Computer Control of a Butane Hydrogenolysis Reactor: Part I. State Space Reactor Modeling," *AIChE J.*, **23**, No. 5, 732 (1977a).
Jutan, A., J. F. MacGregor, and J. D. Wright, "Multivariable Computer Control of a Butane Hydrogenolysis Reactor: Part II. Data Collection, Parameter Estimation and Stochastic Disturbance Identification," *AIChE J.*, **23**, No. 5, 742 (1977b).
Jutan, A., J. D. Wright, and J. F. MacGregor, "Multivariable Computer Control of a Butane Hydrogenolysis Reactor: Part III. On-Line Linear Quadratic Stochastic Control Studies," *AIChE J.*, **23**, No. 5, 751 (1977c).
Khanna, R., and J. H. Seinfeld, "Model Development of a Non-Adiabatic Packed-Bed Reactor," 1982 AIChE Annual Meeting, Los Angeles (Nov. 1982).
MacGregor, J. F., and A. K. L. Wong, "Multivariate Model Identification and Stochastic Control of a Chemical Reactor," *Technometrics*, **22**, No. 4 (1980).
Silva, J. M., P. H. Wallman, and A. S. Foss, "Multibed Catalytic Reactor Control Systems: Configuration Development and Experimental Testing," *I. and E.C. Fund.*, **18**, No. 4, 383 (1979).
Sorensen, J. P., "Experimental Investigation of the Optimal Control of a Fixed Bed Reactor," *C.E.S.*, **32**, 763 (1977).
Wallman, P. H., J. M. Silva, and A. S. Foss, "Multivariable Integral Controls for Fixed-Bed Reactors," *I and E.C. Fund.*, **18**, No. 4, 392 (1979).

Manuscript received June 29, 1982; revision received March 25, and accepted March 28, 1983.

Optimal Operation of Integrated Processing Systems

Part II: Closed-Loop On-Line Optimizing Control

C. E. GARCIA and MANFRED MORARI

Chemical Engineering Department
University of Wisconsin
Madison, WI 53706

A method for tracking the economically optimal operating conditions of a chemical process in the presence of constraints is developed. The technique is based on an on-line search rather than a fundamental model. The most profitable operating point is found by fitting a dynamic model of the process based on data obtained from experimental moves on the plant.

This model is used to compute gradients of the economic objective and of the constraints so that a direction of economic improvement inside the allowed operating region of the plant is always obtained. Constraint violations during the transients are prevented by a multivariable regulator. A new regulation method (Internal Model Control) is used which permits explicit handling of constraints and which can be made robust against modelling errors. This combined optimization/regulation approach is tested in a demonstrative simulation example and shown to be reliable for following a moving optimum and safely handling complex constraint moves.

SCOPE

In the presence of disturbances, the optimal operating point of a process usually shifts from one set of active constraints to another. To keep the operation close to the constraints without violating them, regulatory controllers are generally employed. A standard approach for transfering the operation from one point to another is to switch between different sets of Single Input-Single Output (SISO) loops to tie the constraints which become active (Arkun, 1979). However, due to dynamic interactions and plant changes the SISO loops are difficult to tune and likely to become unstable, rendering the switching procedure unreliable.

In this paper a multivariable control scheme is suggested which considers all constraints, outputs and manipulated variables simultaneously. Since it takes into account all interactions, the transfer of operating points occurs smoothly. The scheme also allows to enhance the robustness in the face of changes in the process. This regulator interfaces with the adaptive on-line optimization algorithm introduced in Part I of this series (Garcia and Morari, 1981) to produce a method which addresses the overall control requirements of a process. A simulation example shows the ability of the scheme to safely track the optimal operating point of a plant in the event of nontrivial constraint shifts.

CONCLUSIONS AND SIGNIFICANCE

A new optimizing controller for plants with constraints was developed. A simulation example of a benzene production plant was used to demonstrate the ability of the scheme to keep the operation at its economically optimal conditions as constraints are shifted due to unmeasured disturbances. The on-line optimizer based on an identified dynamic model of the plant was shown to exhibit fast speed of tracking and to be reliable to

Correspondence concerning this paper should be addressed to M. Morari.

determining optima inside the allowable operating region. The multivariable regulator used for constraint handling exhibited superb performance in keeping the operation close to hard constraints without incurring violations. In addition, the robustness of the regulator in the face of plant changes was clearly demonstrated.

The tolerance against modelling errors and reliability in determining optimal operating points and implementing them safely with minimal *a-priori* process information make this overall plant control method superior to conventional approaches and a unique approach for handling complex industrial control problems.

INTRODUCTION

As a result of soaring energy costs, the most profitable operation of a process is often achieved at the operating point of lowest utility costs allowed by equipment limitations or by other operation constraints. As disturbances affect the process, these constraints shift, changing the optimal operating point or forcing the operation outside the allowed operating region. In the face of these upsets, the objective of plant control is to maintain an economically optimal operation while avoiding constraint violations. This safe tracking of the optimal operating point of the process when disturbances occur is termed optimizing control.

Due to long optimization horizons, only lasting disturbances of periods larger than the process settling time are important. Therefore, the plant can be assumed to be at a quasisteady state for *optimization* purposes. On the other hand, short-term upsets do have an impact on system transients and can drive the operation beyond constraints. For this reason, *regulation* is used to prevent the system from violating constraints and thus ensure operation safety.

In Part I of this series (Garcia and Morari, 1981) an algorithm for tracking the optimal operating point of a plant was developed. The method did not make use of fundamental models for computing the optimal inputs, but rather a model was adapted on line based on experiments performed on the plant. As key features, the algorithm was shown to exhibit fast speed of tracking and to be insensitive to significant measurement noise.

However, only plants with unconstrained operating regions were considered. In this paper we combine the adaptive optimization algorithm with a multivariable regulator for constraint handling. For a regulation scheme to be useful for this purpose, it should exhibit excellent dynamic characteristics to be able to reject short-term disturbances while being robust to changes in the plant during the search. A new regulatory control concept suitable for this application is used here for the design of such a regulator. This combined optimization/regulation scheme proves to be an attractive tool for industrial applications where operation safety and profitability are guaranteed with minimal process information.

PROBLEM FORMULATION: OPEN- VS. CLOSED-LOOP OC

The steady-state optimum of a process is given by the solution of

$$\begin{aligned} &\min_{m} P(y,m) \\ &f(y,m) = 0 \\ &g(y,m) \le 0 \\ &y = y(m,d) \end{aligned} \tag{1}$$

where

P = economic objective of operation
y = system outputs (assume measurable)
m = manipulatable system inputs
d = disturbances to plant
f = hard constraints which must be satisfied at all times (i.e., "equality" constraints)
g = inequality constraints, which determine the limits of operating region

It is assumed that the funtions $f(\cdot)$ & $g(\cdot)$ are known and that their instantaneous value can be either directly measured or indirectly determined from process measurements.

The job of the optimizing controller (OC) is to find the inputs m which solve problem (Eq. 1) for a specific disturbance level and drive the system from the present operating point to the new one without violating the constraints. This search should be done rapidly before the next set of disturbances affects the system.

To avoid violations during the transients, a regulator is commonly employed for implementing the operating points. The objective of this regulator is to control a function $r(y_1,m)$ at a setpoint r_s such that the satisfaction of the equality and inequality constraints is implied

$$\{r(y_1,m) = r_s\} \rightarrow \begin{cases} f(y,m) = 0 \\ g(y,m) = \gamma \le 0 \end{cases} \tag{2}$$

where

$$\dim m = \dim r. \tag{3}$$

We have partitioned the vector y into y_1, the controlled and y_2, the uncontrolled measured outputs. The resulting structure for the OC with regulation is shown in Figure 1, where the optimization variables m have been substituted by the controlled variables r.

Depending on the choice of r, different types of OC arise.

Open-Loop OC

Selecting $r = m$, the optimizer prescribes the optimal inputs m to the process directly. This approach is risky because when fast disturbances affect the process, the slow optimizer may not be able to correct m in time to avoid serious constraint violations. In addition, constraint satisfaction hinges on the accuracy of the plant model available to the optimizer. Introduction of feedback can alleviate these drawbacks.

Closed-Loop OC

In the presence of fast process disturbances and modelling errors, regulatory control of certain outputs is needed to ensure a safe operation. In particular, selecting

$$r(y,m) = \begin{bmatrix} r_f(y,m) \\ r_g(y,m) \end{bmatrix} = \begin{bmatrix} f(y,m) \\ g(y,m) \end{bmatrix} \tag{4}$$

Steady-state satisfaction of constraints is obtained by having the optimizer specify.

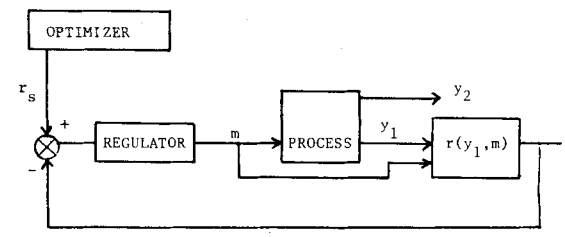

Figure 1. Structure of optimizing controller.

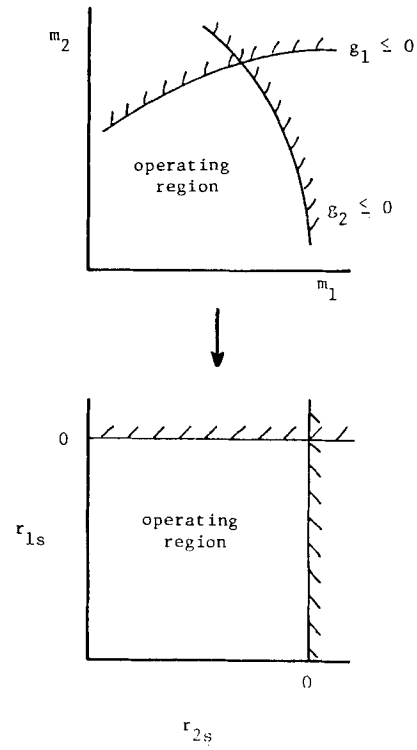

Figure 2. Modification of optimization search space due to use of a subordinate regulator.

$$r_s = \begin{bmatrix} 0 \\ \gamma \end{bmatrix}, \quad \gamma \leq 0 \tag{5}$$

Such selection of r (denoted as *constraint control*) is advantageous for two reasons:

- Since it is generally possible with minimal modelling to design a controller which guarantees $r = r_s$, satisfaction of constraints and offset-free behavior is achieved despite model ignorance.
- By having the regulator act more frequently than the optimizer, fast disturbances can be rejected and thus violations during the transients avoided.

However, for many applications, the number of constraints usually exceeds the number of availale manipulatable inputs. Let us consider two cases.

i) $dim\ m \leq dim\ f + dim\ g$. If sufficient inputs are available to control all of the constraints, the optimizer solves

$$\begin{aligned} &\min_{r_s} P[y_2, r_s] \\ &\text{s.t. } f = r_{fs} = 0 \\ &\quad g = g_{gs} \leq 0 \\ &\quad y_2 = y_2(r_s, d) \end{aligned} \tag{6}$$

where f,g are independent of d. Note that the function $y_2(R_s,d)$ must be supplied to the optimizer to solve Eq. 6. We also note that the search region is transformed from having general nonlinear boundaries in m (problem 1) to a linear polytope (Figure 2). This allows the use of efficient constrained search algorithms for on-line OC. A detailed discussion is given in a later section.

ii) $dim\ m < dim\ f + dim\ g$. For this case, not enough degrees of freedom exist to fix all f and g by regulation. It is assumed here that at least

$$\dim m \leq \dim f$$

so that the equality constraints are always controlled. The procedure to handle the inequality constraints is to regulate a subset of the g, leaving g_u constraints uncontrolled. Then the optimizer solves

$$\begin{aligned} &\min_{r_s} P(y_2, r_s) \\ &\text{s.t. } r_{fs} = 0 \\ &\quad r_{gs} \leq 0 \\ &\quad g_u(y_2, r_s) \leq 0 \\ &\quad y_2(r_s, d) = y_2 \end{aligned} \tag{7}$$

where both P and the constraints g_u are functions of the unregulated outputs. This implies that g_u is not disturbance independent and therefore $y_2(\cdot)$ must be known by the optimizer in order to prescribe points inside the allowed operating region.

A common method for dealing with this case where the number of constraints exceeds the number of manipulated variables is to have dim m SISO loops regulating a subset of g. Should an element of g_u become active, a loop or a set of loops of the original set is broken to tie the new active constraint (Arkun, 1979). This method requires careful sequencing of setpoint changes to transfer the operation to a new set of boundaries. Besides the inconvenience of having to redefine the control structure essentially for each disturbance level, this scheme could fail due to performance degradation of SISO loops as a result of interactions and plant changes. The multivariable regulation scheme to be introduced later proves to be a superior alternative to the constraint handling problem.

Having considered all possible cases and defined the corresponding problems to be solved by the on-line optimizer, we now proceed to present the algorithms to be used by the optimizer and regulator.

ALGORITHM FOR CLOSED-LOOP OPTIMIZING CONTROL

Most closed-loop OC schemes reported in the literature (e.g., Prett and Gillette, 1979; Duncanson and Youle, 1970; Shah and Stillman, 1970) employ fundamental models to simulate the steady-state characteristics of the system. Having a measurement of d, problem 1 is solved for m, determining the constraints g which are active. Then the lower layer regulators are tuned to control these constraints.

This *off-line* method suffers from two drawbacks. First, the model at hand might not be precise enough to allow an accurate computation of the optimum. Secondly, in many situations process disturbances are unmeasurable or unpredictable preventing the determination of constraint position and/or optimal point.

The *on-line* OC algorithm in Part I employs an identified model of the process which is updated based on experimental plant data. Adaptation of the model makes it insensitive to unmeasurable disturbances, while the model is improved as more data are gathered. This algorithm has been extended to the constrained case by Bhattacharya and Joseph (1982). However, the operating points are prescribed in an open-loop fashion and consquently safety problems could arise if implemented in an industrial environment as explained above.

In the following section an algorithm for finding optimal operating points in the face of constraints is discussed. Then the new multivariable regulation scheme for constraint handling is presented.

Optimization of Constrained Plants via System Identification

Since in any on-line OC scheme the search is performed directly on the process, the optimization routine used must produce optimization variables which lie inside the allowed region of operation (i.e., "feasible") at each iteration. However, the most efficient gradient algorithms of nonlinear programming do not exhibit this property and can only ensure feasibility at an optimum. A detailed discussion on convergence properties of constrained optimization algorithms lies beyond the scope of this paper. Good references on this subject include the review by Sargent (1980) and other classical nonlinear programming texts (Avriel, 1976; Bazaraa and Shetty, 1979).

As discussed above, when constraints are regulated the original nonlinear problem 1 is generally transformed to a linearly constrained problem (Figure 2). The following algorithm takes maximum advantage of this property and guarantees feasibility at each iteration (Zoutendijk, 1960; Mangasarian, 1969).

Let the problem to be solved be

$$\min_{m_0} P(m_0)$$
$$\text{s.t. } g(m_0) \leq 0 \tag{8}$$

where m_0 is the vector of independent decision variables; for the OC we usually have $m_0 = r$. At an operating point $(m_0)_l$, a direction q_l which points towards the interior of the feasible region and decreases the objective P, is given by the solution of the following linear program:

$$\min_{\epsilon,q} \epsilon$$
$$\text{s.t. } q^T \nabla_{m_0} P((m_0)_l) < \epsilon$$
$$q^T \nabla_{m_0} g_J((m_0)_l) < \epsilon\, e \tag{9}$$
$$|q_i| < 1 \quad i = 1,\dim(m_0)$$
$$e = (1\,1 \ldots 1)^T$$

where

$\nabla_{m_0} P$ = gradient of objective function
$\nabla_{m_0} g$ = Jacobian matrix of constraints
q = constrained to avoid unbounded solutions
J = $\{j \mid -\sigma_1 < g_j((m_0)_l) \leq 0,\ \sigma_1 > 0\}$ = set of σ-active constraints at $(m_0)_l$

The new operating point $(m_0)_{l+1}$ is found by selecting a stepsize μ_l along q_l, which is usually chosen as the maximum positive value μ such that

$$(m_0)_{l+1} = (m_0)_l + \mu q_l \tag{10}$$

lies inside the allowed operating region. The existence of a q_l and μ_l is guaranteed if g_J satisfies certain qualifications (Mangasarian, 1969), which we assume here to be fulfilled.

To implement this algorithm, the gradients of P and g need to be known. As in Part I, a dynamic model of the plant is identified and from it the required steady-state information is extracted. The optimizer will then solve Eq. 9 for q_l and find the stepsize μ_l in Eq. 10 based on the identified paerameters.

We should note here that when constraints are regulated (problem 6), the g_J used in Eq. 9 are truly linear and known. Therefore, only $\nabla_{m_0} P$ needs to be determined from the identified plant model. This minimizes the computational effort at the optimizer layer. But perhaps more importantly, since the constrained region is *exactly* known, the stepsize selection (Eq. 10) is a trivial task since we can always be sure the operating point is feasible.

On the contrary, whenever uncontrolled constraints need to be identified as in problem 7, only a linear approximation of g in the vicinity of $(m_0)_l$ is at hand. The linear program (Eq. 9) can be solved for q_l without difficulty. However, the stepsize selection is not so trivial since infeasibility may occur. Thus we suggest the following stepsize:

$$\mu_l = \max \mu\epsilon(\bar{\mu}, \bar{\mu}/2, \bar{\mu}/4, \ldots)$$
$$\text{s.t. } g((m_0)_l) + \nabla g((m_0)_l)(\mu q_l) < -\rho_l \tag{11}$$
$$\rho_l > 0$$

where ρ_l is a tolerance factor and $\bar{\mu}$ a maximum bound on the stepsize. How ρ_l is to be chosen depends on the quality of the identification and the severity of the nonlinearities in $g(m_0)$.

Multivariable Regulatory Control Algorithm

For the closed-loop OC proposed above to operate reliably in an industrial environment, it requires a regulatory controller of excellent dynamic characteristics for the implementation of setpoints. Particularly, the regulation layer is required to:

i) Prevent violation of constraints during the transients either by controlling the constraints tightly or by predicting a violation of uncontrolled constraints

ii) Be robust to changes in the plant due to the fact that the operating point is changed constantly by the optimizer

A new regulatory control design concept, Internal Model Control (IMC), has been developed (Garcia and Morari, 1982; Garcia, 1982; Garcia and Morari, 1983). IMC allows one to influence control quality and robustness to modelling errors directly in the design, making it the most suitable candidate for OC applications. Only the key aspects of the design procedure as applied to constraint handling and robustness are presented here. For details, the reader is referred to the above cited references.

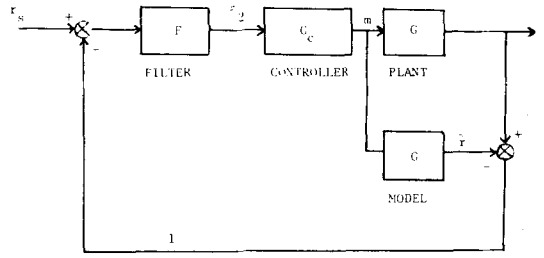

Figure 3. Internal model control structure.

The structure of the IMC regulator is given in Figure 3 where G denotes the controlled system transfer function matrix, G_c a predictive controller for constraint handling, F a filter to ensure robustness, and $\tilde{G}$ is the system model. Although this structure is completely equivalent to a standard feedback loop with controller

$$C = [I - G_c F \tilde{G}]^{-1} G_c F$$

the design of F and G_c is more transparent than that of C in terms of robustness and dynamic handling of constraints, respectively. Let us describe each design step.

Robustness (Design for F). From Figure 3 we observe that in case the model matches the plant ($G = \tilde{G}$) the feedback signal ϵ_1 is exactly zero. For the exact model case the scheme is essentially open-loop and F can be thought of as a reference trajectory for the setpoint y_{1s}. On the other hand, when $G \neq \tilde{G}$, ϵ_1 is not zero and F is visualized as an attenuator of the feedback signal to guarantee stability (Garcia and Morari, 1982).

One practical filter selection is the following

$$F = \text{diag}\,\frac{1-\alpha_i}{1-\alpha_i z^{-1}}, \quad 0 \leq \alpha_i < 1 \tag{12}$$

which is the z-transform of a first-order exponential filter. Under the assumption that the plant is open-loop stable and when a certain condition on the steady-state gain matrices of the system G and the model $\tilde{G}$ is satisfied, it can be shown (Garcia, 1982) that for a given model/plant mismatch ($G - \tilde{G}$) there exist an $\bar{\alpha}$ sufficiently large such that if $\alpha_i > \bar{\alpha}$ the closed-loop system is stable. Physically, by increasing α_i the bandwidth of the low pass exponential filter is decreased sufficiently enough to attenuate ϵ_1.

Dynamic Performance/Constraint Handling (Design for G_c). To prevent constraint violations the controller must be able to predict encroachments and update inputs m accordingly. The following predictive control formulation is used (Garcia, 1982):

$$\min_{\substack{m_1(\bar{k}),m_1(\bar{k}+1),\\ \ldots m_1(\bar{k}+M-1)}} \sum_{l=1}^{P} \|\epsilon_2(\bar{k}+l) - \tilde{r}(\bar{k}+l)\|_{\Gamma_l} + \|m(\bar{k}+l-1)\|_{B_l}$$
$$\text{s.t. } \tilde{r}(\bar{k}+l) = \sum_{i=1}^{N} H_i m(\bar{k}+l-i) \tag{13}$$

where

$\|X\|_Q$ = $X^T Q^T Q X$,
P = controller horizon
M = number of allowed input moves ($1 \le M \le P$)
ϵ_2 = filtered error prediction (Figure 3)
Γ_l, B_l = weighting matrices
$\tilde{r}$ = a model prediction of controlled outputs

In this formulation we assume $\tilde{r}$ to be given by an impulse response model of truncation order N with coefficient matrices H_i.

Solution of this problem at the present time $\bar{k}$ produces a sequence of inputs moves $m(\bar{k}), m(\bar{k}+1) \ldots m(\bar{k}+M-1)$ into the future. This computed sequence can be used to predict $\tilde{r}$ over a horizon M and thus prevent any future constraint violation. We note that for the case where some constraints are not regulated (problem 7), the model for y_2 should also be made available to the controller to prevent violations of g_u. Two approaches can be used for handling violations:

Indirect Method. In the event of a predicted encroachment at time $\bar{k}$ the input weights B_l can be updated on line and the control sequence m recomputed until the predicted outputs do not indicate a violation. This approach has been used successfully by Shell Oil Co. in the OC of a fluid cat cracker (Prett and Gillette, 1979). They employed a multivariable control scheme denoted by Dynamic Matrix Control (DMC) (Cutler and Ramaker, 1979) which can be shown to be an IMC controller with special tuning parameters in problem 13 (Garcia, Morari, 1982).

Direct Method. A more direct approach to this problem involves the explicit inclusion of constraints in the formulation of problem 13, as follows

$$\min_{m(\bar{k}), \ldots m(\bar{k}+M-1)} \sum_{l=1}^{P} \|\epsilon_2(\bar{k}+l) - r(\bar{k}+l)\|_{\Gamma_l} + \|m(\bar{k}+l-1)\|_{B_l}$$

$$\text{s.t.} \begin{bmatrix} \tilde{r} \\ \tilde{y}_2 \end{bmatrix} = \sum_{i=1}^{N} H_i' m(\bar{k}+l-i) \qquad (14)$$

$$\tilde{r}_f(\bar{k}+l) = 0 \quad l = 1, \ldots P$$

$$\tilde{r}_g(\bar{k}+l) \le 0$$

$$g_u(\tilde{r}(\bar{k}+l), \quad \tilde{y}_2(\bar{k}+l)) \le 0$$

which requires a model for outputs y_2. This problem has a quadratic objective with general nonlinear constraints g_u, and is consequently not trivial to solve. However, for the case that $g_u(\ )$ is linear, problem 14 is a quadratic program (QP) solvable by standard routines.

Since Eq. 14 has to be solved on line, computation time could be a limiting factor. On-line computations can be reduced by minizing a linear objective function instead. Then problem 14 becomes a linear program (LP).

Some variations of the QP and LP control computation in the face of constraints have appeared in the literature (Mehra et al., 1982; Brosilow, 1981) but they do not include the filter which we regard to be essential for on-line robustness adjustments. On the other hand, the algorithm by Chang and Seborg (1981). though using an LP, does not employ the IMC structure and can therefore suffer from instability problems when the inputs are constrained.

IMPLEMENTATION OF THE OC SCHEME

To implement the optimization/regulation algorithms explained above, some *a-priori* modelling and on-line selection of tuning parameters are required. Let us briefly discuss each algorithm's requirements.

Optimizer

In Part I of this series, several tuning parameters were suggested to influence the speed of search of the adaptive algorithm. It was found that because of measurement noise the parameter estimation and thus the speed of search needs to be slowed down to avoid divergence. In addition to the recursive instrumental variables algorithm forgetting factor (λ) other parameters are needed for the constrained version.

Search Stepsize (μ_l). At an unconstrained point $(m_0)_l$ problem 9 does *not* produce the direction q_l of steepest descent. It is recommended to use the steepest descent algorithm as proposed in Part I until a constraint is encountered. Then the feasible directions scheme is used whenever the operating points are constrained. When the search is not constrained a constant $\mu_l = \mu$ was found to be sufficient for convergence of the OC. For the constrained case we suggest the use of stepsize (Eq. 10) where the upper bound $\bar{\mu}$ can be the same stepsize used for steepest descent.

Tolerances (σ_l, ρ_l). These parameters essentially determine how close to the constraints we want to bring the system. If some constraints need to be identified and are strongly nonlinear or when the lower layer regulator does not perform satisfactorily, larger tolerances must be selected to avoid violations.

Regulator

To design an IMC regulator, impulse response coefficients (H_j in Eq. 13) relating r and m are required. In practice, such parameters are obtained by performing step-response experiments on the system. Once the model is at hand, the present input $m(\bar{k})$ is computed by solving Eq. 13 for a specific selection of P, M, Γ_l, B_l. Rules for tuning these parameters to obtain a stable control law are given by Garcia (1982).

Solution of Eq. 13 at each sampling time yields a control law:

$$m(k) = C\epsilon_2(k) - \sum_{i=1}^{N-1} D_i m(k-i) \qquad (15)$$

Unless a constraint handling method is used, Eq. 15 remains fixed throughout the operation. However, since the optimizer drives the process over all the search space, the model output $\tilde{r}$ will differ from r, and instability may occur. Robustness is ensured in IMC by introducing the filter (Eq. 12) with $\alpha_i > 0$. These filter time constants are the only tuning parameters that need to be changed on line for the implementation of the IMC regulator.

SIMULATION EXAMPLE

We have applied the new algorithm for closed-loop OC to a simulated benzene process. This plant has been identified before as a likely candidate for on-line OC due to its tradeoffs and cost intensity (Latour, 1979; Latour et al., 1981). Although simple, the plant has enough constraints and interactions to present a challenging optimization problem.

Benzene Plant: Flowsheet

As the test system for the proposed OC method a plant for the production of high-purity benzene has been selected (Figure 4). Thermal hydro-dealkylation of alkyl-benzenes (toluene and *m*-

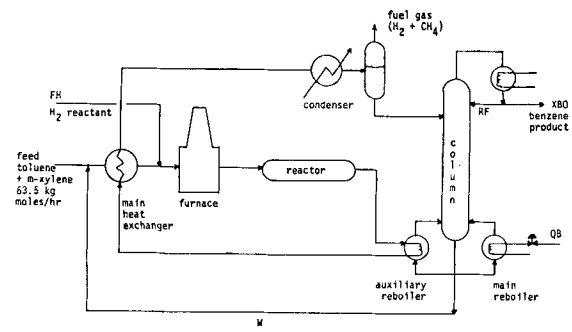

Figure 4. Benzene production plant used in simulation example.

TABLE 1. OPERATING COST FUNCTION OF BENZENE EXAMPLE PLANT

$COST = QFUR \times PF + QFLS \times PC - FMOUT \times PMF - FHOUT \times PHF + FH \times PH + QB \times PS$ (\$/HR)

Process Variables	Units	Price Index	Assumed Values	Units	Description
QFUR	MMKCAL/HR	*PF*	15.0	\$/MMKCAL	Furnace Fuel
QFLS	"	*PC*	76.0	"	Condenser Refrigerant
FMOUT	KGMOLES/HR	*PMF*	0.4	\$/KGMOLE	Methane Product
FHOUT	"	*PHF*	0.4	"	Unreacted Hydrogen
FH	"	*PH*	1.4	"	Hydrogen Reactant
QB	MMKCAL/HR	*PS*	10.0	\$/MMKCAL	Reboiler Steam

xylene are chosen here) occurs at a temperature of 950 K under 39 atm (3.95 MPa). The process consists of a furnace which brings the preheated feed mixture of toluene (T) and m-xylene (X) with pure H_2 to the reaction temperature. The exothermic reaction then takes place in a 70 m^3 tubular reactor increasing the stream temperature even further. Hot product gases are first used to boil part of the bottoms of the downstream benzene column (~510 K) and then to preheat the feed. The quenched stream is then further cooled to flash the unreacted H_2 together with product $CH_4(M)$, which is sold as fuel gas. Product benzene (B) is then separated at about 99.5% purity in a 40 tray column operated at 2 atm (200 kPa).

Column bottoms consist mainly of heavy aromatics which are produced in an equilibrium reaction with benzene. These have no commercial value and consequently are recycled to maintain an equilibrium level.

Specifically, the reactions occurring are

$$T + H_2 \rightarrow B + M$$

$$X + H_2 \rightarrow T + M$$

$$B + B \rightleftarrows D + H_2$$

where product diphenyl (D) is used to represent the heavy aromatics formed.

At high reaction temperatures there is an almost complete conversion of T and X to B. However, large concentrations of B at high temperatures increase the diphenyl equilibrium level. We will consider the reaction temperature to be fixed at 950 K in our simulations. Further details on the process and kinetics are given by Button and Kirby (1971) and Silsby and Sawyer (1956), respectively. For details on the dynamic and steady-state equations used in the simulation the reader is referred to Garcia (1982).

Optimization Problem

During the operation, long-term changes in feedstock quality (T/X ratio) change the optimal operating conditions of the benzene process. The on-line optimizer has the task of finding the economic optimum without having information on these disturbances.

Considering the reaction temperature and column pressure to be fixed, there remain three variables for manipulation: the hydrogen flowrate (FH), the column boilup (QB) and the column reflux ratio (RF).

The operating cost function of the process includes the terms shown in Table 1. Note that FH and QB have both direct and indirect effects on the cost. Due to the recycle, interactions are such that the optimization problem is far from trivial. FH affects the product distribution of the reactor and consequently the recycle composition, the flash condenser duty ($QFLS$) and the furnace duty ($QFUR$). QB and RF influence not only the separation but also the recycle flow and in turn $QFUR$ and $QFLS$. For example, for fixed RF and FH, an increase in QB decreases the recycle heavy aromatics composition causing $QFUR$ to decrease. Therefore, there is a tradeoff between reboiler stream consumption and furnace fuel costs.

Since the main goal of the operation is to maintain the benzene product (column overheads) within specifications, the hard constraint

$$XBO(QB,FH,RF) = XBOS \tag{16}$$

is imposed. Also, due to high costs of reactant H_2 assumed in Table 1, an economic analysis indicates that the optimal operating point generally lies at low levels of FH. This is likely to produce undesirably high levels of diphenyl in the system. Since diphenyl is not removed from the plant, the recycle flowrate W (and column liquid flow) increases. Therefore, an upper operating bound $WMAX$ on W must be employed: QB, FH and RF should not varied such that

$$W(QB,FH,RF) - WMAX \leq 0 \tag{17}$$

Other operating constraints can be introduced. An increase in QB or RF due to changes in feedstock may produce large internal column flows. This may bring the operation close to flooding conditions. By monitoring the pressure drop across the stripping section ($DELP$, psi/tray) the operation can be changed so that

$$DP(QB,FH,RF) = DELP(QB,FH,RF) - DELPMX \leq 0 \tag{18}$$

where $DELPMX$ is a maximum pressure drop above which flooding occurs. In addition, design dimensions put an upper limit on QB:

$$QB - QBMX \leq 0. \tag{19}$$

In summary, the benzene plant steady-state optimization problem becomes:

$$\min_{QB,FH,RF} \; COST\,(QB,FH,RF)$$

$$\text{s.t.} \quad XBO(QB,FH,RF) - XBOS = 0 \tag{20}$$

$$W(QB,FH,RF) - WMAX \leq 0$$

$$DELP(QB,FH,RF) - DELPMX \leq 0$$

$$QB - QBMX \leq 0$$

OC Structure

Given the hard constraint (Eq. 16) XBO is necessarily selected as controlled variable. In addition, since the recycle flow rate constraint (Eq. 17) is expected to be active for most levels of disturbance, W could be controlled at a setpoint WS which is kept less than $WMAX$. Therefore, we must choose from among QB,FH and RF two inputs to be used for regulation.

As explained above, W reaches its upper bound mainly due to low FH which has a strong effect on reactor product distribution.

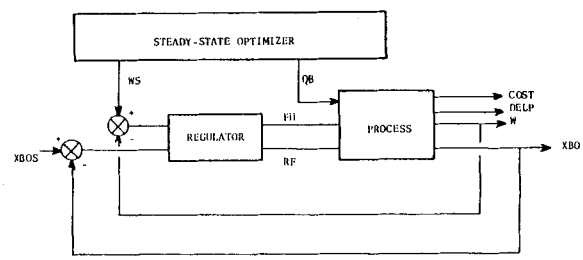

Figure 5. Two-layer OC structure for example benzene plant.

TABLE 2. SUMMARY OF MEASURED AND MANIPULATED VARIABLES

	Variable	Description	Units
Regulator Manipulated Variables	*FH*	Reactant H_2 Flow Rate	KGMOLES/HR
	RF	Column Reflux Ratio	—
Regulator Controlled Variables	*XBO*	Benzene Mole Fraction at Column Overhead	—
	W	Recycle Flow Rate	KGMOLES/HR
Free Manipulated Variables	*QB*	Column Boilup	MMKCAL/HR
Uncontrolled Constraint	*DELP*	Column Stripping Section Pressure Drop	psi/Tray [1 psi = 6.893 kPa]

Thus *FH* should be selected as a regulator manipulated variable. Also, both *QB* and *RF* have a direct influence on *XBO*. However, there is less deadtime in the response to *RF* and therefore it is preferred over *QB*.

The resulting OC structure is shown in Figure 5. Setpoints *XBOS* and *WS* have substituted inputs *RF* and *FH*. However, since product specifications must be met, *XBOS* is fixed, reducing the degrees of freedom for optimization to two variables; the recycle flowrate setpoint *WS* and the column boilup *QB*. (For a summary of the variables used see Table 2.)

QB could have been used to additionally regulare *DELP*. In this example, the flooding constraint is not regulated to demonstrate the constraint handling features of the *OC* via system identification. The problem to be solved by the optimizer is

$$\min_{QB,WS} \text{COST}\,(QB,WS)$$

$$\text{s.t.} \quad WS\text{-}WMAX \le 0 \tag{21}$$

$$DELP(QB,WS) - DELPMX \le 0$$

$$QB - QBMX \le 0$$

Note that both the cost function and the flooding constraint need to be identified to apply the adaptive optimizing control scheme.

Implementation of Closed-Loop OC

Multivariable IMC Design. A sampling time of 5 minutes was selected for regulation. By performing impulse response experiments on the simulated process, the matrices H_i in Eq. 13 were obtained. Then tuning parameters P, M, Γ_1, B_1 were selected employing an interactive program developed by the authors to obtain a stable control law (Garcia, 1982). No retuning of the controller is done on line for constraint handling.

QB is considered as a known disturbance to the *IMC* loop. Since changes in *QB* are frequent, a feedforward compensator is used to reduce its effects on *XBO* and *W*. For details, the reader is referred to Garcia (1982).

Optimizer with System Identification. The gradients of the cost function (COST(*QB*,*WS*)) and of the flooding constraint (*DP*(*QB*,*WS*)) with respect to *QB* and *WS* are needed in the proposed OC search algorithm. These are obtained via the identification of a dynamic model using an instrumental variables estimator (Part I).

Certain contributions to the cost equation have slow dynamics as compared to others. Consequently, a second order model for identification is justified where COST 1 collects the slow dynamic contributions and COST 2 the faster ones.

The following models are identified.

$$\begin{bmatrix} \frac{\text{COST 1}}{100}(k+1) \\ \frac{\text{COST 2}}{100}(k+1) \end{bmatrix} = A \begin{bmatrix} \frac{\text{COST 1}}{100}(k) \\ \frac{\text{COST 2}}{100}(k) \end{bmatrix} + B \begin{bmatrix} QB(k) \\ \frac{WS}{50}(k) \end{bmatrix} + C \tag{22}$$

$$DP(k+1) = a_3 DP(k) + b_{31} QB(k) + b_{32} \frac{WS}{50}(k) + c_3$$

where

$$\begin{aligned} \text{COST 1} &= QFUR \times PF + QFLS \times PC - FMOUT \times PMF \\ \text{COST 2} &= FH \times PH - FHOUT \times PHF \\ \text{COST} &= \text{COST 1} + \text{COST 2} + QB \times PS \\ DP &= DELP - DELPMX. \end{aligned}$$

Note that COST 1, COST 2 and *WS* were scaled for improved numerical behavior of the recursive identification algorithm. The search is performed in the scaled variables *QB*, *MS*/50. From Eq. 22 the optimizer uses the steady-state gradients:

$$\nabla \frac{\text{COST}}{100} = (1\ \ 1)(I - A)^{-1}B + \begin{pmatrix} PS \\ \overline{100} \end{pmatrix} 0 \tag{23}$$

$$\nabla DP = \begin{bmatrix} \frac{b_{31}}{1-a_3} & \frac{b_{32}}{1-a_3} \end{bmatrix}$$

The sampling time for identification is 2.5 minutes, which is half of that used for regulation. This value was selected to increase the number of data points between optimization moves. For all the runs 14 samples are taken before updating the inputs every 35 minutes.

If no preliminary estimates of the system are available it is suggested to use *PRBS* signals around an operating point, as described in Part I. This was done here prior to activating the optimizer.

Simulation Runs

As the process feed composition changes the optimum of the benzene plant shifts due primarily to constraint moves. Upper bounds on *W* of *WMAX* = 30. KGMOLES/HR and on *QB* of *QBMX* = 1.275 MMKCAL/HR are imposed on the system. Then for a 90% toluene mole fraction in the plant feed and *XBO* = 0.995 the optimum is at both maximum values (*QB*,*WS*) = (1.275,30.) (Figure 6). When the toluene feed mole fraction decreases to 40% a larger *RF* is needed therefore increasing *DELP*. In our plant *DELPMX* = 1.85 psi tray (12.8 kPa/tray) at which 95% of flooding occurs. This 95% flooding constraint then shifts as shown, rendering the operating point 1 infeasible. The optimum corresponding to

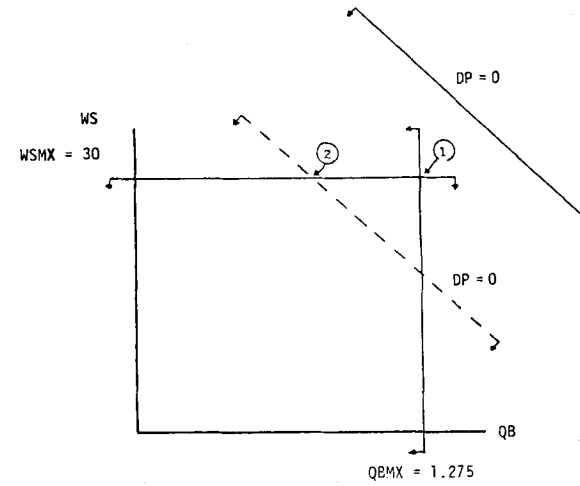

Figure 6. Shifting of optimal operating points for benzene plant optimum at 90% toluene in feed, ①: flooding constraint (—), ②: 40% toluene in feed, flooding (- - -).

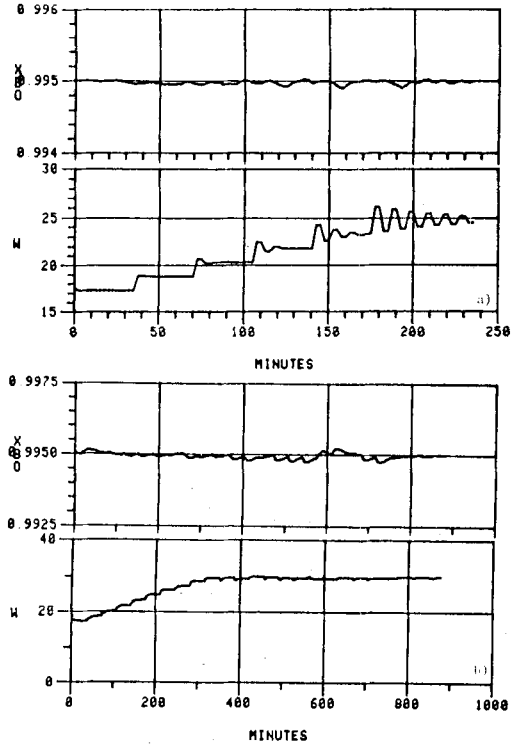

Figure 7. Effect of modelling errors on regulator performance. IMC outputs during search for optimum ① of Figure 7 with $\lambda = 0.90$, $\mu = 0.03$, $\sigma_l = \rho_l = 0.01$. a) No robustness filter ($\alpha_f = 0$); b) $\alpha_f = 0.65$.

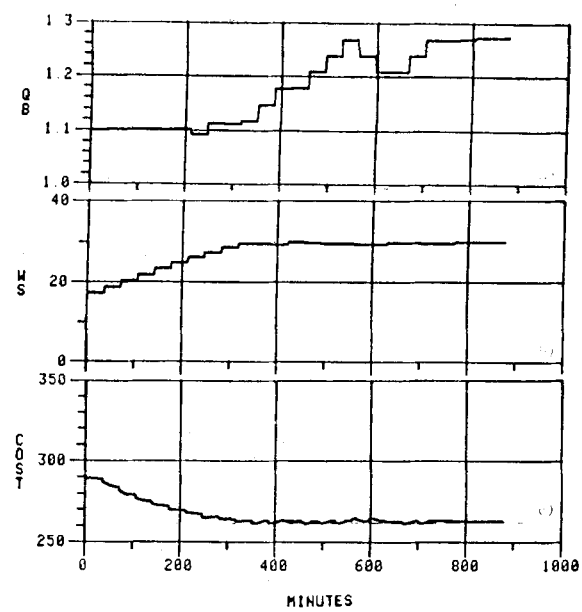

Figure 8. Transients during search for optimum ① of Figure 6 ($\lambda = 0.9$, $\mu = 0.03$, $\sigma_l\rho_l = 0.01$, $\alpha_f = 0.65$). a) column boilup, b) recycle flow rate, c) operating cost (- - - steady-state optimum).

the new disturbance level is now at $(QB,WS) = 1.250,30.$) which is where the flooding constraint intersects $W = WSMX = 30$.

We should note here that these changes in the optimal operating point are nontrivial to predict, particularly the position of the flooding constraint. In addition, system safety is always at stake since the operating point essentially "travels" on the flooding limit at the maximum recycle flow. For our case, the constraint DP is not controlled and thus we are testing the ability of the OC to maintain the system within allowable bounds based on its identified model.

Despite so many constant changes the regulator must maintain XBO at its specification level $XBOS$. But more importantly it must prevent W from exceeding the upper limit $WMAX$ by keeping it close to its set point. This must be done despite changes in the system due to a shift from the nominal steady-state where the dynamic model used for design is valid.

On-Line Search: 90% Toluene Inlet Composition. Initially, the plant is operating at $(QB,WS) = (1.10,17.5)$ for a toluene feed composition of 90%, $XBOS = 0.995$. Using tuning parameter values of $\lambda = 0.90$, $\mu = 0.03$, $\sigma_l = \rho_l = 0.01$ the search was attempted using no filtering action in IMC ($\alpha_f = 0$). The resulting responses of the regular outputs are shown in Figure 7a. The model parameters used for regulator design were taken at $(RF,FH,QB) = (1.4,125.0,1.13)$ which is close to the starting operating point. However, as the process is driven away by the OC this model does not apply any more and in the absence of a filter instability occurs.

The search was repeated using filter time constants of $\alpha_f = 0.65$. The resulting response is shown in Figure 7b which remains stable throughout the search, despite plant changes.

The sequence of changes in the optimization variables is given in Figure 8. Note that the steepest descent direction indicates essentially an increase in WS, until its upper limit is reached. Then the feasible direction search scheme moves the operating point basically along $WS = 30.0$ increasing QB until the upper bound $QBMX$ is encountered.

We note that the process is essentially at the optimum after 300 min which is approximately ten system time constants. Any experimental search which employs on-line static gradient measurements would have performed about four steady-state experiments during this time allowing at most two input moves. Due to the on-line identification of the dynamic model (Eq. 22), there is no need to wait for steady state and thus the speed of search is dramatically improved. Also, by using the antidivergence provision given in Part I, the OC is able to maintain the operating point at a constant value for this noise-free system without failure of the recursive identification algorithm.

We should note the simplicity of introducing an exponential filter for robustness. Should undesirable closed-loop behavior start to show during the operation as in Figure 7, the operators can easily retune the filter to restore stability. This is done without changing the control law and the resulting response has the physical significance of an exponential trajectory.

On-Line Search: 40% Toluene Inlet Composition. After convergence of the OC at the optimum for 90% toluene feed, a change in toluene feed composition down to 40% was introduced. In Figure 9, the sequence of changes in the optimization variables produced by the OC is shown. Note that it brings the boilup QB to a lower value to avoid flooding, finally converging at the optimum. Due to the selected tolerances σ_l, ρ_l, the OC converges to values of QB,WS lower than the true optimal levels.

Figure 10 shows the transients during the search for the new optimum. Due to the disturbance, the operating cost shows a net increase. After about 140 minutes the column is at 95% flooding

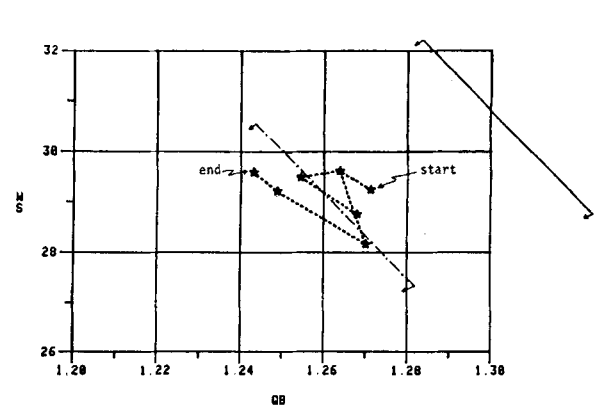

Figure 9. Search for optimum when toluene feed composition is reduced to 40%, after convergence to optimum ① of Figure 6. (—) flooding constraint at 90% toluene feed, (- · · · -) flooding constraint at 40% toluene feed.

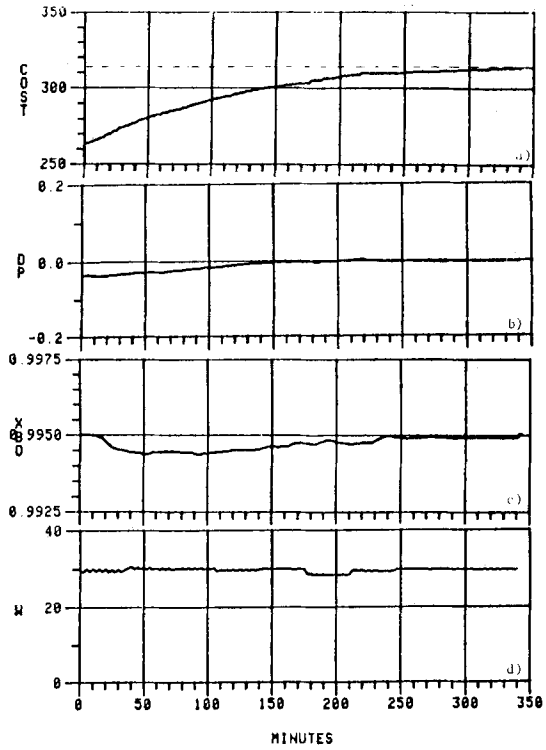

Figure 10. Transients during the search of Figure 9: a) operating cost (--- steady-state optimum); b) flooding constraint; c) column overhead benzene composition; d) recycle flow rate.

and the OC is able to maintain the operating safely at the flooding imit during the rest of the search. This demonstrates the reliability of the on-line identifier in determining constraint position. Also we observe the excellent performance of the IMC regulator as it keeps W at its maximum value while the flooding constraint shifts. The overhead composition shows some deviation from 0.995 during the transients. This could be avoided if a constraint handling procedure at the regulator level (e.g., by using the QP (Eq. 14)] were employed.

ACKNOWLEDGMENT

Partial support from the National Science Foundation (CPE-8115022) is gratefully acknowledged.

NOTATION

P = economic objective function
$f(\cdot)$ = equality ("hard") constraints
$g(\cdot)$ = inequality constraints
y = vector of system outputs
m = vector of manipulatable system inputs
d = disturbances to the plant
r = function of outputs and inputs which is controlled
m_o = independent optimization variable
q = optimization parameter defined in Eq. 9
G = system transfer matrix
G_c = predictive controller transfer matrix
$\tilde{G}$ = system model transfer matrix
F = robustness filter transfer matrix
P = controller horizon
N = model truncation order
M = number of input moves allowed by controller in future

Greek Letters

α_i = parameter in exponential filter
B_e = weighting matrix
Γ_e = weighting matrix
μ = optimization step size
ζ = tolerance factor defined in Eq. 11
σ = tolerance factor defined in Eq. 9

Subscripts

f = related to equality constraint f
g = related to inequality constraint g
s = specified, i.e., set point
l = search index in optimization

LITERATURE CITED

Arkun, Y., "Design of Steady-State Optimizing Control Structures for Chemical Processes," PhD Thesis, University of Minnesota (1979).

Avriel, M., *Nonlinear Programming Analysis and Methods*, Prentice-Hall, NJ (1976).

Bazaraa, M. S., and C. M. Shetty, *Nonlinear Programming Theory and Algorithms*, Wiley, New York (1979).

Bhattacharya, A., and B. Joseph, "On-line Optimization of Chemical Processes," JACC Proceedings (1982).

Brosilow, C., Proceedings Sponsors' Meeting, Control of Industrial Systems Research Program, Case Western Reserve University, Cleveland, OH (Dec., 1981).

Button, D. L., and L. J. Kirby, "Thermal Hydrodealkylation Process," U.S. Patent No. 3,607,960 (1971).

Chang, T. S., and D. E. Seborg, "Process Control in the Presence of Constraints," (1982).

Cutler, C. R., and B. L. Ramaker, "Dynamic Matrix Control—A Computer Control Algorithm," Paper 51b, AIChE 86th National Meeting (April, 1979).

Duncanson, L. A., and P. V. Youle, "On-line Control of Olefin Plant," *Chem. and Proc. Eng.*, **51**, 49 (May, 1970).

Garcia, C. E., "Studies in Optimizing and Regulatory Control of Chemical Processing Systems," PhD Thesis, University of Wisconsin-Madison (1982).

Garcia, C. E., and M. Morari, "Optimal Operation of Integrated Processing Systems. Part I: Open-loop On-line Optimizing Control," *AIChE J.*, **27**, 960 (1981).

Garcia, C. E., and M. Morari, "Internal Model Control. 1. A Unifying Review and Some New Results," *IEC Proc. Des. and Dev.*, **21**, 308 (1982).

Garcia, C. E., and M. Morari, "Internal Model Control. 2: The Multivariable Case," *IEC Proc. Des. and Dev.* (1983).

Latour, P. R., "On-line Computer Optimization. 2: Benefits and Implementation," HP July, 219 (1979).

Latour, P. R., G. D. Martin, and L. A. Richard, "Closed-loop Optimization of Distillation Energy," Paper 45c, AIChE 88th National Meeting (April, 1981).

Mangasarian, O. L., *Nonlinear Programming*, R. Krieger Publ., Huntinton, NY (1969).

Mehra, R. K., R. Rouhani, J. Eterno, J. Richalet, and A. Rault, "Model Algorithmic Control (MAC); Review and Recent Developments," *Chemical Process Control 2*, T. F. Edgar and D. F. Seborg, Eds., Engineering Foundation (1982).

Prett, D. M., and R. D. Gillette, "Optimization and Constrained Multivariable Control of a Catalytic Cracking Unit," Paper 51c, AIChE 86th National Meeting (April, 1979).

Sargent, R. W. H., "A Review of Optimization Methods for Nonlinear Problems," *ACS Symp. Ser.*, **124**, 37 (1980).

Shah, M. J., and R. F. Stillman, "Computer Control and Optimization of a Large Methanol Plant," *IEC*, **62**, 59 (1970).

Silsby, R. I., and E. W. Sawyer, "The Dealkylation of Alkyl Aromatic Hydrocarbons Part I," *J. Appl. Chem.*, **6**, 347 (1956).

Zoutendijk, G., *Methods of Feasible Directions; A Study in Linear and Nonlinear Programming*, Elsevier, New York (1960).

Manuscript received July 30, 1982; revision received April 1, and accepted April 5, 1983.