial. Indeed, for non-linear systems the issues of model error (robustness) and unmeasured disturbances become indistinguishable.

In general, all blocks in Fig. 2(a) or (c) can be nonlinear. The process model is a simulation program where the non-linear differential equations are solved on-line in parallel with the process.

As shown by Economou *et al.* (1986) the stability properties of the IMC structure carry over to non-linear systems when the appropriate definitions are made. Indeed, as for linear systems, all stabilizing controllers for the non-linear plant P are generated from stable IMC controllers Q.

A general technique for the design of the non-linear controller Q is not available to date. Three attempts reported in the literature will be discussed next.

Non-linear optimal control

In analogy with (12)–(15) one can define the general non-linear problem

$$\min_u G[x(t_f)] + \int_{t_0}^{t_f} F[x(t), u(t)] \, dt \quad (52)$$

subject to

$$\dot{x} = f[x(t), u(t)], \; x(t_0) = x_0 \quad (53)$$

$$h(x, u) = 0 \quad (54)$$

$$g(x, u) \leq 0. \quad (55)$$

The solution to this problem when (54) is not present and $g(x, u)$ is a function of u only can be found in all classical references on optimal control (Athans and Falb, 1966; Lee and Markus, 1967). The variational methods become extremely complex when inequalities involving the states are present. Small examples are shown by Bryson *et al.* (1963) and Denham and Bryson (1964). Because of the computational complexity these methods are not suitable for on-line use.

It is more promising to discretize the control vector with respect to time and to convert (52)–(55) into a non-linear programming problem ("mathematical programming approach" (Bryson and Ho, 1975)). It is also possible to use a "black box" simulation model instead of (53) and to compute the gradients necessary for the mathematical program numerically. This has been done successfully for the on-line optimization of gas pipeline networks by Marqués (Marqués, 1985; Marqués and Morari, 1986).

Linearized optimal control

García (1984) linearized a problem similar to (52)–(55) and solved the linear problem by the DMC/QDMC approach. The linear model is updated as the state of the process changes and used to obtain the step response coefficients. The non-linear model is used to perform the model prediction. That is, the non-linear differential equations are integrated on-line, in parallel with the process, while the controller is a linear QDMC controller. The application to a batch reactor produced excellent results.

Inversion

The implicit objective of (12)–(14) is to make the system output y track the reference trajectory r, while keeping the manipulated variables u at a reasonable level. In the IMC context the penalty term on the inputs is omitted. Instead, r, is processed through the IMC filter so that r can be tracked exactly without putting too much strain on the inputs. Thus the control algorithm should determine u such that

$$Pu = r \quad (56)$$

where r is the filtered reference. In other words the right inverse of p is to be determined

$$u = P^{-1}r. \quad (57)$$

From this point of view the key questions for the design of non-linear MPC involve the existence, uniqueness and numerical construction of the inverse of the non-linear operator P. For specific static nonlinearities, saturation constraints and multiplicative nonlinearities the construction of (approximate) stable inverses is discussed by Frank (1974). Hirschorn (1979) derived conditions for the existence of the inverse of quite a general class of non-linear operators and also described an analytic procedure for their inversion. Though mathematically rigorous, the method implies the use of higher order derivatives and is therefore very sensitive to noise and/or numerical errors. It is unsuitable for on-line use.

Economou and co-workers (Economou *et al.*, 1986; Economou, 1986) experimented with various methods (contraction mapping, Newton's method) to solve the non-linear operator equation (56) for u. Simulation experiments were very successful. Indeed, stable performance of the non-linear control scheme was found under conditions where any linear controller would make the system unstable.

In an extension of this work, Economou (1986) and Economou and Morari (1985) applied the ideas of non-linear operator inversion directly to the design of feedback controllers.

ROBUSTNESS

By robustness we mean roughly that the quality of performance of the feedback system is preserved when the dynamic behavior of the real plant is different from that assumed in the model. The importance of the robustness problem has become recognized universally in the last decade and much progress has been made toward its systematic solution (at least for linear systems: very little is known about non-linear systems). A detailed review and discussion is beyond the scope of this paper. The reader is referred to Morari and Doyle (1986).

In its linear form, unconstrained MPC is equivalent to classic feedback and therefore all the developed robustness analysis procedures can be applied. Whether MPC is robust or not depends on how it is designed. MPC is not inherently more or less robust than classic feedback as has been falsely claimed (Mehra *et al.*, 1982). Also, the large number of parameters in a step or impulse response model as opposed to the three or four parameters in a parsimonious transfer function model does not add any robustness. By that argument one could include some parameters in the model which have no effect on the input–output description at all and obtain even more robustness.

On the other hand there is no doubt that MPC can be adjusted more easily for robustness than classic feedback. For example, the filter in MAC and IMC has a very direct effect on robustness (García and Morari, 1982, 1985a, b). This fact might be responsible for the misconceptions regarding MPC robustness.

CONCLUSIONS

Process control is driven by the need to respond competitively to a rapidly changing marketplace. Contrary to aerospace control the dynamics of the underlying system are simple (overdamped, no high frequency resonances). The difficulties arise from model uncertainty and the requirement to satisfy dynamically a number of constraints on states and inputs, which define the economically optimal operating condition. MPC is an attractive tool for addressing these issues. Tuning for robustness is direct and the constraints are considered explicitly in the algorithm. The future outputs predicted by the internal model could be displayed to the operator to give him confidence in the algorithm and to allow transparent on-line tuning.

Despite the progress during the last decade there are quite a few unresolved research issues.

• *Linear systems.* The tuning procedures necessary to achieve robust performance for ill-conditioned MIMO systems (for an example see the study of Skogestad and Morari (1986)) are very complex. We lack the physical understanding of the connections between uncertainty and performance to suggest simpler techniques.

• *Constraints.* The LP and QP algorithms which have been employed for MPC are basically "off-the-shelf" with minor modifications. For large-scale applications more efficient tailor made algorithms will be needed. In the constrained algorithm model uncertainty is neglected. Thus, there is no guarantee that the real process variables will satisfy the constraints when the model variables do. Also, when there are constraints on both the inputs and states it is possible that the LP/QP is unable to find a feasible solution when a disturbance pushes the process outside the usual operating region. It is unclear how to recover gracefully in these situations. Some ideas have been presented by Gutman (1982), García and Morshedi (1984) and Brosilow and Zhao (1986).

• *Non-linear systems.* The surface has barely been scratched. All important questions like nominal stability/model uncertainty and constraint handling are unanswered. The computational requirements are expected to be very high and the power of special computer architectures (e.g. hypercube) will have to be utilized.

Acknowledgement—Partial support from the National Science Foundation and the Department of Energy are gratefully acknowledged. The authors would like to thank E. Zafiriou for some helpful comments and criticism.

REFERENCES

Arkun, Y. (1978). Design of steady-state optimizing control structures for chemical processes. Ph.D. Thesis, University of Minnesota.

Arkun, Y., J. Hollett, W. M. Canney and M. Morari (1986). Experimental study of internal model control. *Ind. Engng Chem. Process Des. Dev.*, **25**, 102–108.

Athans, M. and P. L. Falb (1966). *Optimal Control: an Introduction to the Theory and its Applications.* McGraw-Hill, New York.

Bristol, E. H. (1980). After DDC. Idiomatic control. *Chem. Engng Prog.* **76**(11), 84–89.

Brosilow, C. B. (1979). The structure and design of Smith Predictors from the viewpoint of inferential control. *Proc. Joint Aut. Control Conf.*, Denver, Colorado.

Brosilow, C. B., G. Q. Zhao and K. C. Rao (1984). A linear programming approach to constrained multivariable control. *Proc. Am. Control Conf.*, San Diego, California, pp. 667–674.

Brosilow, C. B. and G. Q. Zhao (1986). A linear programming approach to constrained multivariable process control. *Adv. Control Dyn. Syst.*, **24**.

Bryson, A. E., W. F. Denham and S. E. Dreyfus (1963). Optimal programming problems with inequality constraints—I. Necessary conditions for extremal solutions. *AIAA J.*, **1**, 2544–2550.

Bryson, A. E. and Y. Ho (1975). *Applied Optimal Control.* Hemisphere, Washington, DC/Wiley, New York.

Caldwell, J. M. and G. D. Martin (1987). On-line analyzer predictive control. Sixth Annual Control Expo Conf., Rosemont, Illinois, 19–21 May.

Campo, P. J. and M. Morari (1986). ∞-Norm formulation of model predictive control problems. *Proc. Am. Control Conf.*, Seattle, Washington, pp. 339–343.

Chang, S. S. L. (1961). *Synthesis of Optimum Control Systems*. McGraw-Hill, New York.

Chang, T. S. and D. E. Seborg (1983). A linear programming approach to multivariable feedback control with inequality constraints. *Int. J. Control*, **37**, 583–597.

Clarke, D. W. and C. Mohtadi (1987). Properties of generalized predictive control. *Proc. 10th IFAC World Congress*, Munich, Germany, Vol. 10, pp. 63–74.

Clarke, D. W., C. Mohtadi and P. S. Tuffs (1987a). Generalized predictive control—I. The basic algorithm. *Automatica*, **23**, 137–148.

Clarke, D. W., C. Mohtadi and P. S. Tuffs (1987b). Generalized predictive control—II. Extensions and interpretations. *Automatica*, **23**, 149–160.

Cutler, C. R. and B. L. Ramaker (1979). Dynamic matrix control—a computer control algorithm. AIChE National Mtg, Houston, Texas (1979); also *Proc. Joint Aut. Control Conf.*, San Francisco, California (1980).

Cutler, C. R. and R. B. Hawkins (1987). Constrained multivariable control of a hydrocracker reactor. *Proc. Am. Control Conf.*, Minneapolis, Minnesota, pp. 1014–1020.

DeKeyser, R. M. C. and A. R. van Cauwenberghe (1982). Typical application possibilities for self-tuning predictive control. IFAC Symp. on Identification and System Parameter Estimation, Washington, DC, Vol. 2, pp. 1552–1557.

DeKeyser, R. M. C. and A. R. van Cauwenberghe (1985). Extended prediction self-adaptive control. IFAC Symp. on Identification and System Parameter Estimation, York, U.K., pp. 1255–1260.

DeKeyser, R. M. C., Ph. G. A. VandeVelde and F. A. G. Dumoriter (1985). A comparative study of self-adaptive long-range predictive control methods. IFAC Symp. on Identification and System Parameter Estimation, York, U.K., pp. 1317–1322.

Denham, W. F. and A. E. Bryson (1964). Programming problems with inequality constraints—II. Solution by the steepest ascent. *AIAA J.*, **2**, 25–26.

Economou, C. G. and M. Morari (1985). Newton control laws for nonlinear controller design. *Proc. IEEE Conf. on Decision and Control*. Fort Lauderdale, Florida, pp. 1361–1366.

Economou, C. G. (1986). An operator theory approach to nonlinear controller design. Ph.D. Thesis, California Institute of Technology.

Economou, C. G., M. Morari and B. Palsson (1986). Internal model control—5. Extension to nonlinear systems. *Ind. Engng Chem. Process Des. Dev.*, **25**, 403–411.

Enns, D. (1984). Model reduction for control system design. Ph.D. Thesis, Stanford University.

Frank, P. M. (1974). *Entwurf von Regelkreisen mit vorgeschriebenem Verhalten*. G. Braun, Karlsruhe (1974).

Gantmacher, F. R. (1959). *The Theory of Matrices*, Vol. I. Chelsea, New York.

García, C. E. (1982). Studies in optimizing and regulatory control of chemical processing systems. Ph.D. Thesis, University of Wisconsin, Madison.

García, C. E. and M. Morari (1982). Internal model control—1. A unifying review and some new results. *Ind. Engng Chem. Process Des. Dev.*, **21**, 308–323.

García, C. E. (1984). Quadratic dynamic matrix control of nonlinear processes. An application to a batch reaction process. AIChE Annual Mtg, San Francisco, California.

García, C. E. and A. M. Morshedi (1984). Quadratic programming solution of dynamic matrix control (QDMC). *Proc. Am. Control Conf.*, San Diego, California (1984); also *Chem. Engng Commun.*, **46**, 73–87 (1986).

García, C. E. and M. Morari (1985a). Internal model control—2. Design procedure for multivariable systems. *Ind. Engng Chem. Process Des. Dev.*, **24**, 472–484.

García, C. E. and M. Morari (1985b). Internal model control—3. Multivariable control law computation and tuning guidelines. *Ind. Engng Chem. Process Des. Dev.*, **24**, 484–494.

García, C. E. and D. M. Prett (1986). Advances in industrial model—predictive control. In M. Morari and T. J. McAvoy (Eds), *Chemical Process Control—CPC III*, pp. 249–294. CACHE and Elsevier, Amsterdam.

Giloi, W. (1959). Zur Theorie und Verwirklichung einer Regelung für Laufzeitstreken nach dem Prinzip der ergänzenden Rückführung. Ph.D. Thesis, TH Stuttgart.

Grosdidier, P. (1987). Setpoint, Inc., Houston, Personal communication.

Gutman, P. O. (1982). Controllers for bilinear and constrained linear systems. Ph.D. Thesis, Lund Institute of Technology.

Hirschorn, R. M. (1979). Invertibility of Multivariable nonlinear control systems. *IEEE Trans. Aut. Control*, **24**, 855–865.

Horowitz, I. M. (1963). *Synthesis of Feedback Systems*. Academic Press, London.

Kwakernaak, H. and R. Sivan (1972). *Linear Optimal Control Systems*. Wiley-Interscience, New York.

Lee, E. B. and L. Markus (1967). *Foundations of Optimal Control Theory*. Wiley, New York.

Levien, K. L. (1985). Studies in the design and control of coupled distillation columns. Ph.D. Thesis, University of Wisconsin, Madison.

Levien, K. L. and M. Morari (1987). Internal model control of coupled distillation columns. *A.I.Ch.E. Jl*, **33**, 83.

Marqués, D. (1985). On-line optimization of large scale dynamic systems. Ph.D. Thesis, University of Wisconsin, Madison.

Marqués, D. and M. Morari (1986). Model predictive control of gas pipeline networks. *Proc. Am. Control Conf.*, Seattle, Washington, pp. 349–354.

Martin, G. D., J. M. Caldwell and T. E. Ayral (1986). Predictive control applications for the petroleum refining industry. *Japan Petroleum Institute—Petroleum Refining Conf.*, Tokyo, Japan, 27–28 October.

Matsko, T. N. (1985). Internal model control for chemical recovery. *Chem. Engng Prog.*, **81**(12), 46–51.

Mehra, R. K. and R. Rouhani (1980). Theoretical considerations on model algorithmic control for nonminimum phase systems. *Proc. Joint Aut. Control Conf.*, San Francisco, California.

Mehra, R. K., R. Rouhani, J. Eterno, J. Richalet and A. Rault (1982). Model algorithmic control: review and recent development. Engng Foundation Conf. on Chemical Process Control II, Sea Island, Georgia, pp. 287–310.

Morari, M. and G. Stephanopoulos (1980). Minimizing unobservability in inferential control schemes. *Int. J. Control*, **31**, 367–377.

Morari, M. and J. C. Doyle (1986). A unifying framework for control system design under uncertainty and its implications for chemical process control. In M. Morari and T. J. McAvoy (Eds), *Chemical Process Control—CPC III*, pp. 5–52. CACHE and Elsevier, Amsterdam.

Morari, M. and C. Scali (1989). A comparison of internal model control and linear quadratic optimal control for SISO systems. *Ind. Engng Chem. Res.* (submitted).

Morari, M. and E. Zafiriou (1989). *Robust Process Control*. Prentice-Hall, Englewood Cliffs, New Jersey.

Morshedi, A. M., C. R. Cutler and T. A. Skrovanek (1985). Optimal solution of dynamic matrix control with linear programming techniques (LDMC). *Proc. Am. Control Conf.*, Boston, Massachusetts, pp. 199–208.

Newton, G. C., L. A. Gould and J. F. Kaiser (1957). *Analytical Design of Feedback Controls*. Wiley, New York.

Parrish, J. R. and C. B. Brosilow (1985). Inferential control applications. *Automatica*, **21**, 527–538.

Peterka, V. (1984). Predictor-based self-tuning control. *Automatica*, **20**, 39–50.

Prett, D. M. and R. D. Gillette (1979). Optimization and constrained multivariable control of a catalytic cracking unit. AIChE National Mtg, Houston, Texas; also *Proc. Joint Aut. Control Conf.*, San Francisco, California.

Prett, D. M. and C. E. García (1988). *Fundamental Process Control*. Butterworths, Stoneham, Massachusetts.

Propoi, A. I. (1963). Use of LP methods for synthesizing sampled-data automatic systems. *Automn Remote Control*, **24**, 837.

Reid, J. G., R. K. Mehra and E. Kirkwood (1979). Robustness properties of output predictive dead-beat control: SISO case. *Proc. IEEE Conf. on Decision and Control*, Fort Lauderdale, Florida.

Richalet, J. A., A. Rault, J. L. Testud and J. Papon (1978). Model predictive heuristic control: applications to an industrial process. *Automatica*, **14**, 413–428.

Ricker, N. L. (1985). The use of quadratic programming for constrained internal model control. *Ind. Engng Chem. Process Des. Dev.*, **24**, 925–936.

Skogestad, S. and M. Morari (1986). Control of ill-conditioned plants: high purity distillation. AIChE Annual Mtg, Miami Beach. Florida.

Smith, O. J. M. (1957). Closer control of loops with dead time. *Chem. Engng Prog.*, **53**(5), 217–219.

Vidyasagar, M. (1985). *Control System Synthesis—a Factorization Approach*. MIT, Cambridge, Massachusetts.

Ydstie, B. E. (1984). Extended horizon adaptive control. 9th World Congress of the IFAC, Budapest, Hungary.

Youla, D. C., J. J. Bongiorno and H. A. Jabr (1976a). Modern Wiener–Hopf design of optimal controllers—I. The single-input-output case. *IEEE Trans. Aut. Control*, **AC-21**, 3–13.

Youla, D. C., H. A. Jabr and J. J. Bongiorno (1976b). Modern Wiener–Hopf design of optimal controllers—II. The multivariable case. *IEEE Trans. Aut. Control*, **AC-21**, 319–338.

Zadeh, L. A. and B. H. Whalen (1962). On optimal control and linear programming. *IRE Trans. Aut. Control*, **7**(4), 45.

Zafiriou, E. and M. Morari (1985a). Digital controllers for SISO systems. A review and a new algorithm. *Int. J. Control*, **42**, 855–876.

Zafiriou, E. and M. Morari (1985b). Robust digital controller design for multivariable systems. AIChE National Mtg, Chicago, Illinois.

Zafiriou, E. and M. Morari (1986a). Design of the IMC filter by using the structured singular value approach. *Proc. Am. Control Conf.*, Seattle, Washington, pp. 1–6.

Zafiriou, E. and M. Morari (1986b). Design of robust controllers and sampling time selection for SISO systems. *Int. J. Control*, **44**, 711–735.

Zafiriou, E. and M. Morari (1987). Digital controller design for multivariable systems with structural closed-loop performance specifications. *Int. J. Control*, **46**, 2087–2111.

Zames, G. (1981). Feedback and optimal sensitivity: model reference transformations, multiplicative seminorms, and approximate inverses. *IEEE Trans. Aut. Control*, **AC-26**, 301–320.

Zirwas, H. C. (1958). Die ergänzende Rückführung als Mittel zur schnellen Regelung von Regelstrecken mit Laufzeit. Ph.D. Thesis, TH Stuttgart.