

STATE-SPACE PREDICTIVE CONTROL

JENS G. BALCHEN, DAG LJUNGQUIST and STIG STRAND

Division of Engineering Cybernetics, The Norwegian Institute of Technology, 7034 Trondheim, Norway

(Received for publication 6 August 1991)

Abstract—This paper deals with a predictive control strategy based on state-space models. Important issues concerning inherent model identification and optimal control computation are briefly discussed. Predictive control relies heavily on a model with satisfactory predictive capabilities. An off-line identification procedure must be accomplished to obtain a proper model structure and a parameter set, which is required for on-line adjustment. The control calculation is based on a general performance index and parameterization of the control variables in a nonlinear model, which includes the relevant constraints. This results in a finite-dimensional optimization problem which can be repetitively solved on-line. Simulation studies on two very different, typical industrial processes are presented. The simulations show that this MPC technique offers a major improvement in the control of many industrial processes.

INTRODUCTION

Model-based predictive control methods have been the subject of intensive research for about 10–15 years. However, these methods have much longer history. Model-based prediction was one of the main concepts introduced by Wiener (1942) nearly 50 years ago. The Pontryagin maximum principle developed in the mid-1950s (Pontryagin *et al.*, 1962) is an elegant mathematical solution to model-based predictive control. Nevertheless industrial applications have come slowly, mainly because the required computing capacity was not available at an acceptable price until about 15 years ago. This enabled established theoretical solutions to be given practical realizations even though many of the basic results were “reinvented” by contributors who had not studied the available literature too thoroughly.

It is a widely-accepted fact that only the *future* behavior of a process can be *controlled*. The *present state* of a system is the end of the *past*. Future behaviour will be governed by three dominating phenomena:

- the autonomous system behavior
- future independent external disturbances
- future control actions.

System responses belonging to the first category can be predicted using a dynamic model of the system and the more precise the model is, the better the predictions. Predicting future disturbances is more difficult and can be done in two ways: either by direct measurement of a phenomenon, which produces the disturbance after some delay; or by estimating the disturbance based upon a model of its basic phenomena. An example of the measurement of a phenomenon which is delayed is a transportation lag (in pipeline or transportation facility) bringing material into a process from a remote location where a measurement is made. Another type of predictable disturbance is a cyclic one which is relatively easy to estimate using a model of the cyclic process. Future control

actions present no difficulty since they are being generated by the control system itself.

When designing model-predictive control systems, an important issue is related to the choice of model representation. Some of the available products from control vendors utilize input-output models in the form of step-response matrices, impulse-response matrices etc. (Richalet *et al.*, 1978; Cutler and Ramaker, 1979). Since these are linear descriptions, the unconstrained predictive control then will be a linear approximation. In fact it can easily be shown that a predictive control system based upon a linear model and a least-squares predictive control towards a prescribed trajectory is totally equivalent to LQ control (linear/quadratic-optimal control). This matter is further discussed by Morari and Lee (1991), and in the papers referred there.

When a process is highly nonlinear and both control and state variables are subject to constraints, a state-space description is preferable for the system model. Most dynamic processes can be modelled in state-space, and the number of parameters necessary to describe a certain set of phenomena, can become much smaller when this type of description is used rather than an input-output description. Furthermore, when a state-space model is employed in a general dynamic optimization scheme, it is possible to determine the future control actions which are optimal in terms of any performance criterion. In process control the objective will most often be maximization of the net profit. Moreover the close relationship between the model and the real plant makes a state-space predictive control system conceptually simple and appealing to process operators because the predicted *physical quantities* corresponding to the computed control variables can be displayed to them.

There is a growing interest in the use of nonlinear state-space models in model-predictive control. Peterson *et al.* (1989) and Hernández and Arkun (1990) propose to use such models to calculate improved step responses when DMC is applied to pro-

cesses with highly nonlinear dynamics. Schmid and Biegler (1990) demonstrate the advantage of using a multistep Newton-type controller rather than QDMC in controlling a simulated FCC unit. This algorithm can be regarded as one of the many possible methods for solving nonlinear optimal control problems. The application of nonlinear state-space models in predictive control has been proposed also by Jang *et al.* (1987), Joseph *et al.* (1989), Sistu and Bequette (1990), Balchen *et al.* (1988, 1989) and Patwardhan *et al.* (1988).

A model-based predictive control method will now be outlined which has the above properties. Its application will be demonstrated on two very different processes:

- the FCC for the production of light hydrocarbons from a heavy feed stock and
- the production of a light metal by electrolysis in a large number of units connected in series.

The first of these processes is a highly nonlinear process with very strong cross-couplings; thus a constant ordinary control algorithm does not perform satisfactorily. The latter is highly constrained, very interconnected, nonlinear and with partly discontinuous control actions. Ordinary linear control-theory based systems have little to offer for this process.

The following presentation starts with a description of the SSPC scheme, which highlights the order of the involved computations. The model identification and dynamic optimization issues are very essential for the performance of the proposed controller, and have been focused in two doctoral theses (Ljungquist, 1990; Strand, 1991). These parts will be elaborated in some detail in the subsequent sections of this paper. Then the two simulation studies with some illustrating results will be presented.

CONTROLLER DESCRIPTION

The goal of this section is to provide a description of the main functions performed by the SSPC system of Fig. 1. Details of the model identification and inherent open-loop optimal control issues are discussed in later sections.

The moving horizon concept illustrated in Fig. 2 is common to all predictive controllers. Figure 2 also shows some features which are specific for the proposed SSPC system. It is supposed that the time δt is needed to solve the open-loop optimal control problem. This can be longer than the sample time τ of the process control system. The relationship $\delta t = 3\tau$ is indicated in Fig. 2. An efficient way to solve the open-loop optimal control problem is to parameterize the control variables, for instance as a piecewise linear function as shown in Fig. 2. This means that quite few parameters are needed to get a satisfactory resolution of the control variables. The implemented control can naturally be piecewise constant (discretized) if this is a convenient representation.

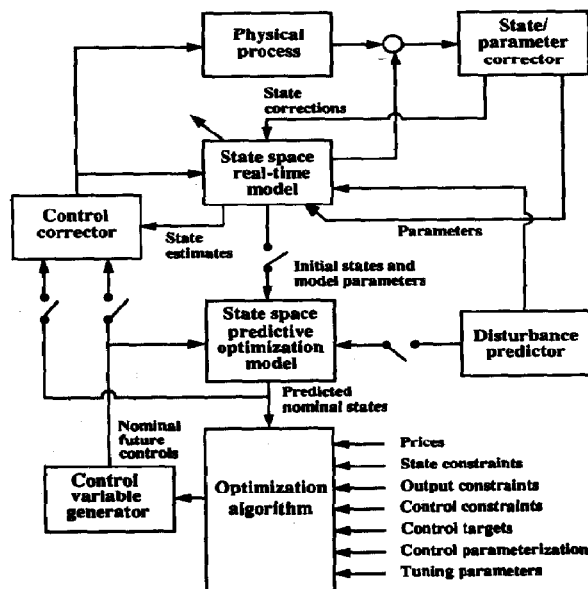


Fig. 1. Schematic SSPC overview.

If the model provides a very good description of the real process in the frequency domain of interest, the control variables that result from the dynamic optimization can be used unmodified, or discretized as in Fig. 2. This is also true if the time interval δt is very small compared to the desired closed-loop response times, either due to a slow process or high computing capacity. In the general case, however, it is preferable to include feedback around the optimized state trajectories. This is expressed by a feedback matrix G in the following presentation.

Let $M_e(t; t')$ be defined as the estimated (real-time) state-space model at time t' , while $M_p(t; t')$ is the corresponding predictive optimization model. Available disturbance predictions are included in these models. It may sometimes be advantageous to let M_p have a simpler structure than M_e , when this can speed up the optimization computations so that δt can be reduced. Both models can be used to predict dynamic responses due to manipulation of control variables when initial state variables have been specified.

With reference to Fig. 2, suppose that the nominal control and state trajectories $u^0(t; t_0)$ and $x^0(t; t_0)$ are available at real time t_0 . They are the solutions of the open-loop optimal control problem with initial horizon time t_0 .

Computation of nominal solutions from future time $t_0 + \delta t$

The following functions are performed during the real-time interval $(t_0; t_0 + \delta t)$ in order to have solutions $u^0(t; t_0 + \delta t)$ and $x^0(t; t_0 + \delta t)$ available at real time $t_0 + \delta t$:

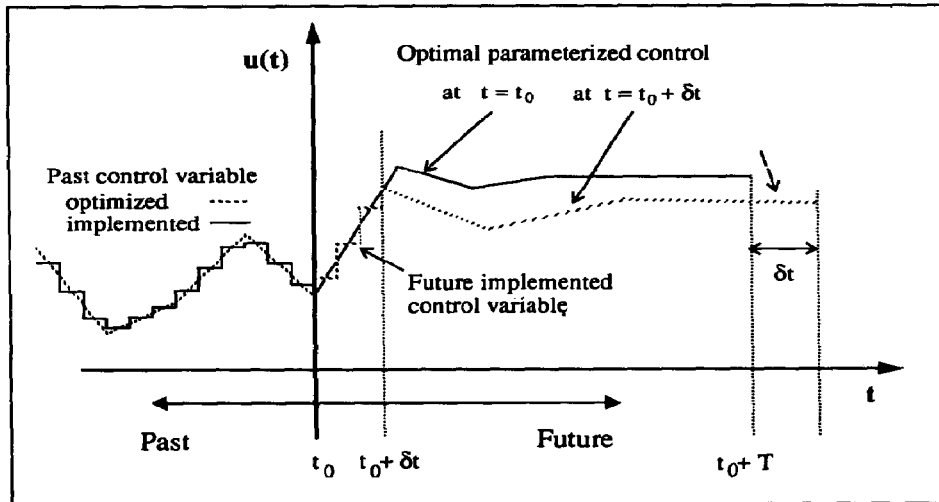


Fig. 2. Control calculation in SSPC.

- Update the real-time model at time t_0 , by use of the available process measurements. This gives $M_e(t; t_0)$ and the state estimate $\hat{x}(t_0)$.
- Predict the initial state variables $\bar{x}(t_0 + \delta t; t_0)$ for the subsequent optimization:

$$\hat{x}(t_0), M_e(t; t_0), \bar{u}_d(t; t_0) \rightarrow \bar{x}(t_0 + \delta t; t_0) \quad (1)$$

$$\bar{u}_d(t; t_0) = u_d^0(t; t_0) + G[\bar{x}(t; t_0) - x^0(t; t_0)]. \quad (2)$$

$u_d^0(t; t_0)$ is the discretized version of $u^0(t; t_0)$, and eq. (2) is calculated at discrete time instants corresponding to the future sampling instants of the process control system.

The feedback matrix is used in the initial state prediction because it will also be used in the control of the real process for the same time interval. It has an effect because the model M_e has been updated (three times in Fig. 2) since the computation of $x^0(t; t_0)$.

- Solve the open-loop optimal control problem with initial horizon time $t_0 + \delta t$:

$$\bar{x}(t_0 + \delta t; t_0), M_p(t; t_0) \rightarrow u^0(t; t_0 + \delta t), x^0(t; t_0 + \delta t) \quad (3)$$

$$u^0(t; t_0 + \delta t), M_e(t; t_0) \rightarrow x_e^0(t; t_0 + \delta t). \quad (4)$$

The dynamic optimization is performed in eq. (3). The function of eq. (4) is not needed if the models M_p and M_e are identical. If they are structurally different it may happen that the mismatch between the state responses in eqs (3) and (4) is so large that the present structure of M_p should be reconsidered when process conditions change.

Control of the process during the real-time interval ($t_0, t_0 + \delta t$)

The operations listed above must be finished or interrupted at real time $t_0 + \delta t$. In the meantime, the process is controlled by the following operations per-

formed at each sampling instant $t_0 + i\tau$, $i = 0, 1, 2$ (when $\delta t = 3\tau$).

- Update the real-time model by use of the process measurements, which gives.

$$M_e(t; t_0 + i\tau) \text{ and } \hat{x}(t_0 + i\tau).$$

- Implement the discretized control $u_d(t_0 + i\tau; t_0)$ according to

$$u_d(t_0 + i\tau; t_0) = u_d^0(t_0 + i\tau; t_0) + G[\hat{x}(t_0 + i\tau) - x^0(t_0 + i\tau; t_0)]. \quad (5)$$

Comments on the control calculations

The computation of the feedback matrix G in eqs (2) and (5) is based on the available model and the precomputed nominal trajectories. It is also possible to adapt the structure to the actual active constraints. This matter is further discussed by Strand (1991).

The usual practice in process control is to measure the plant response and then calculate the control actions, giving a *dead time in the feedback loop*. When the involved computational load is modest, as for instance when PID-controllers are used, this dead time will usually not contribute significantly to the negative phase angle. However, the control calculations in the dynamic optimization part of the SSPC scheme are so time consuming that the controller dead time would degrade the control quality. Starting the optimization at a time interval δt in the future in order to be finished when real time has advanced that interval, does not remove the problem. However, the better the model, the less the error in the initial states prediction.

The nominal control actions and responses for the time interval $t_0 + \delta t$ to $t_0 + 2\delta t$ are calculated by the model that is available at time t_0 . In the worst case, this leads to a time delay of $2\delta t$ from an abrupt

process change till this knowledge is included in the optimization. This indicates why a control correction with faster sampling rate is often necessary. If a *severe process change* is recognized, it will probably be advantageous to start a new optimization immediately, using old nominal control actions with correction till the nominal controls based on the fresh knowledge are available. This gives a maximum time delay of δt for severe changes to be accounted for by the nominal controls.

MODEL IDENTIFICATION

MPC techniques rely on a model which has satisfactory predictive capabilities, defining the goal of the identification procedure. This procedure has to be divided into two separate parts; off-line and on-line identification. The following subjects are dealt with in off-line experiments:

- choice of model set and model structure;
- determination of reasonable model parameters;
- definition of a parameter set to be adjusted on-line;
- choice of an identification algorithm to be used for the on-line identification.

As already mentioned the most suitable process description to be used in an MPC scheme is the nonlinear state-space model. The state-space model is established using physical knowledge, but allows for inclusion of input/output models of poorly known process parts. Disturbances may be modelled based on physical insight and represented in the model by state variables, or they may be represented by time-varying parameters.

Finding a set of parameters which makes the model fit the process measurements, is often referred to as an "inverse problem". However, in most cases this is an ill-posed problem because the number of parameters is usually too large to be uniquely given from the process measurements. One of the important features of a state-space model developed from "first principles" is that most of the model parameters are either given, or reasonable values can be found from physical considerations. The number of unknown parameters is then reduced to a minimum. Once the set of unknown parameters is decided, off-line experiments can be designed to maximize the identifiability of these parameters. It is beyond the scope of this paper to give details on this subject. Such work is available elsewhere [e.g. Ljung (1987)].

When the model structure and the set of unknown parameters are given, the best model within this model set is usually defined by the parameter combination which minimizes some norm of the difference between the process and model measurements. A quite general formulation of this minimization problem which is referred to as an *estimation problem*, is:

$$\hat{\theta}_N = \arg \min_{\theta \in D_M} V_N(\theta) = \arg \min_{\theta \in D_M} \frac{1}{N} \sum_{k=1}^N l[\varepsilon(k)] \quad (6)$$

where

$$\begin{aligned} \varepsilon(k) &= y(k) - y_m(k) \\ y_m(k) &= g[\bar{x}(k, \hat{\theta}), \hat{\theta}] \\ \bar{x}(k, \hat{\theta}) &= \Phi[\hat{x}(k-1), \hat{\theta}] \\ \hat{x}(k) &= \bar{x}(k) + K_x(k)\varepsilon(k) \end{aligned}$$

and

$\hat{\theta}$	vector of parameter estimates, dimension p
$\hat{\theta}_N$	optimal vector of parameter estimates, dimension p
D_M	connected, open subset of \mathbf{R}^p
$V_N(\theta)$	error norm/minimization criterion
N	number of samples used in the minimization
$l(\cdot)$	a scalar valued positive function
$\varepsilon(k)$	vector of prediction errors, dimension m
k	discrete time
$y(k)$	vector of process measurements, dimension m
$y_m(k)$	vector of model measurements, dimension m
$g(\cdot)$	vector of nonlinear measurement functions, dimension m
$\bar{x}(k)$	vector of <i>a priori</i> state variable estimates, dimension n
$\Phi(\cdot)$	vector of discrete state propagation functions, dimension n
$\hat{x}(k)$	vector of <i>a posteriori</i> state variable estimates, dimension n
$K_x(k)$	state update matrix.

The control variables are suppressed in the expression above. It should be noted that the discrete state propagation vector $\Phi(\cdot)$ does not have to be computed analytically. Application of first principles will usually result in a nonlinear state-space model of the form:

$$\frac{dx}{dt} = f(x, \theta, u, t) \quad (7)$$

where, $f(\cdot)$ is a vector of nonlinear functions and u is the control vector. This continuous model is simulated from $k-1$ to k using an ordinary differential equation (ODE) solver. The discrete form in eq. (6) is used because most process measurements are discrete by nature.

Clearly, the number of samples used in eq. (6) will influence the maximum number of parameters which can be identified. Moreover the time variation of the process inputs, i.e. the control variables and the measurable disturbances, are of vital importance. In fact the control variables should be designed to maximize the identifiability of the model parameters in the off-line experiments.

The quadratic error norm given by

$$l(\varepsilon) = \varepsilon^T \hat{E}^{-1} \varepsilon \quad (8)$$

where \hat{E} is an estimate of the covariance matrix of the prediction error ε , is frequently used in eq. (6). Then there is a close relationship between the estimate from eq. (6) and the maximum likelihood estimate (MLE). This is a desired property since the parameter covariance matrix then can be estimated from the inverse Hessian of the minimization problem (6). Furthermore, there is obviously a close relationship between the condition number of the Hessian and the identifiability of the parameter set. It can be shown that the parameter combination corresponding to the smallest eigenvalue of the Hessian matrix in a given experiment gives the parameter combination which is most difficult to identify (Ljungquist, 1990). This information can be used to pick out the parameter set which is to be estimated on-line as will be illustrated in the FCC example.

In the off-line identification procedure, different model structures, minimization techniques, criteria etc. are tried out to find a satisfactory model. The minimization techniques used to find a set of constant parameters over the time horizon of N samples, may be similar to those used in optimal control computation, discussed in the next section. It may be advantageous to set the state update matrix equal to zero in off-line experiments in order to improve the predictive capabilities of the model. When it comes to on-line estimation, the parameters must be allowed to change in order to keep the prediction errors within acceptable bounds, and the state update matrix should be included to increase the robustness of the algorithm. An estimation scheme as defined in eq. (6) may be implemented on line with a moving time horizon of N samples, but this strategy is not yet implemented in practical processes due to the high optimization time consumption. Instead algorithms similar to that in eq. (6) are formulated recursively in order to update the state and parameters using just one measurement sample each time. In addition to the inherent lack of robustness in recursive computation, recursive algorithms are sensitive to measurement noise because they are based on quadratic criteria (Poljak and Tsytkin, 1980). However, recursive algorithms may perform well in some applications. Because of their simplicity at least one recursive algorithm should be tried out in order to justify the implementation of a more advanced estimation scheme. The computing capacity available on modern computers allows for on-line implementation of an estimation scheme based on eq. (6) with a robust criterion in many industrial processes. It is believed that such a strategy will be applied within a few years.

In order to solve the problem of simultaneous estimation of both states and unknown parameters recursively, the state vector can be augmented with the unknown parameters and the extended Kalman filter used to solve the resulting nonlinear filtering problem.

This strategy is often referred to as the augmented Kalman filter (AKF). However, the assumptions leading to optimal filters are usually not valid in the nonlinear case, which means that the AKF may diverge. More accurate nonlinear filters, such as the iterated AKF and the truncated second-order filter (Jazwinski, 1970; Gelb, 1974), have been successfully implemented to solve nonlinear filtering problems, but still poor convergence properties may arise when the noise characteristics are complex and unknown.

Using the innovations state-space representation, Ljung (1979) derived a Newton-type stochastic gradient algorithm for the minimization of eq. (6) and showed that the resulting algorithm, the recursive prediction error method (RPEM), has improved convergence properties compared to the AKF in the case of a linear model. However the algorithm suffers from the fact that simultaneous determination of both model parameters and the parameterized state update matrix generally leads to an overparameterized minimization problem. Because both the sensitivity matrix and the optimal state update gain may change drastically if either the parameters or the control variables change in nonlinear models, the parameter update direction computed from the sensitivity matrix estimate in the RPEM may be nonprofitable. Then the RPEM will generally diverge. Simulation results indicate that the AKF is more robust than the RPEM when it comes to simultaneous state and parameter estimation on nonlinear state-space models.

In the simulation of the FCC example, the AKF implemented in continuous discrete form is used for the combined parameter and state estimation. It is important to notice the similarity between the model states and the model parameters in the AKF formulation. As the information which can be extracted from the measurements of an industrial process under continuous operation is usually restricted, the dimension of the augmented state vector should be kept to a minimum. Sensitivity analysis (Gelb, 1974) can be a useful tool, but a model reduction procedure should be based on physical considerations. The experience gained from the FCC implementation showed that the following points are important:

- the dimension of the augmented state variable vector;
- the scaling of the state variables and the parameters;
- a numerically robust algorithm in the covariance updating equations (Bierman, 1977) and a robust ODE solver (variable step size/stiff solvers);
- well-tuned user chosen covariance matrices.

Further details on the practical implementation of AKFs can be found in Maybeck (1979, 1982).

OPEN-LOOP OPTIMAL CONTROL COMPUTATION

The open-loop optimal control problem to be repetitively solved in the SSPC strategy can be formally

written as

$$\min_{u(t)} J = \int_{T_p} L[x(t), u(t), t] dt \quad (9)$$

such that

$$\frac{dx}{dt} = f[x(t), u(t), \bar{v}(t)] \quad t \in T_p$$

$x(t_i)$ known by prediction

$$u(t) \in [u_l, u_u] \quad t \in T_p$$

$$h[x(t), u(t), t] \leq 0 \quad t \in T_p$$

$$\psi[x(t_f)] \leq 0$$

$$t_i = t_0 + \delta t$$

$$t_f = t_i + T$$

$$T_p = \{t | t_i \leq t \leq t_f\}$$

where

L	cost function, control objective
J	performance index, performance criterion, objective functional
T_p	optimization time interval
x	vector of state variables, dimension n
u	vector of control variables, dimension r
t	time
\bar{v}	disturbance predictions
u_l, u_u	lower and upper constant control bounds
h	vector of state and control trajectory constraint functions
ψ	vector of state constraints at t_f
t_0	real process time
δt	optimization time consumption
T	length of optimization time horizon.

Numerical algorithms computing open-loop optimal controls have been proposed by numerous contributors from the early 60s, and one possible classification is function space versus control vector parameterization (CVP) algorithms. In the former class the controls are time functions which are determined by specific values for all time instants. A prescribed parametric control function form is used in the latter class, which means that the controls are completely specified by a limited number of parameters.

Function space algorithms

The Pontryagin maximum principle provides a control gradient value at each time instant. With conjugate gradient and variable metric types of function space algorithms the control variables are updated at every time instant according to these gradient values. Extensions of finite-dimensional conjugate gradient algorithms to function space were proposed by Lasdon *et al.* (1967), while Lasdon (1970), Tripathi and Narendra (1970) and Edge and Powers (1976) proposed similar extensions for the variable metric or quasi-Newton class. Quintana and Davison (1974) and Edge and Powers (1976) handled control bounds

explicitly by a clipping-off method. These classical function space algorithms encounter great difficulties when there are constraints imposed on some parts of the state trajectory. Penalty functions have been proposed to treat fixed final states or state inequality path constraints (Kelley, 1962; Lasdon *et al.*, 1967; Quintana and Davison, 1974), which may result in poor convergence. Furthermore, the penalty function techniques must be carefully implemented and tuned (Strand and Balchen, 1990). Although these algorithms may work well on unconstrained or control bound constrained optimal control problems, they are not recommended as general tools in the SSPC system, because of their inefficiency in handling general constraints.

Other types of function space algorithms have also been suggested, but will not be discussed here. This is because different types of CVP algorithms seem to offer the necessary constraint handling efficiency, and the control parameterization may be used as a controller design tool as well.

Control vector parameterization

Using CVP the original infinite-dimensional optimal control problem is transformed into a parametric optimization problem in a finite-dimensional space. This means that well-developed nonlinear programming optimization codes may be applied, and state constraints in the original problem become nonlinear constraints in the parametric optimization problem.

When the ordinary differential equations (ODEs) system is solved from initial to final time as one initial-value problem by the application of a numerical integration scheme, the technique is often called *single shooting*. In *multiple shooting* the time horizon is divided in subintervals, on which the ODEs are locally integrated as independent initial-value problems. Then the ODEs for a given control trajectory are not solved before state continuity at the subinterval knots is achieved. CVP has been combined with both single and multiple shooting techniques to solve optimal control problems.

Control parameterization with single shooting has been proposed by many authors, and the main differences are in choice of control approximating functions, treatment of state trajectory and end point constraints, and the gradient computation techniques.

CVP and single shooting

Sargent and Sullivan (1978) provide an excellent description of CVP combined with single shooting and analytical gradient computations. Each control variable is approximated by piece-wise polynomials, with parameter subsets to describe the polynomial in each subinterval. The subinterval knots, or control switching times, which may be optimized, are common for all control variables. Applying the Pontryagin maximum principle, analytical gradient equations are outlined for the optimization parameters. State trajectory constraints are handled by the

introduction of an integral equality constraint. Efficient ODE solvers for stiff systems are also discussed.

All the constraints of the optimal control problem (9) may be incorporated in Sargent and Sullivan's strategy, and their proposed control parameterization is quite general. However, there are some further considerations concerning:

- differentiability
- time consumption
- physical instability.

Deriving the necessary analytical gradient expressions may be quite laborious for some optimal control problems, sometimes impossible. When it comes to actual process implementation in the SSPC system, analytical gradients should be derived whenever possible, with regard to reliability and computational efficiency. In the design or feasibility study phases, however, and when analytical derivatives are not obtainable, an alternative approach is numerical gradient calculations. This can be done by cyclic parameter perturbation and corresponding solution of the initial-value problem to measure the perturbation responses on the performance index and constraint values. This strategy calls for high accuracy in the initial-value solver and balanced choice of perturbation intervals in order to get sufficient accuracy in the gradient values. The application of internal numerical differentiation (Bock, 1983) is more efficient; this is claimed to save 60 to 80% gradient computational time compared to the former numerical strategy. On the other hand, the cyclic control parameter perturbation and the initial-value solution strategy are ideally suited for implementation in the parallel processing systems which are available today. These often have very low costs compared to the potential increase in process earnings from more advanced control. Even if analytic gradients are derived, numerical verifications should be performed.

CVP and multiple shooting

CVP combined with multiple shooting (Bock and Plitt, 1985) instead of single shooting leads to more optimization parameters. This is because the initial-state variables for all subinterval initial-value problems inside the time horizon are added to the set of optimization parameters. Besides, one, generally nonlinear equality constraint is added to the problem for each extra optimization parameter, to demand state continuity at the subinterval knots. Despite this increased dimensionality, the multiple shooting technique may be preferable to single shooting. If gradients are to be calculated numerically, the effect of control parameter perturbation is strictly local if the control approximating functions are local, giving much less integration effort for each control parameter. However, internal numerical differentiation in single and multiple shooting does not make this advantage clear, and when analytical gradients are available the time consumption in either method is about

the same for gradient calculations. The possibility of removing the convergence influence of bad initial control parameter guesses is more important. This is because the subinterval initial-state values may be prescribed along a reasonable solution trajectory and the control parameters only have a local effect in the first integration. In the SSPC scheme proposed here, the initial control guesses are usually good if the model does not change very much between two optimization runs. What may be a problem, however, is local physical model instabilities occurring during the optimization. If that is likely to happen, multiple shooting should be preferred to single shooting. Of course, a system which is unstable must be stabilized before it is applied in an open-loop optimization.

A clear disadvantage of combining CVP with multiple rather than single shooting in the SSPC system is the invalid state variable trajectories for a given control before optimization convergence is attained. This fact degrades the information which can be given to the operator during the optimization run, and increases the difficulties in judging the effects of a timing interrupt of the optimization procedure, either from an automatic timer or from the operator.

CVP with polynomial state approximation and collocation

In all the methods discussed so far, the state variable differential equations are solved by a numerical integration scheme, completely solved for every control variable pattern in single shooting, and locally solved on the subintervals in multiple shooting. In the latter, continuity of the state variables is obtained while the optimization converges. By approximating the state variables explicitly by polynomials, *collocation techniques* can be used to decide the parameters in the polynomial expansions to get the best fit to the model. In fact, this is nothing else but implicit numerical integration, and for many collocation schemes the similarity to the Runge-Kutta formulas is obvious. Hence, the collocation technique may be used to solve the state equations in the optimal control problem, but nothing is gained from applying multiple or single shooting if the equations are completely solved for every control pattern. The idea is to solve the differential equations and the optimization problem simultaneously by adding the state approximating parameters to the set of optimization parameters. The collocation equality equations become equality constraints in the optimization problem.

Neuman and Sen (1973) applied cubic splines to approximate both control and state variables, i.e. third-order polynomials were chosen on a finite number of subintervals. A quadratic programming code was applied to solve the problem, since the problem was linear-quadratic with a state variable inequality path constraint which was linear in the state. Later contributions on collocation techniques to solve optimal control problems are those by Tsang *et al.* (1975), Biegler (1984), Cuthrell and Biegler (1987), Renfro

et al. (1987) and Vlassenbroeck and van Dooren (1988).

The collocation scheme may be differentiated analytically and combined with analytical or numerical differentiation of the original problem constraints and the right-hand sides of the ODEs. An important observation is that the possible lack of accuracy in the ODE solutions affected by the choice of approximating polynomials does not degrade the optimization convergence through inaccurate gradients.

The dimension of the optimization problem is drastically increased by this approach, while the computational effort spent on the model solution for each optimization iteration is reduced by a fair amount compared to single and multiple shooting techniques. The method relies on nonlinear programming solvers which efficiently handle nonlinear equality constraints. Here, sequential quadratic programming (SQP) codes will be useful. Further, codes which exploit the sparsity of the constraints Jacobian may be advantageous in practical applications.

As in the multiple shooting technique, the model solution with a given control pattern is not valid before optimization convergence is attained. Whether this methodology should be preferred in the SSPC system, must be judged from increased optimization dimensionality versus reduced gradient computational effort, from an operator's point of view, and from system installation and maintenance considerations. Performance comparisons of different types of optimal control algorithms demonstrate the potential of the collocation techniques as general tools (Biegler, 1984; Strand and Balchen, 1990).

Nonlinear programming algorithms

SQP algorithms are most frequently used to solve nonlinear programs (NLP), which results from the application of CVP technique to original optimal control problem. Well-developed codes are NPSOL (Gill *et al.*, 1983) and NLPQL (Schittkowski, 1985). If integer variables are added to the problem, by discrete control variable values, for instance, the arising optimization problem is termed mixed integer nonlinear programming (MINLP). In the case of discrete control values, a satisfactory solution may be to approximate the control by continuous variables in the optimization and pick the nearest realizable value when implementing the calculated controls on the process. If this is not satisfactory, the discrete values, or even the on/off values, must be iterated in an outer optimization loop, while the continuous problem resulting from actual outer choices may be solved by an NLP solver (Duran and Grossman, 1986; Floudas *et al.*, 1989).

Control parametrization considerations in the SSPC context

There are several considerations which should be kept in mind when the CVP strategy is to be discussed:

- The process model to be employed in the open-loop optimization is at best a reasonable approximation of process behavior.
- Only the first part of the optimized controls will ever be applied on the process.
- The model is usually changing from one optimization to the next.
- The CVP must provide enough degrees of freedom to meet the constraints.
- Batch process control usually demands more flexibility in the control pattern than the control of a continuous process.
- The total computational effort spent on gradient calculations grows much faster in numerical than analytical calculations with an increasing number of parameters describing each control.
- The control must be parameterized according to realizable patterns. This means for instance that an integrated control should not be allowed to change in steps; it should rather be continuous.

So far only piecewise constant and linear parameterizations have been considered. These are believed to provide enough approximating accuracy from the above considerations. In the FCC example three to six control intervals favorably balanced the computation time and performance index value. In the light metal electrolysis example, the minimum number of control parameters was merely governed by the satisfaction of the constraints.

Typical questions that arise when deciding the CVP are:

- Piecewise linear or constant control?
- Number of control intervals?
- Control switching time optimization?
- Common or individual switching times for the control variables?
- Minimum length of different control intervals?
- Optimize initial value if piecewise linear control?
- Demand constant control in the last control interval if the controls are linearly parameterized?

These choices influence the achievable value of the performance index, and may influence the controllability with regard to constraints or set points. They may also be used to balance the controller bandwidth against the time delay introduced in the optimization loop.

SSPC OF A FLUIDIZED CATALYTIC CRACKING PROCESS

A modern FCC process consists mainly of three parts, a riser, a separator and a regenerator as illustrated in Fig. 3. The heavy feed oil is mixed with hot catalyst from the regenerator and reacts endothermically in the riser to form lighter hydrocarbons as well as coke. The lighter hydrocarbons, mainly gasoline, are separated from the catalyst in the separator and sent to the fractionator, while the coke is deposited on the catalyst which is returned to the regenerator. In order to maintain an acceptable catalyst activity the

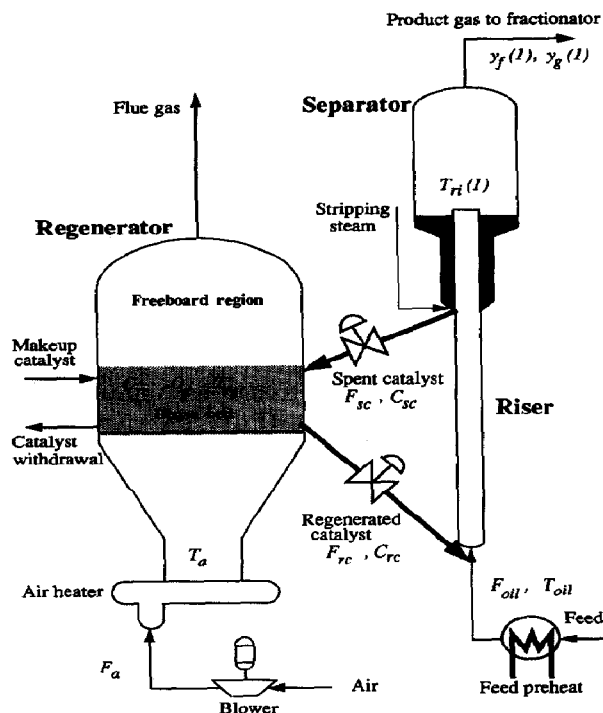


Fig. 3. Schematic diagram of the cracking plant.

coke is burned off in the regenerator. The air needed for this combustion is supplied at the bottom of the regenerator in such a way that the catalyst is kept fluidized in the rising air. This ensures good air and catalyst mixing in the so-called dense bed of the regenerator. Irreversible deactivation and loss of catalyst in the flue gas mean that makeup catalyst and catalyst withdrawal are needed during continuous operation of the FCC unit.

The FCC unit exposes complex dynamics due to strong cross-couplings between the endothermic riser and the exothermic regenerator sections and nonlinearities in the reaction rates. In addition, nonlinearities will occur due to constraints on the process variables:

- The rate of air supply to the regenerator is limited due to the process equipment, and the fractionator capacity limits the rate of feed oil.
- The temperatures on the feed oil and the air to the regenerator can only be varied within narrow constraints.

Modelling the FCC plant

Although major oil companies claim to have sophisticated models of the cracker, the amount of published work in the open literature on dynamic catalytic cracking models is rather restricted. Luyben and Lamb (1963) described a simplified model using first-order kinetics for both the cracker and the regenerator. Kurihara (1967) presented a model with more

realistic kinetics for the cracking reaction, but his assumption that the cracking takes place in a perfectly mixed tank with negligible cracking in the riser is not valid for modern zeolite catalysts. Isles-Smith (1985) developed a detailed model, but it is questionable whether such a complex model is necessary when the important issue is to compare different control strategies. Lee and Groves (1985) have published a relatively simple model where the kinetics in the riser is modeled using Weekman's "three-lump" model (Weekman and Nace, 1970), extended by Shah *et al.* (1977) to account for cracking in transfer line under adiabatic, endothermic conditions. The regenerator model comes from Errazu *et al.* (1979), while the carbon buildup on the catalyst is from Kurihara (1967). Huang *et al.* (1989) use a model similar to that of Lee and Groves, except that a simple stripper model is included and the regenerator afterburning is accounted for. As pointed out by Isles-Smith (1985) it may be necessary to include the afterburning effect and the pressure drop across the standpipe in order to get correct long-range predictions. However, complicated models cannot be assured to be more precise than simpler ones due to lack of appropriate model parameters. Moreover a simple model may be used in an on-line estimation scheme and it may be extended if it turns out that neglected phenomena are more important to the model dynamics than expected. In the simulations below, the process measurements are computed from a model which is identical to that of Lee and Groves (1985) except for some minor modifications. Some of the numerical values given there may be somewhat unrealistic, but are used here to simplify controller comparisons.

A simplified FCC model

Lee and Groves' model is quasi-dynamic because the riser equations are steady-state ordinary differential equation which are solved every 0.5 min, while the regenerator is modeled using unsteady-state energy and mass balances in the dense bed. If the regenerator temperature is kept between 920 and 1040 K and ideal level control of the stripper is assumed, the following model will describe the FCC process reasonably well (Ljungquist, 1990):

Regenerator:

$$W \frac{dC_{rc}}{dt} = F_{sc}(C_{sc} - C_{rc}) - kO_d C_{rc} W \quad (10)$$

$$W_a \frac{dO_d}{dt} = R_a(O_{in} - O_d) - \frac{n+2+(n+4)\sigma}{4M_c(1+\sigma)} kO_d C_{rc} W \quad (11)$$

$$Wc_{ps} \frac{dT_{rg}}{dt} = T_{rl}(1) F_{sc} c_{ps} + T_a F_a c_{pa} - T_{rg}(F_{sc} c_{ps} + F_a c_{pa}) - \Delta H \frac{kO_d C_{rc}}{M_c} W \quad (12)$$

where t is the time and

$$C_{sc} = C_{rc} + C_{cat} \quad (13)$$

$$C_{cat} = k_c^1 \sqrt{\frac{t_c}{C_{rc}^N} \exp\left(-\frac{E_{cf}}{RT_{ri}(1)}\right)} \quad (14)$$

$$k = k_{com} \exp\left[\left(\frac{1}{960} - \frac{1}{T_{rg}}\right) \frac{E_{cb}}{R}\right] \quad (15)$$

$$\Delta H = -h_1 - h_2(T_{rg} - 960) + 0.6(T_{rg} - 960)^2 \quad (16)$$

$$\sigma = 1.1 + \sigma_2(T_{rg} - 873) \quad (17)$$

Riser:

$$y_f(1) = \frac{y_f(0)\alpha}{\alpha + K_r\phi_0[1.0 - \exp(-\alpha t_c \text{COR})]} \quad (18)$$

$$T_{ri}(1) = \left\{ 1.0 - \frac{\gamma y_f(0) K_r \phi_0 [1.0 - \exp(-\alpha t_c \text{COR})]}{\alpha + K_r \phi_0 [1.0 - \exp(-\alpha t_c \text{COR})]} \right\} \times T_{ri}(0) \quad (19)$$

$$y_g(1) = (1 + R_r) \frac{F_{gi}}{1.0 - I_{gi}} [y_f(1)^{I_{gi}} - y_f(1)] \quad (20)$$

where

$$K_r = k_0 \exp\left(\frac{-E_f}{RT_r}\right) \quad (21)$$

$$T_r = \left(1.0 - \frac{\gamma y_f(0) K_0 \phi_0 [1.0 - \exp(-\alpha t_c \text{COR } z_r)]}{\alpha + K_0 \phi_0 [1.0 - \exp(-\alpha t_c \text{COR } z_r)]} \right) \times T_{ri}(0) \quad (22)$$

$$K_0 = k_0 \exp\left(\frac{-E_f}{RT_{ri}(0)}\right) \quad (23)$$

$$\phi_0 = 1 - mC_{rc} \quad (24)$$

$$\gamma = \frac{\Delta H_f F_{oil}}{T_{ri}(0) (F_{rc} c_{ps} + F_{oil} c_{po} + \lambda F_{oil} c_{pd})} \quad (25)$$

$$T_{ri}(0) = \frac{(c_{po} F_{oil} + \lambda F_{oil} c_{pd}) T_{oil} + c_{ps} F_{rc} T_{rg}}{c_{po} F_{oil} + \lambda F_{oil} c_{pd} + c_{ps} F_{rc}} \quad (26)$$

Typical numerical values are given together with the symbol definitions in the Notation. z_r may be interpreted as the location in the riser where the average reaction temperature occurs, and should be tuned to make the algebraic riser equations agree with the more accurate differential description (Shah *et al.*, 1977). The gasoline eq. (20) is from Kurihara (1967). Equations (5)–(7), which form a rather stiff set of differential equations, will be referred to as model 1. With the numerical values given, the largest time constant is about 75 min while the smallest is about 0.008 min. The smallest time constant is closely related to the oxygen balance in the dense bed, eq. (6). One way to reduce this stiffness is to increase W_a , i.e. to use a fictitious air holdup in regenerator. With a sample interval equal to 0.5 min, this variable may be increased to 500 with negligible effect on the model responses. The smallest time constant is then in-

creased to about 0.2 min. Another strategy is to replace the dynamic equation (6) with the steady state

$$O_d = \frac{R_a O_{in}}{R_a + \frac{n+2+(n+4)\sigma}{4M_c(1+\sigma)} k C_{rc} W} \quad (27)$$

The resulting model, referred to as model 2, gives approximately the same responses as model 1.

An experimental study of the computational effort in the numerical integration of the two models is presented in Strand (1991), where the performance of a stiff solver is compared with several nonstiff solvers for both models. The stiff solver is a third-order variable stepsize semi-implicit Runge–Kutta method due to Nørsett and Thomsen (SIMPLE, 1987). The nonstiff solvers are explicit Runge–Kutta methods, ranging from the forward Euler to the fifth-order variable step-size method of Dormand and Prince [Dopri 5(4), 1980].

The stiff solver is most efficient for low-accuracy (10^{-2}) integration of model 1, while a moderate accuracy requiring (10^{-4}) makes the stiff solver and a second-order variable stepsize explicit Runge–Kutta–Fehlberg [RKF2(3)] method about equally efficient. The RKF2(3) method performs best for low-accuracy integration of model 2, while the moderate accuracy request makes Dopri5(4) the best method for this model. For the optimal choices of integration method, model 2 is solved about five times faster than model 1 for both accuracies.

Identification of the FCC model

The state variables in a state-space model contain information about the history of the past and reflect the model order. Real processes are theoretically of infinite order, and it may be questionable to reduce the state dimension too much. It is felt that a second-order model may be insufficient to describe a real FCC reasonably well. Thus model 1 is used as the real-time model in the SSPC scheme, Fig. 1. When it comes to long-range predictions, the smaller time constant may not be so important, and therefore model 2 is used as the predictive optimization model in the SSPC scheme to save computational costs. Because the fractionation system is not included in the model, it is not realistic to assume that the components in the product gas are measured. The measurement variables then are:

- riser outlet temperature;
- oxygen mole fraction in regenerator dense bed;
- regenerator dense bed temperature.

A reasonable parameter vector to be adjusted online is:

$$\theta = [k_0 k_c^1 k_{com} \sigma_2 h_1 h_2 E_{cb}/R]^T \quad (28)$$

where

- | | |
|---------|--|
| k_0 | rate constant for gas oil cracking |
| k_c^1 | rate constant for catalytic coke formation |

k_{com}	rate constant for coke burning
σ_2	linear CO_2/CO dependence on the temperature
h_1, h_2	parameters in the ΔH approximation
E_{cb}/R	activation energy for coke burning reaction.

Using the optimal parameter vector, model 1 and moderate excitation, the eigenvalues corresponding to an estimate of the Hessian over a time horizon of 250 min are:

$$\text{eig}(F_{7 \times 7}) = \{5.17 \times 10^2, 3.74 \times 10^1, 1.36 \times 10^1, 1.29, 4.31 \times 10^{-1}, 8.41 \times 10^{-2}, 1.33 \times 10^{-4}\}. \quad (29)$$

There is a large step between the two smaller eigenvalues, which indicates that only six parameters should be estimated in this experiment. The eigenvector corresponding to the smaller eigenvalue is dominated by the fifth element, and if h_1 is kept constant, the new eigenvalues are:

$$\text{eig}(F_{6 \times 6}) = \{5.11 \times 10^2, 3.52 \times 10^1, 1.35 \times 10^1, 1.28, 4.00 \times 10^{-1}, 7.65 \times 10^{-2}\}. \quad (30)$$

This result shows that a reduced parameter set leads to a more well-conditioned Hessian, and simulation results verify that it is hard to estimate h_1 on-line using an AKF. Simulation tests show that if reasonable measurement noise is added to the simulated process measurements, the number of parameters to be adjusted online should be less than or equal to four when a recursive estimation algorithm is used (Ljungquist, 1990).

The system sample time used in the simulations is 0.5 min and the AKF is run for 8 h with correct initial parameters before time equal to 0.0 to attain reasonable initial values for the covariance matrix of the augmented state. In all the simulations where the AKF is used, k_0 , k_c^1 , k_{com} , σ_2 and h_2 are estimated. If the estimation algorithm performs well, measurement noise will not significantly influence the resulting control strategy. Measurement noise is therefore not included in the simulations shown in Figs 4–22, where an increase of 4% in the coking nature of the feed occurring at time equal to 1.0 h is used to compare different control strategies. This means that k_c^1 is increased from 0.019 to $0.01976 \text{ s}^{-1/2}$ in the process. Note that the model parameters differ slightly from the parameters used in the process, as illustrated in Fig. 14, because of the structure difference. The innovations corresponding to the two simulations shown in Figs 14 and 15 were almost identical, indicating that it is sufficient to adjust five parameters online.

FCC control

The objectives of an FCC control scheme may be summarized as:

- For a given catalyst and a given feed, the riser exit temperature should be kept as close to

a given temperature as possible in order to maximize the profit.

- The temperature in the regenerator dense bed should be kept low to minimize the irreversible catalyst deactivation and high to maximize the coke burning rate.
- The oxygen concentration in the regenerator dense bed should be kept low to avoid excessive afterburning in the freeboard region and a potential dangerous situation, but it should be as high as possible to maximize the coke burning rate.

A short review of the control strategies reported up to 1980 can be found in Balchen and Mumme (1988). Usually the feed rate, the feed temperature and the temperature of the air to the regenerator are set at constant values while the mass flow rate of spent catalyst is used to control the mass holdup in the stripper section. This leaves two control variables to control the three measurements. In the conventional control structure (Hicks *et al.*, 1966) the riser exit temperature is controlled by adjusting the recirculation rate of regenerated catalyst, referred to as the catalyst circulation rate from now on, while the dense-bed oxygen fraction is controlled by adjusting the inlet air rate. This control scheme results in oscillatory behavior, and the integral action in the PI-controller has to be weak in order to avoid instability as illustrated in Figs 4–7, where the control parameters used can be found in Lee and Groves (1985). Due to the nonlinearity of the FCC process, the controllers have to be retuned if the operational conditions change. If the PI-controller simulation in Figs 4–7 is repeated with a corresponding decrease in the coking nature, the controller responses are very bad.

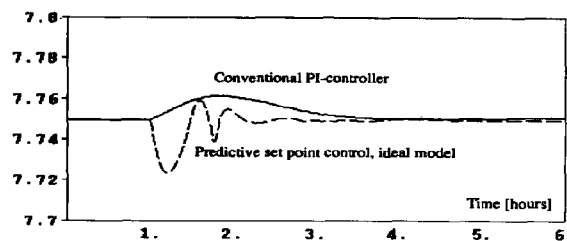


Fig. 4. Riser outlet temperature, $T_{r1}(1)$ ($\times 100 \text{ K}$) (4.0% increase in carbon formation rate, k_c^1 , at $t = 1.0 \text{ h}$).

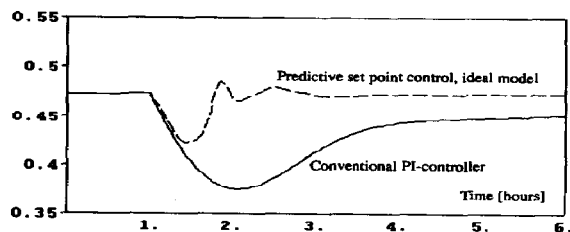


Fig. 5. Oxygen mole fraction in regenerator dense bed, O_d (mol%) (4.0% increase in carbon formation rate, k_c^1 , at $t = 1.0 \text{ h}$).

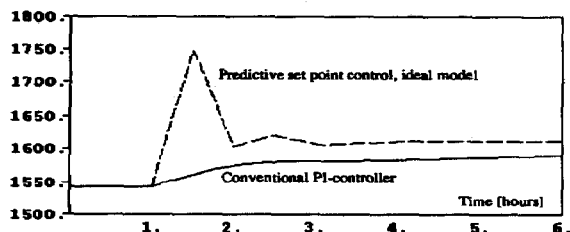


Fig. 6. Air rate to regenerator, F_a (kg/min) (4.0% increase in carbon formation rate, k_c^1 , at $t = 1.0$ h).

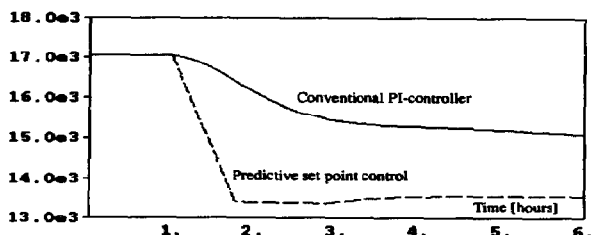


Fig. 7. Regenerated catalyst recirculation rate, F_{rc} (kg/min) (4.0% increase in carbon formation rate, k_c^1 , at $t = 1.0$ h).

Other control schemes based on PI-controllers have been proposed (Kurihara, 1967; Lee and Weekman, 1976; Browley and Ward, 1981) and simulation results indicate that they perform much better than the conventional controller. Even so, they are not widely used for practical processes. LQ controllers have been proposed (Schuldt and Smith, 1971; Isles-Smith, 1985), but it turns out that these do not have the desired robustness because they rely on a linear model for the state estimation and feedback matrix computation. More recently MPC techniques have been successfully implemented on real FCC processes (Rhemann *et al.*, 1989).

The present FCC simulation study is performed to illustrate some features of the proposed SSPC technique. The simulated FCC process serves as a reasonably complicated test case, and some mismatch is introduced between the simulated process and the nonlinear models which are used in the SSPC scheme. The obtained results will indicate the advantages of using the SSPC strategy to control a more complicated real FCC process. A simulation study is the natural way to quantify these advantages.

In the simulation experiments to be discussed next, the SSPC is implemented with the following two criteria [cf. eq. (9)]:

Predictive set-point control:

$$L1 = P_{r1}(T_{r1}(1) - T_{r1}^0)^2 + P_O(O_A - O_A^0)^2 + P_{Tg}(T_{Tg} - T_{Tg}^0)^2 \quad (31)$$

Optimal predictive control:

$$L2 = F_{oil} \{ y_g(1) Pr_0 - [1.0 - y_f(1)] Pr_1$$

$$+ [1.0 - y_f(1) - y_g(1)] Pr_2 \} - F_{oil}(T_{oil} - 400.0) Pr_3 - Co_{cat} - P_{Tfg} - O_A Pr_4 - F_a Pr_5 \quad (32)$$

where

$$Co_{cat} = \begin{cases} Co_1 & \text{if } T_{rg} < 960 \\ Co_1 + Co_2(T_{rg} - 960)^2 & \text{if } T_{rg} \geq 960 \end{cases}$$

$$P_{Tfg} = \max \{ 0.0, (T_{rg} + 4500 O_A) - 1010 \} P_1. \quad (33)$$

The numerical values used are given together with the symbol definitions in the notation list. The profit-based criterion $L2$ includes product stream values and costs related to the use of control variables. In addition, a simple catalyst cost model is included to account for the increased irreversible catalyst deactivation for high regenerator temperatures. The potentially dangerous after burning is considered by a linear penalty on the flue gas temperature as shown in eq. (33). It might be more correct to use some combination of a penalty and a hard constraint to account for this effect in a real process. This profit function can be used for steady-state optimization for the computation of set points, but in the present context it is also used to optimize the profit dynamically. Then the profit function $L2$ is the integrand of eq. (9).

In all the SSPC simulations, a minimum control interval length of 0.5 h is used, and the control approximations are piecewise linear. The initial values of the control variables are kept constant in the optimization in order to attain realizable solutions, and three to five control intervals are used in the experiments with moving time horizons. A single shooting technique with analytical gradient calculation is applied, giving a maximum optimization time consumption of about 20 s on a Sun SPARC station 1 (Strand, 1991). Despite this, 10 min is used as the sample time δt for the optimization algorithm, which means that about 1/30 of the SPARC computing capacity will be sufficient for an on-line implementation with this model complexity.

Criterion $L1$ corresponds to predictive set-point control like MAC and DMC except that a nonlinear state-space model is used. Figures 4–7 compare predictive set-point control using a perfect model, i.e. correct parameters and no estimation, and with no weight on the regenerator temperature, to a conventional controller. Although seven control intervals are used in this single horizon optimization, Figs 6 and 7 indicate that three control intervals may be sufficient in moving horizon applications. It is seen from Figs 4 and 5 that the responses are significantly faster than those of the conventional controller, especially when it comes to the oxygen mole fraction response. Note that the amplitude of the SSPC temperature response can be made smaller at the cost of the oxygen fraction amplitude by increasing the weight on the temperature deviation in $L1$, and that the responses can be made faster by reducing the minimum control interval length in the optimization.

When the SSPC scheme is applied with state estimation only, both the state initial values used in the optimization and the predictions of the optimization model are incorrect compared to the process. The computed controls make the prediction model reach the set points, while the process measurements deviate. When a feedback matrix is included the deviations become smaller. This is illustrated in Figs 8 and 9 where the feedback matrix is constant and computed using a linearized model at the set points.

Figures 10 and 11 show that the responses are satisfactory when the parameters are estimated. The feedback matrix is used in this simulation as well. In all the predictive control simulations shown in Figs 14–22 both the parameter estimator and the feedback matrix are included.

If the regenerator temperature is weighted instead of the riser outlet temperature in $L1$, the controller performance is improved as shown in Figs 12 and 13. This result agrees with the fact that the Kurihara controller perform better than the conventional controller at this set point. Isles-Smith (1985) shows that if the FCC process is operated in the high-temperature regeneration mode, the conventional controller may be advantageous to the Kurihara controller. A predictive set-point controller takes advantage from the fact that the pairing problem is avoided, and the number of control variables and measurements do not have to be the same. In the FCC process all three measurements should be weighted in the criterion. If a disturbance occurs and the set points cannot be reached due to controllability problems, all the measurements will then deviate from their set points. However, the deviations are weighted against each

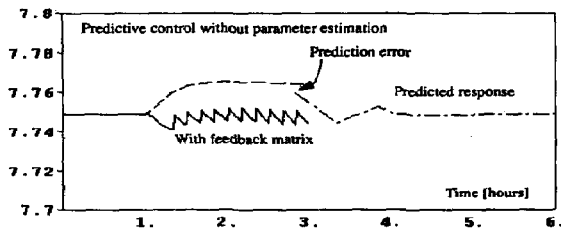


Fig. 8. Riser outlet temperature, $T_{ri}(1)$ ($\times 100$ K) (4.0% increase in carbon formation rate, k_c^1 , at $t = 1.0$ h).

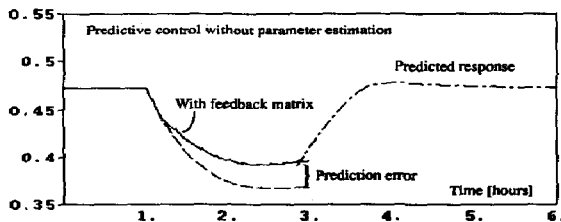


Fig. 9. Oxygen mole fraction in regenerator dense bed, O_d (mol%) (4.0% increase in carbon formation rate, k_c^1 , at $t = 1.0$ h).

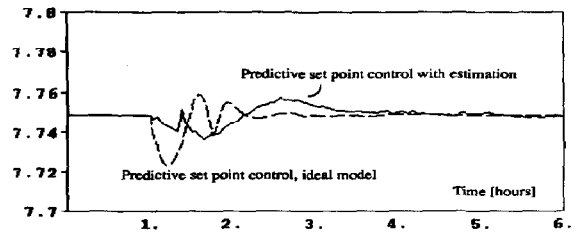


Fig. 10. Riser outlet temperature, $T_{ri}(1)$ ($\times 100$ K) (4.0% increase in carbon formation rate, k_c^1 , at $t = 1.0$ h).

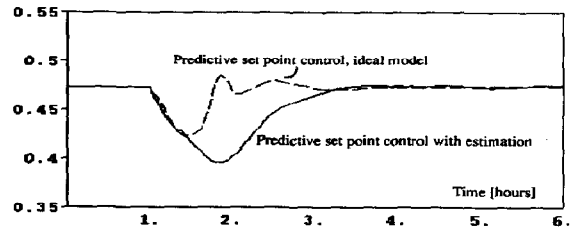


Fig. 11. Oxygen mole fraction in regenerator dense bed, O_d (mol%) (4.0% increase in carbon formation rate, k_c^1 , at $t = 1.0$ h).

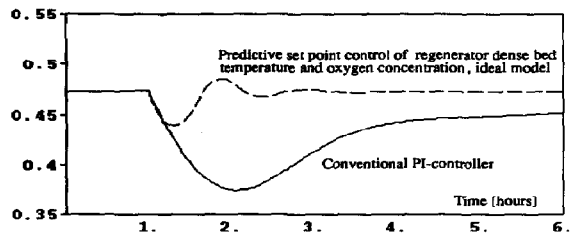


Fig. 12. Oxygen mole fraction in regenerator dense bed, O_d (mol%) (4.0% increase in carbon formation rate, k_c^1 , at $t = 1.0$ h).

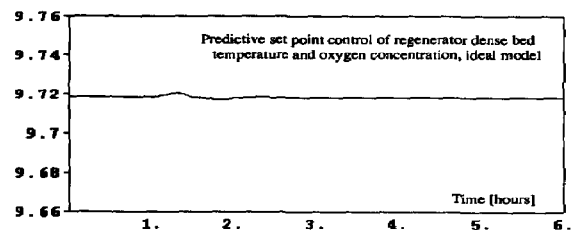


Fig. 13. Regenerator dense bed temperature, T_{rg} ($\times 100$ K) (4.0% increase in carbon formation rate, k_c^1 , at $t = 1.0$ h).

other according to the criterion $L1$, instead of forcing just two measurements to their set points. The advantage of such a control strategy is illustrated in Figs 16–21. The initial set points are found from the steady-state optimal set point computed from $L2$ with a coke formation rate of $0.019 \text{ s}^{-1/2}$. When the formation rate is increased by 4%, the net profit rate will decrease as illustrated in Fig. 21. It is then assumed that the set point is changed to the new optimal set

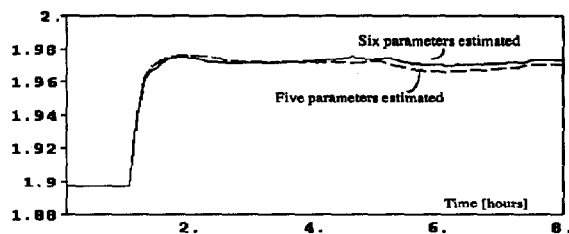


Fig. 14. Estimated rate constant for catalytic coke formation, k_c^1 ($\times 0.01 \text{ s}^{-1/2}$) (4.0% increase in carbon formation rate, k_c^1 , at $t = 1.0$ h, set-point change at $t = 5.0$ h).

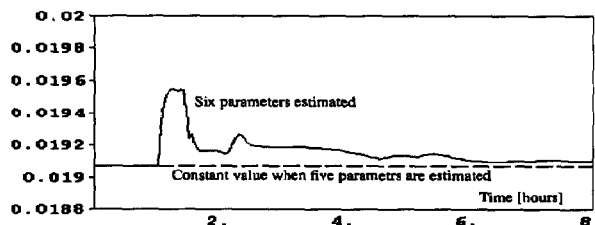


Fig. 15. Estimated activation energy for coke burning reaction, E_{cb}/R ($\times 10^7 \text{ K}$) (4.0% increase in carbon formation rate, k_c^1 , at $t = 1.0$ h, set-point change at $t = 5.0$ h).

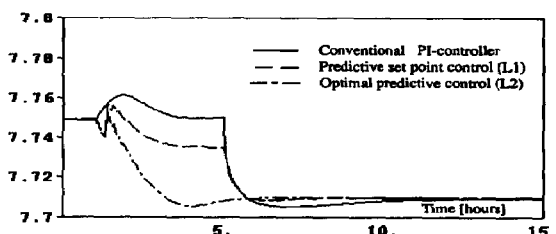


Fig. 16. Riser outlet temperature, T_{r1} (1) ($\times 100 \text{ K}$) (4.0% increase in carbon formation rate, k_c^1 , at $t = 1.0$ h, set-point change at $t = 5.0$ h).

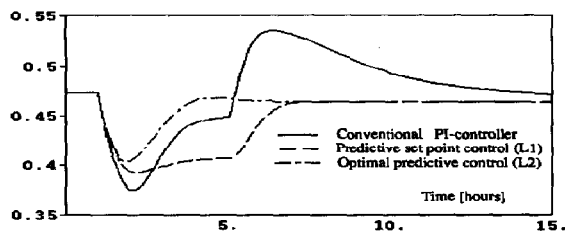


Fig. 17. Oxygen concentration in regenerator dense bed, O_d (mol%) (4.0% increase in carbon formation rate, k_c^1 , at $t = 1.0$ h, set-point change at $t = 5.0$ h).

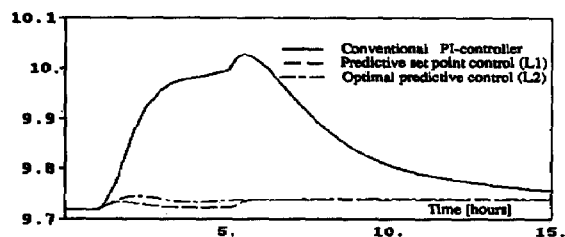


Fig. 18. Regenerator dense bed temperature, T_{rg} ($\times 100 \text{ K}$) (4.0% increase in carbon formation rate, k_c^1 , at $t = 1.0$ h, set-point change at $t = 5.0$ h).

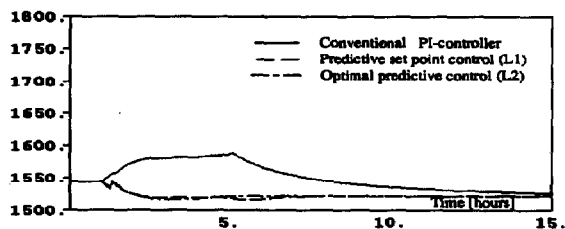


Fig. 19. Air rate to regenerator, F_a (kg/min) (4.0% increase in carbon formation rate, k_c^1 , at $t = 1.0$ h, set-point change at $t = 5.0$ h).

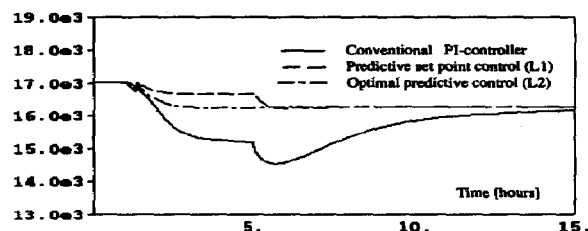


Fig. 20. Regenerated catalyst recirculation rate, F_r (kg/min) (4.0% increase in carbon formation rate, k_c^1 , at $t = 1.0$ h, set-point change at $t = 5.0$ h).

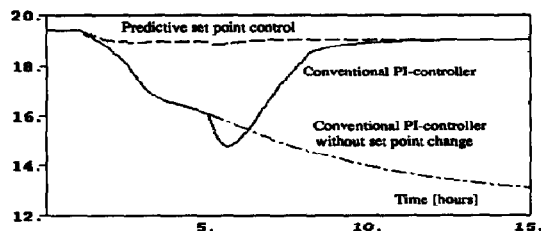


Fig. 21. Net profit rate ($\times 1000 \text{ NOK/h}$) (4.0% increase in carbon formation rate, k_c^1 , at $t = 1.0$ h, set-point change at $t = 5.0$ h).

point 4 h after the feed coking nature is changed. This decision may be based on the corresponding parameter estimates shown in Figs 14 and 15. Figure 21 illustrates the net profit rate based on $L2$ and clearly shows the advantage gained from using predictive set-point control compared to PI-control. The main reason is that the conventional controller results in

a very high regenerator temperature before the set points are changed (Fig. 18).

The simulation experiments discussed above demonstrate some of the advantages of predictive control. However they do not illustrate all the advantages of using a criterion based on net profit rate, $L2$. From Figs 16–22 it is seen that the resulting optimal pre-

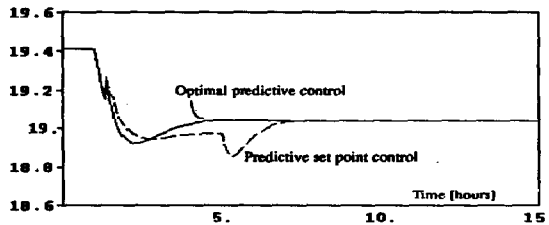


Fig. 22. Net profit rate ($\times 1000$ NOK/h) (4.0% increase in carbon formation rate, k_c^1 , at $t = 1.0$ h, set-point change at $t = 5.0$ h).

dictive control results in a better profit rate than any other control scheme. A lot of simulations were performed using this criterion (Strand, 1991), and the main results may be summarized as:

- The net profit rate is increased by including the oil feed rate in the optimization. But then the control variables must be constrained because maximization of the profit rate results in maximum rate of either air to regenerator or oil feed to riser.
- When a profit function like L_2 is used, state constraints are not necessary for the simulated FCC process. However, a different catalyst cost function could change this conclusion.
- Even though one control variable lies on its upper bound, the control strategy has no difficulty in handling process disturbances.

SSPC OF A MULTIUNIT ELECTROMETALLURGICAL PROCESS

In order to highlight some of the capabilities of the proposed SSPC strategy, the multiunit electrometallurgical (electrolysis) process described by Balchen *et al.* (1989) is revisited. Part of the process is schematically drawn in Fig. 23. The units are structurally identical, and are modeled by energy and mass balance differential equations. What really makes this process an extremely strong candidate for SSPC, are the strong interactions between the units through the control variables (main direct current and demand on total use of feed aggregate A) and hard constraints on the temperature and concentration in each unit. The control actions and limitations are:

- main direct current through all units;
- additional direct current for each row of units, 0–8% of nominal main current;
- individual heating element for each unit, discrete in five levels;
- bypass possibility to cool a unit, which means that nothing is produced in that unit;
- feed aggregate A, with temperature about the same as that of the unit;
- feed aggregate B, chemically identical to A, but at much lower temperature;
- feed aggregate A is produced at an upstream stage with no storage facility. Hence, all available

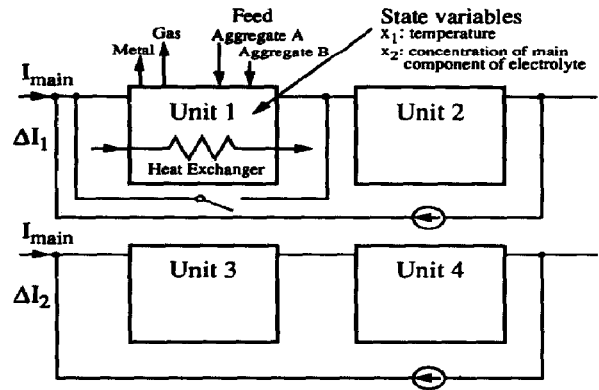


Fig. 23. The multistage electrometallurgical process.

aggregate A must be used immediately, and there is certainly a prediction possibility of future available feed due to the production time delay;

- there is a rather small stock capacity of feed aggregate B. In the present simulations a constant flow rate into the stock is assumed. Feed aggregate B costs about 50% more than aggregate A.

The goal of the process is metal production, at a high net profit rate. Though there are certain state-dependent conditions, such as increasing process yield with decreasing temperature, and decreasing internal resistance with increasing temperature and decreasing concentration, production continues as long as the state variables stay between their upper and lower bounds. These bounds are due to operational reasons, and keeping the state variables in the required domain is an absolute demand. This is why the unproductive control actions represented by additional heating and bypass are included in the process.

Under normal operating conditions it suffices to describe the process by two state variables for each unit (temperature and concentration). Under abnormal conditions a more detailed model is necessary, the structure and parameters of which are known and can be implemented in an industrial system. In this paper these abnormal aspects have been neglected in order to make the presentation tractable.

When this process is to be optimized over a finite-time horizon and the profit function consists of metal price and costs of raw material and energy, the unit concentrations are driven to their lower bound at the end of the time horizon. Naturally, it is optimal to reduce the "stock" of raw material in the unit to a minimum when a finite-time horizon is considered, but it is not a desirable solution since the process is to be continuously operated. Concentration at the lower bound is also favorable in terms of energy costs because of low resistance, while the temperature tends to approach the lower bound to increase the process yield.

Process variables which approach bounds when pure profit functions are maximized should not be too surprising. Because of process disturbances and the strong control couplings, it is not desirable to operate with process variables at their bounds. If possible from computational point of view, uncertainties could have been modeled and Monte Carlo simulations performed to decide stochastic optimal control strategies, and the process variables would certainly not have been driven to their bounds. A more practical approach is to introduce quadratic penalties on deviation from some prescribed values, which has actually been done in the present simulations.

The mass and energy balance differential equations are equal to those presented in Balchen *et al.* (1989), while the optimization problem is slightly differently formulated by:

Profit function (each unit):

- metal price \times process yield \times direct current
- dc energy price \times voltage across unit \times direct current
- price feed aggregate A \times feed rate aggregate A
- price feed aggregate B \times feed rate aggregate B
- price additional heating \times heating power
- linear cost of temperatures exceeding mean temperature due to equipment
- cost of temperature variations due to equipment.

Penalties (each unit):

- quadratic cost of temperature derivations from mean temperature
- + quadratic cost of concentration deviations from mean concentration
- + cost of concentration variation, which increases linearly with time.

Stock penalty:

- quadratic cost in upper and lower thirds of feed aggregate B stock.

Optimal control problem:

$$\max J = \int_{T_p} \left[\sum_{\text{units}} (\text{profit} - \text{penalties}) - \text{stock penalties} \right] dt \quad (34)$$

under the constraints: (35)

- temperature within upper and lower bounds;
- concentrations within upper and lower bounds;
- all units in one row are subject to the same direct current;
- additional heating is discrete in five levels;
- all available feed aggregate A must be used;
- stock capacity of feed aggregate B is fixed.

Solution strategy

Single shooting with numerical gradient calculations is applied in the simulation studies. The control variables are piecewise constant, with three to six independent control intervals. Feed aggregates A and B have equal control interval time distributions for all units to simplify the treatments of the two last constraints defined above. Temperature and concentration constraints are pointwise evaluated with a 10 h interval, while the feed constraints are evaluated at the control switching times. The discrete additional heating control has been regarded as continuous in the optimization, and the closest realizable level is implemented on the simulated process.

Simulation experiment

A lot of simulations have been done, and the SSPC strategy performs satisfactorily in the sense that the constraints are met if the control actions have enough degrees of freedom for the actual situation. Further, the pure profit maximization and controller flexibility maintenance are balanced according to the chosen criterion weights.

The FCC control simulations illustrate the modeling and identification aspects of the SSPC strategy, in addition to controller performance in a multivariable, highly nonlinear and dynamically coupled process. The present example highlights the SSPC advantages in the handling of constraints and disturbance predictions. Therefore, the optimization model and the simulated process are equal, except for the discrete additional heating control variables and the chosen disturbance if this is not predicted in the model.

In the simulation, presented in Figs 24–32, a disturbance in the amount of available feed aggregate A is applied. This realistic disturbance may be predicted because of the time delay in the feed production. The simulations show the resulting process control with and without disturbance predictions.

In order to make the message clear, just one process unit is simulated. This is nevertheless a valid assumption, since the disturbances in feed availability would have been equally distributed to each unit if the initial states and all process parameters were equal in all units. The optimization time horizon is 120 h, which is in fact quite long for this process, also compared with the time horizon movement of 4 h between each optimization.

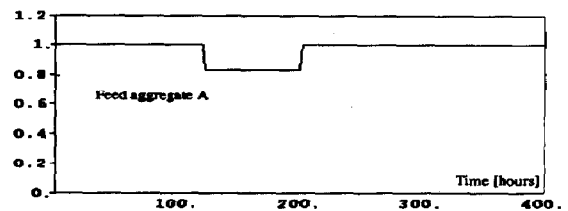


Fig. 24. Disturbances in available feed aggregate A.

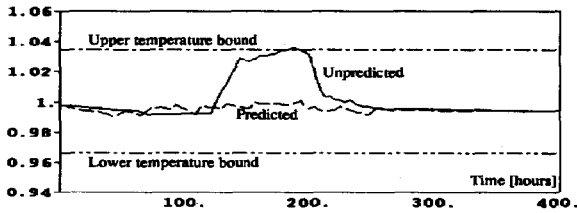


Fig. 25. Unit temperature (predicted and unpredicted disturbances in available feed aggregate A).

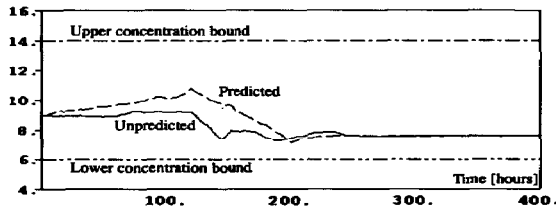


Fig. 26. Unit concentration (predicted and unpredicted disturbances in available feed aggregate A).

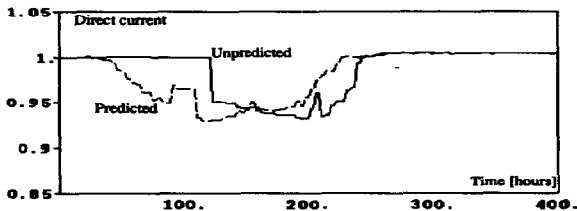


Fig. 27. Optimized direct current (predicted and unpredicted disturbances in available feed aggregate A).

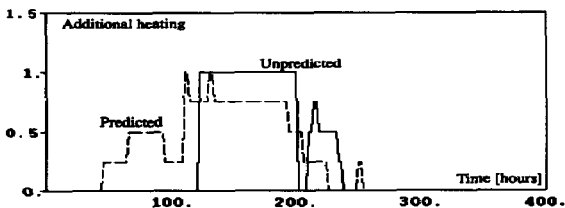


Fig. 28. Optimized additional heating (predicted and unpredicted disturbances in available feed aggregate A).

The obvious advantages of taking the predictable disturbances into account in the SSPC strategy in these simulations are:

- Unit temperature is kept at a more desirable level.
- Unit concentration prepares for the disturbance.
- Stock level of feed aggregate B prepares for the disturbance and is more desirable throughout the period of lowered availability of feed aggregate A.
- The net profit is increased.

When subtracting the total penalty, Fig. 32, from the net profit rate, Fig. 31, the resulting quantity is

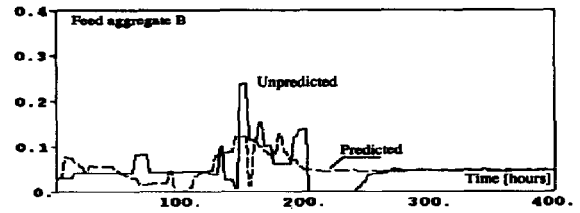


Fig. 29. Optimized feed aggregate B (predicted and unpredicted disturbances in available feed aggregate A).

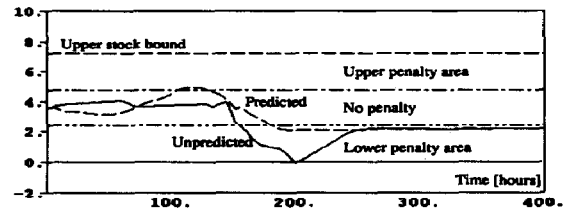


Fig. 30. Stock content feed aggregate B (predicted and unpredicted disturbances in available feed aggregate A).

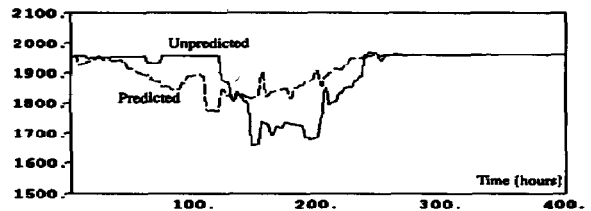


Fig. 31. Net profit rate (NOK/h) (predicted and unpredicted disturbances in available feed aggregate A).

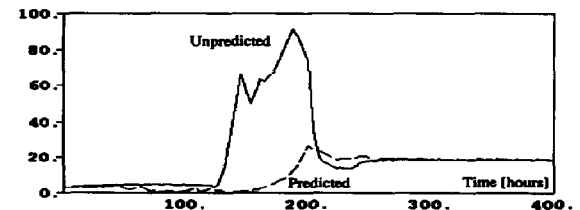


Fig. 32. Penalty cost (NOK/h) (predicted and unpredicted disturbances in available feed aggregate A).

that being optimized. This clearly illustrates that it is favorable to take predictable disturbances into account in the optimization.

Computational time consumption on the SPARC station makes the SSPC applicable to the real process even when all process units are included in the optimization (Strand, 1991).

CONCLUSIONS

The well-known MPC concepts are extended to account for severe process nonlinearities and general

constraints on process variables. This is achieved by introducing the SSPC strategy. A further advantage with the SSPC compared to earlier MPC strategies is the great flexibility in defining the optimality criterion.

Well-known theoretical and practical results from the areas of state-space modeling, model identification and optimal control are applied in developing the SSPC scheme for process control. SSPC is probably the most "model-based" process control system ever suggested. Thus, the initial modeling, which is mostly based on physical process knowledge and identification phases, must be carefully accomplished and might be quite laborious for some processes. When it comes to implementation, it is necessary to include an on-line estimation scheme capable of correcting unmeasured state variables and estimating the chosen set of unknown model parameters. This can be done by applying a recursive scheme such as AKF or by more time consuming, but consequently more robust, methods like direct on-line implementation of the maximum likelihood estimation strategy. Clearly, the long-range predictive capabilities of the model and the identifiability of the parameter set must be stressed.

The prediction model is used in an open-loop optimal control calculation. This can be done by control vector parameterization (CVP) combined with single or multiple shooting or even state polynomial approximation and collocation to solve the set of first-order differential equations. If constraints on state, control or output variables are present, the resulting problem is generally most efficiently solved by a sequential quadratic programming algorithm. The actual CVP was found to be an important tuning factor in the SSPC strategy.

The optimality criterion does not at all have to be quadratic; instead it could reflect the economical as-

pects of the process operation. When a criterion based on net profit rate is used, some of the process variables will often approach their limits. Penalties then have to be introduced in the optimality criterion to maintain enough controller flexibility to handle process disturbances.

The time delay in the open-loop optimal control calculation requires that a feedback from the deviation between the nominal predicted state variables and the real-time estimated state variables has to be introduced to account for the unmodeled disturbances. Optimal feedforward from predictable disturbances is achieved by taking them into account in the optimal control calculation.

On the FCC example both modeling and identification concepts are illustrated. The advantages of using SSPC as a set-point controller are quite clear compared to the conventional single-loop PID controllers. Further, the simulations demonstrate the possibility of computing the controls from an optimality criterion established from economical considerations. The light-metal electrolysis simulation highlights the inherent constraint handling capacity of the SSPC, and shows the improved performance gained from taking predictable disturbances into account in the open-loop optimization.

Even though the methods combined into the SSPC strategy have been known for quite a long time, no implementation of such a scheme is known. The constant increase in computing capacity makes it reasonable to suppose that the SSPC will be a future control strategy. The computer time consumptions in the FCC and the light-metal electrolysis examples were small enough to allow for SSPC in these processes.

Acknowledgement—This research has been sponsored by the Royal Norwegian Council for Scientific and Industrial Research (NTNF) through the MIP program.

NOTATION

List of abbreviations

AKF	augmented Kalman filter
CVP	control vector parameterization
DMC	dynamic matrix control
FCC	fluid catalytic cracking
LQ	linear quadratic
QDMC	quadratic DMC
MAC	model algorithmic control
MINLP	mixed integer nonlinear program(ming)
MLE	maximum likelihood estimate
MPC	model-predictive control
NLP	nonlinear program(ming)
ODE	ordinary differential equation
RPEM	recursive prediction error method
SQP	sequential quadratic programming
SSPC	state-space predictive control

FCC notation

		Nominal values
C_{cat}	weight fraction of coke deposited on catalyst, kg/kg	0.0069
COR	catalyst to oil weight ratio, kg/kg	6.98
C_{rc}	weight fraction of coke on regenerated catalyst, kg/kg	0.00385
C_{sc}	weight fraction of coke on spent catalyst, kg/kg	0.01075
c_{pa}	heat capacity of air, kJ/kg K	1.074
c_{pd}	heat capacity of dispersing stream, kJ/kg K	1.9
c_{po}	heat capacity of oil, kJ/kg K	3.1335
c_{ps}	heat capacity of catalyst, kJ/kg K	1.005
CO_1	constant factor in catalyst cost model, NOK/min	50.0
CO_2	cost factor regenerator temperature in catalyst cost model, NOK/K ² min	0.03125
Co_{cat}	cost of catalyst consumption, NOK/min	
E_{cf}	activation energy for coke formation, kJ/mol	41.79
E_{cb}	activation energy for coke burning reaction, kJ/mol	158.6
E_f	activation energy for gas oil cracking, kJ/mol	101.5
F_a	mass flow rate of air to regenerator, kg/min	1543.6
F_{oil}	mass flow rate of gas oil feed, kg/min	2438.0
F_{rc}	mass flow rate of regenerated catalyst, kg/min	17,023.9
F_{sc}	mass flow rate of spent catalyst, kg/min	17,023.9
F_{gl}	gasoline yield factor of catalyst	1.0
h_1	parameter in the approximation for ΔH , kJ/kmol	521,150.0
h_2	parameter in the approximation for ΔH , kJ/kmol K	245.0
I_{gl}	gasoline cracking intensity	0.9
k_c^1	rate constant for catalytic coke formation, s ^{-1/2}	0.01897
k_{com}	rate constant for coke burning, 1/min	29.338
k_0	rate constant for gas oil cracking, 1/s	96,200.0
K_r	approximate reaction rate for gas oil cracking at z ,	
$L1, L2$	predictive control criteria	
M_c	molecular weight of coke, kg/kmol	14.0
M_{car}	molecular weight of carbon, kg/kmol	12.0
m	empirical deactivation parameter	80.0
N	empirical exponent in coke formation equation	0.4
n	hydrogen content in coke	2.0
O_d	oxygen mole fraction in regenerator dense bed	0.00472
O_d^0	set point for the mole fraction	0.00472
O_{in}	oxygen mole fraction in air	0.2136
P_1	penalty factor flue gas temperature, NOK/K min	2.0
P_{Tfg}	penalty on afterburning rate/flue gas temperature. NOK/min	
P_{ri}	weight factor on quadratic riser outlet temperature deviation	0.01/0.0
P_O	weight factor on quadratic dense bed oxygen fraction deviation	100,000.0
P_{rg}	weight factor on quadratic dense bed temperature deviation	0.01/0.0
Pr_0	price of gasoline, NOK/kg	1.34
Pr_1	price of feed, NOK/kg	0.84
Pr_2	price of light products, NOK/kg	0.70
Pr_3	price of feed preheat, NOK/kg K	0.00017
Pr_4	price of oxygen concentration for regularization, NOK/min	800.0
Pr_5	price of inlet air rate to regenerator, NOK/kg	0.0
R	universal gas constant, kJ/(mol K)	0.0083143
R_a	molar flow rate of air to regenerator, kmol/min	53.50
R_r	recycle ratio	0.0
T_a	temperature of air to regenerator, K	320.0
T_{oil}	temperature of gas oil feed, K	420.0
T_{rg}	temperature of catalyst in regenerator dense bed, K	972.1
T_{rg}^0	set point for regenerator dense bed temperature, K	972.1
$T_{ri}(0)$	temperature of catalyst and gas oil mixture at riser inlet, K	799.0
T_{ri}^0	set point for the riser outlet temperature, K	774.9
$T_{ri}(1)$	temperature of catalyst and gas oil mixture at riser outlet, K	774.9

T_r	approximate riser temperature at z_r , K	
t	time, min	
t_c	catalyst residence time in riser, s	9.6
W	catalyst holdup in regenerator, kg	175,738.0
W_a	air holdup in regenerator, kmol	20.0
$y_f(0)$	weight fraction of gas oil in feed	1.0
$y_f(1)$	weight fraction of gas oil in product	0.5137
$y_g(1)$	weight fraction of gasoline in product	0.3532
z_r	dimensionless position for average reaction rate in riser	0.0555

Greek letters

α	catalyst decay rate constant, 1/s	0.12
ΔH	heat of reaction of coke burning, kJ/kmol	
ΔH_f	heat of reaction gas oil cracking, kJ/kg	506.0
λ	weight fraction of steam in feed stream to riser, kg/kg	0.035
σ	CO ₂ /CO ratio in regenerator dense bed	
σ_2	linear CO ₂ /CO dependence on the temperature, 1/K	0.006244
ϕ	catalyst decay function	
ϕ_0	activity of regenerated catalyst	

REFERENCES

- Balchen, J. G., Ljungquist, D. and Strand, S., 1988, Predictive control based upon state space models. *Proceedings of the 1988 ACC*, Vol. 3, pp. 2174–2179.
- Balchen, J. G., Ljungquist, D. and Strand, S., 1989, State space predictive control of a multistage electrometallurgical process, in *Proceedings of the IFAC Workshop on Model Based Process Control*, Atlanta, Georgia, June 1988, pp. 47–53. Pergamon Press, Oxford.
- Balchen, J. G. and Mumme, K. I., 1988, *Process Control: Structures and Applications*, pp. 369–375. Van Nostrand Reinhold, New York.
- Biegler, L. T., 1984, Solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation. *Comput. chem. Engng* 8, 243–248.
- Bierman, G. J., 1977, *Factorization Method for Discrete Sequential Estimation*. Academic Press, New York.
- Bock, H. G., 1983, Recent advances in parameter identification techniques for O.D.E., in *Numerical Treatment of Inverse Problems in Differential and Integral Equations* (Edited by P. Deufhard and E. Hairer), pp. 95–121. Birkhäuser, Boston.
- Bock, H. G. and Plitt, K. J., 1985, A multiple shooting algorithm for direct solution of optimal control problems, in *Proceedings of the 9th Triennial World Congress of IFAC*, Budapest 1984, Vol. 3, pp. 1603–1608 Pergamon Press, Oxford.
- Bromley, J. A. and Ward, T. J., 1981, Fluidized catalytic cracker control. A structural analysis approach. *Ind. Engng Chem. Process Des. Dev.* 20, 74–81.
- Cuthrel, J. E., and Biegler, L. T., 1987, On the optimization of differential-algebraic process systems. *A.I.Ch.E. J.* 33, 1257–1270.
- Cutler, C. R. and Ramaker, B. L., 1979, Dynamic matrix control—a computer control algorithm. AIChE 86th National Meeting, Houston, TX, April 1979.
- Dormand, J. R. and Prince, P. J., 1980, A family of embedded Runge–Kutta formulae. *J. Comput. appl. Math.* 6, 19–26.
- Duran, M. A. and Grossmann, I. E., 1986, An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Math. Prog.* 36, 307–339.
- Edge, E. R. and Powers, W. F., 1976, Function-space quasi-Newton algorithms for optimal control problems with bounded controls and singular arcs. *J. Optimization Theory Applic.* 20, 455–479.
- Errazu, A. F., deLasa, H. I. and Sarti, F., 1979, A fluidized bed catalytic cracking regenerator model, grid effects. *Can. J. chem. Engng* 57, 191–197.
- Floudas, C. A., Aggrawal, A. and Ciric, A. R., 1989, Global optimum search for nonconvex NLP and MINLP problems. *Comput. chem. Engng* 13, 1117–1132.
- Gelb, A., 1974, *Applied Optimal Estimation*. The Analytic Science Corporation, MA.
- Gill, P. E., Murray, W., Saunders, M. A. and Wright, M. H., 1983, User's guide for SOL/NPSOL: a Fortran package for nonlinear programming. Technical Report SOL 83-12, Department of Operations Research, Stanford University, California.
- Hernández, E. and Arkun, Y., 1990, Neural network modeling and an extended DMC algorithm to control nonlinear systems, in *Proceedings of the 1990 ACC*, pp. 2454–2459.
- Hicks, R. C., Worell, G. R. and Durney, R. J., 1966, Analog–digital cat-cracking-model development. *Oil Gas J.* 64, 97–105.
- Huang, Z., Sapre, A. V. and Tsiligiannis, C., 1989, Dynamic modeling and control of fluidized catalytic cracking units. AIChE Annual Meeting, paper 22c, San Francisco, November 1989.
- Isles-Smith, P. C., 1985, A study of the modelling, dynamic simulation and control of the modern fluid catalytic cracking process. Ph.D. thesis, University of Leeds.
- Jang, S. S., Joseph, B. and Muka, H., 1987, On-line optimization of constrained multivariable chemical processes. *A.I.Ch.E. J.* 33, 26–35.
- Jazwinski, A. H., 1970, *Stochastic Processes and Filtering Theory*. Academic Press, New York.
- Joseph, B., Jang, S. S. and Mukai, H., 1989, Integrated model based control of multivariable nonlinear systems, in *Proceedings of the IFAC Workshop on Model Based Process Control*, Atlanta, Georgia, June 1988, pp. 121–128. Pergamon Press, Oxford.
- Kelley, H. J., 1962, Method of gradients, in *Optimization Techniques* (Edited by G. Leitman). Academic Press, New York.
- Kurihara, H., 1967, Optimal control of fluid catalytic cracking processes. MIT Electric System Laboratory Report ESL-R-309.
- Lasdon, L. S., 1970, Conjugate direction methods for optimal control. *IEEE Trans. autom. Control* 15, 267–268.
- Lasdon, L. S., Mitter, S. K. and Waren, A. D., 1967, The conjugate gradient method for optimal control problems. *IEEE Trans. autom. Control* 12, 132–138.
- Lee, E. and Groves, F. R. Jr., 1985, Mathematical model of the fluidized bed catalytic cracking plant. *Trans. Soc. Comput. Sim.* 2, 219–236.

- Lee, W. and Weekman, V. W. Jr., 1976, Advanced control practice in the chemical process industry: a view from industry. *A.I.Ch.E. J.* **22**, 31–34.
- Ljung, L., 1979, Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems. *IEEE Trans. autom. Control* **25**, 36–50.
- Ljung, L., 1987, *System Identification, Theory for the User*. Prentice-Hall, Englewood Cliffs, NJ.
- Ljungquist, D., 1990, Online estimation in nonlinear state-space models with application to catalytic cracking. Dr. Ing. thesis. Report no. 90-89-W, Division of Engineering Cybernetics, NTH, Trondheim.
- Luyben, W. and Lamb, D. E., 1963, Feed-forward control of a fluidized catalytic cracker reactor-regenerator system. *Chem. Engng Prog. Symp. Ser.* **59**, 165.
- Maybeck, P. S., 1979, *Stochastic Models, Estimation and Control 1*. Academic Press, New York.
- Maybeck, P. S., 1982, *Stochastic Models, Estimation and Control 2*. Academic Press, New York.
- Morari, M. and Lee, J. H., 1991, Model predictive control: the good, the bad, and the ugly. Preprints of CPC IV, Fourth International Conference on Chemical Process Control, 17–22 February 1991, South Padre Island, TX.
- Neuman, C. P. and Sen, A., 1973, A suboptimal control algorithm for constrained problems using cubic splines. *Automatica* **9**, 601–613.
- Nørsett, S. P. and Thomsen, P. G., 1987, *User Guide for SIMPLE—a Stiff System Solver*. Division of Numerical Math., NTH, Trondheim.
- Pagurek, B. and Woodside, C. M., 1968, The conjugate gradient method for optimal control problems with bounded control variables. *Automatica* **4**, 337–349.
- Patwardhan, A. A., Rawlings, J. B. and Edgar, T. F., 1988, Model predictive control of nonlinear processes in the presence of constraints. AIChE Annual Meeting, Session 129, paper 129b. Washington, D.C.
- Peterson, T., Hernández, E., Arkun, Y. and Schork, F. J., 1989, Nonlinear predictive control of a semi batch polymerization reactor by an extended DMC, in *Proceedings of the 1989 ACC*, Pittsburgh, PA, pp. 1534–1539.
- Poljak, B. T. and Tsytkin, J. Z., 1980, Robust identification. *Automatica* **16**, 53–63.
- Pontryagin, L. S., Boltianskii, V. G., Gamkrelidze, R. V. and Mishchenko, E. F., 1962, *The Mathematical Theory of Optimal Processes*. Wiley, New York.
- Quintana, V. H. and Davison, E. J., 1974, Clipping-off gradients algorithms to compute optimal controls with constrained magnitude. *Int. J. Control* **20**, 243–255.
- Renfro, J. G., Morshedi, A. M. and Asbjornsen, O. A., 1987, Simultaneous optimization and solution of systems described by differential/algebraic equations. *Comput. Chem. Engng* **11**, 503–517.
- Rhemann, H., Schwarz, G., Badgwell, T. A., Darby, M. L., and White, D. C., 1989, On-line FCCU advanced control and optimization. *Hydrocarb. Process.* **68**, 64–71.
- Richalet, J. A., Rault, A., Testud, J. D. and Papon, J., 1978, Model predictive heuristic control: applications to industrial processes. *Automatica* **14**, 413–428.
- Sargent, R. W. H. and Sullivan, G. R., 1978, The development of an efficient optimal control package, in *Optimization Techniques, Proceedings of the 8th IFIP Conference on Optimization Techniques* (Edited by J. Stoer), Wurzburg 1977. Springer, Berlin.
- Schittkowski, K., 1985, NLPQL: a Fortran subroutine solving constrained nonlinear programming problems. *Operations Res. Ann.* **5**, 485–500.
- Schmid, C. and Biegler, L. T., 1990, Application of multistep Newton-type controllers to fluid catalytic cracking, in *Proceedings of the 1990 ACC*, pp. 581–586.
- Schuldt, S. B. and Smith, F. B., Jr., 1971, An application of quadratic performance synthesis techniques to a fluid cat cracker. Joint Automatic Control Conference paper 3-E4, St. Louis, MO, pp. 270–276.
- Shah, Y. T., Huling, G. P., Paraskos, J. A. and McKinney, J. D., 1977, A kinematic model for an adiabatic transfer line catalytic cracking reactor. *Ind. Engng Chem. Process Des. Dev.* **16**, 89–94.
- Sist, P. B. and Bequette, B. W., 1990, Process identification using nonlinear programming techniques, in *Proceedings of the 1990 ACC*, pp. 1534–1539.
- Strand, S., 1991, Dynamic optimization in state-space predictive control schemes. Dr. Ing. thesis. Report 91-11-W, Division of Engineering Cybernetics, NTH, Trondheim.
- Strand, S. and Balchen, J. G., 1990, A comparison of constrained optimal control algorithms. Preprints of the 11th IFAC World Congress, 6, pp 191–199, Tallinn, USSR.
- Tripathi, S. S. and Narendra, K. S., 1970, Optimization using conjugate gradient methods. *IEEE Trans. autom. Control* **15**, 268–270.
- Tsang, T. H., Himmelblau, D. M. and Edgar, T. F., 1975, Optimal control via collocation and nonlinear programming. *Int. J. Control* **21**, 763–768.
- Vlassenbroeck, J. and van Dooren, R., 1988, A Chebyshev technique for solving nonlinear optimal control problems. *IEEE Trans. autom. Control* **33**, 333–340.
- Weekman, V. W., Jr. and Nace, D. M., 1970, Kinetics of catalytic cracking selectivity in fixed, moving and fluid bed reactors. *A.I.Ch.E. J.* **16**, 397–404.
- Wiener, N., 1942, The extrapolation, interpolation and smoothing of stationary time series. NDRC Report. Cambridge, MA.