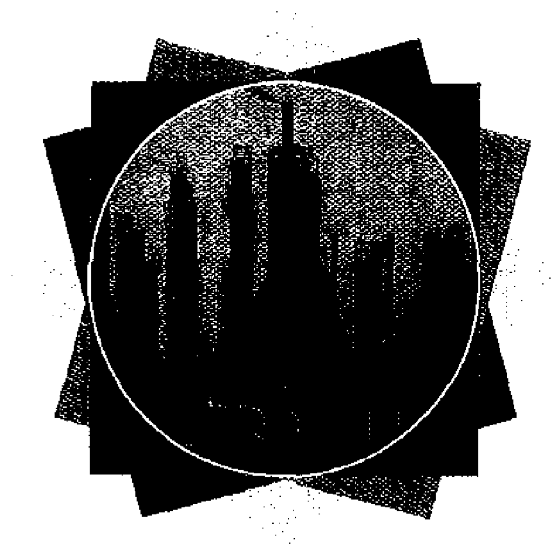

SPECIAL SECTION
Industrial Process Control

Tutorial Overview of Model Predictive Control

By James B. Rawlings

Providing a reasonably accessible and self-contained tutorial exposition on model predictive control (MPC) is the purpose of this article. It is aimed at readers with control expertise, particularly practitioners, who wish to broaden their perspective in the MPC area of control technology. We introduce the concepts, provide a framework in which the critical issues can be expressed and analyzed, and point out how MPC allows practitioners to address the trade-offs that must be considered in implementing a control technology.

The MPC research literature is by now large, but review articles have appeared at regular intervals. We should point these out before narrowing the focus in the interest of presenting a reasonably self-contained tutorial for the nonexpert. The three MPC papers presented at the Chemical Process Control (CPC) V conference in 1996 are an excellent starting point [2]-[4]. Qin and Badgwell present comparisons of industrial MPC algorithms that practitioners may find particularly useful. Chen and Allgöwer and Morari and Lee provide other recent reviews [5], [6]. Kwon provides a very extensive list of references [7]. Moreover, several excellent books have appeared recently [8]-[10]. For those interested in the status of MPC for *nonlinear* plants, [11] would be of strong interest. Finally, Allgöwer and coworkers have presented a recent minicourse covering the area [12].

Models

The essence of MPC is to optimize, over the manipulable inputs, forecasts of process behavior. The forecasting is accomplished with a process model, and, therefore, the model is the essential element of an MPC controller. As discussed subsequently, models are not perfect forecasters, and feedback can overcome some effects of poor models, but start-

The author is with the Department of Chemical Engineering, University of Wisconsin-Madison, Madison, WI 53706-1691, U.S.A. (jbraw@che.wisc.edu). A modified version of this paper was presented at the 1999 American Control Conference [1].

1053-5888/00/$10.00©2000IEEE

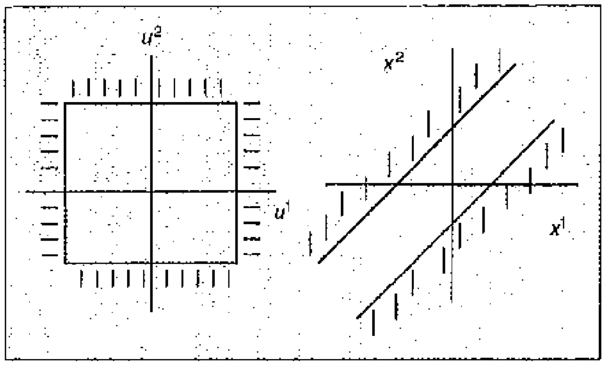

Figure 1. *Example input and state constraint regions defined by (3)-(4).*

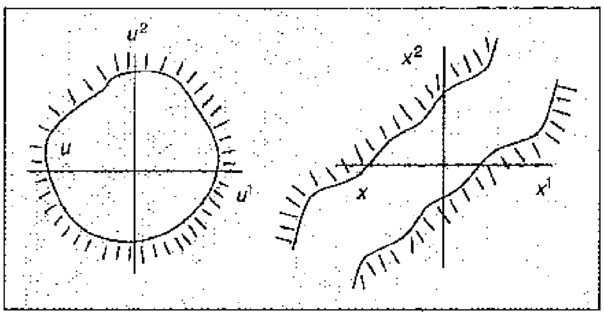

Figure 2. *Example input and state constraint regions defined by (7)-(8).*

ing with a poor process model is akin to driving a car at night without headlights; the feedback may be a bit late to be truly effective.

Linear Models

Historically, the models of choice in early industrial MPC applications were time domain, input/output, step, or impulse response models [13]-[15]. Part of the early appeal of MPC for practitioners in the process industries was undoubtedly the ease of understanding provided by this model form. It has become more common for MPC researchers, however, to discuss linear models in state-space form:

$$\frac{dx}{dt} = Ax + Bu \qquad x_{j+1} = Ax_j + Bu_j$$
$$y = Cx \qquad y_j = Cx_j$$

in which x is the n-vector of states, y is the p-vector of (measurable) outputs, u is the m-vector of (manipulable) inputs, and t is the continuous-time and j is the discrete-time sample number. Continuous-time models may be more familiar to those with a classical control background in transfer functions, but discrete-time models are very convenient for digital computer implementation. With abuse of notation, we use the same system matrices (A,B,C) for either model, but the subsequent discussion focuses on discrete time. Transformation from continuous-time to discrete-time models is available as a one-line command in a language like Octave or MATLAB. Linear models in the process industries are, by their nature, empirical models and identified from input/output data. The ideal model form for identification purposes is perhaps best left to the experts in identification theory, but a survey of that literature indicates no disadvantage to using state-space models inside the MPC controller.

The discussion of MPC in state-space form has several advantages, including easy generalization to multivariable systems, ease of analysis of closed-loop properties, and online computation. Furthermore, starting with this model form, the wealth of linear systems theory—the linear quadratic (LQ) regulator theory, Kalman filtering theory, internal model principle, etc.—is immediately accessible for use in MPC. We demonstrate the usefulness of these tools subsequently.

A word of caution is also in order. Categories, frameworks, and viewpoints, while indispensable for clear thinking and communication, may blind us to other possibilities. We should resist the easy temptation to formulate all control issues from an LQ, state-space framework. The tendency is to focus on those issues that are easily imported into the dominant framework while neglecting other issues, of possibly equal or greater import to practice, which are difficult to analyze, awkward, and inconvenient.

From a theoretical perspective, the significant shift in problem formulation came from the MPC practitioners who insisted on maintaining constraints, particularly input constraints in the problem formulation

$$\frac{dx}{dt} = Ax + B \qquad x_{j+1} = Ax_j + Bu_j \tag{1}$$

$$y = Cx \qquad y_j = Cx_j \tag{2}$$

$$Du \le d \qquad Du_j \le d \tag{3}$$

$$Hx \le h \qquad Hx_j \le h \tag{4}$$

in which D, H are the constraint matrices and d, h are positive vectors. The constraint region boundaries are straight lines, as shown in Fig. 1. At this point we are assuming that $x = 0, u = 0$ is the steady state to which we are controlling the process, but we treat the more general case subsequently.

Optimization over inputs subject to hard constraints leads immediately to nonlinear control, and that departure from the well-understood and well-tested linear control theory provided practitioners with an important, new control technology and motivated researchers to better understand this new framework. Certainly optimal control with constraints was not a new concept in the 1970s, but the moving horizon implementation of these open-loop optimal control solutions subject to constraints at each sample time was the new twist that had not been fully investigated.

Nonlinear Models

The use of nonlinear models in MPC is motivated by the possibility of improving control by improving the quality of the forecasting. The fundamentals in any process control problem—conservation of mass, momentum, and energy; considerations of phase equilibria; relationships of chemical kinetics and properties of final products—all introduce nonlinearity into the process description. Determining the settings in which the use of nonlinear models for forecasting delivers improved control performance is an open issue. For continuous processes maintained at nominal operating conditions and subject to small disturbances, the potential improvement would appear small. For processes operated over large regions of the state space—semibatch reactors, frequent product grade changes, processes subject to large disturbances, for example—the advantages of nonlinear models appear larger.

The essence of MPC is to optimize forecasts of process behavior. The forecasting is accomplished with a process model, and, therefore, the model is the essential element of an MPC controller.

Identification of nonlinear models runs the entire range from models based on fundamental principles with only parameter estimation from data to completely empirical nonlinear models with all coefficients identified from data. We will not stray into the issues of identification of nonlinear models, which is a large topic by itself. The interested reader may consult [16] and [17] and the references therein for an entry point into this literature. Qin and Badgwell's recent survey of vendor MPC products includes those based on several forms of polynomial nonlinear auto-regressive moving-average exogenous (NARMAX) models and nonlinear neural net models [18]. Bequette provides a summary review of the models used in nonlinear MPC [19].

Regardless of the model form and identification method, for tutorial purposes we represent the nonlinear model inside the MPC controller also in state-space form:

$$\frac{dx}{dt} = f(x,u) \qquad x_{j+1} = f\left(x_j, u_j\right) \tag{5}$$

$$y = g(x) \qquad y_j = g(x_j) \tag{6}$$

$$u \in \mathcal{U} \qquad u_j \in \mathcal{U} \tag{7}$$

$$x \in \mathcal{X} \qquad x_j \in \mathcal{X}. \tag{8}$$

If the model is nonlinear, there is no advantage in keeping the constraints as linear inequalities, so we consider the constraints as membership in more general regions $\mathcal{U}, \mathcal{X}$ shown in Fig. 2.

MPC with Linear Models

We focus on formulating MPC as an infinite horizon optimal control strategy with a quadratic performance criterion. We use the following discrete-time model of the plant:

$$x_{j+1} = Ax_j + B(u_j + d) \tag{9}$$

$$y_j = Cx_j + p. \tag{10}$$

The affine terms d and p serve the purpose of adding integral control. They may be interpreted as modeling the effect of constant disturbances influencing the input and output, respectively. Assuming that the state of the plant is perfectly measured, we define MPC as the feedback law $u_j = \mu(x_j)$ that minimizes

$$\Phi = \frac{1}{2}\sum_{j=0}^{\infty}\left(y_j - \bar{y}\right)' Q\left(y_j - \bar{y}\right) + (u_j - \bar{u})' R\left(u_j - \bar{u}\right) + \Delta u_j' S \Delta u_j \tag{11}$$

in which $\Delta u_j \triangleq u_j - u_{j-1}$. The matrices Q, R, and S are assumed to be symmetric positive definite. When the complete state of the plant is not measured, as is almost always the case, the addition of a state estimator is necessary (see the "State Estimation" section).

The vector $\bar{y}$ is the desired output target and $\bar{u}$ is the desired input target, assumed for simplicity to be time invariant. In many industrial implementations, the desired targets are calculated as a steady-state economic optimization at the plant level. In these cases, the desired targets are normally constant between plant optimizations, which are performed on a slower time scale than the one at which the MPC controller operates. In batch and semi-batch reactor operation, on the other hand, a final time objective may be optimized instead, which produces a time-varying trajectory for the system states. Even in continuous operations, some recommend tuning MPC controllers by specifying the setpoint trajectory, often a first order response with adjustable time constant. As discussed by Bitmead et al. [10] in the context of generalized predictive control (GPC), one can pose these types of tracking problems within the LQ framework by augmenting the state of the system to describe the evolution of the reference signal and posing an LQ problem for the combined system.

For a time invariant setpoint, the steady-state aspect of the control problem is to determine appropriate values of (y_s, x_s, u_s). Ideally, $y_s = \bar{y}$ and $u_s = \bar{u}$. Process limitations and constraints, however, may prevent the system from reaching the desired steady state. The goal of the target calculation is to find the feasible triple (y_s, x_s, u_s) such that y_s and u_s are as close as possible to $\bar{y}$ and $\bar{u}$. We address the target calculation below.

The fundamentals in any process control problem—conservation of mass, momentum, and energy; considerations of phase equilibria; relationships of chemical kinetics and properties of final products—all introduce nonlinearity into the process description.

To simplify the analysis and formulation, we transform (11) using deviation variables to the generic infinite horizon quadratic criterion

$$\Phi = \frac{1}{2}\sum_{j=0}^{\infty} z_j' Q z_j + v_j' R v_j + \Delta v_j' S \Delta v_j. \tag{12}$$

The original criterion (11) can be recovered from (12) by making the following substitutions:

$$z_j \leftarrow y_j - Cx_s - p, \quad w_j \leftarrow x_j - x_s, \quad v_j \leftarrow u_j - u_s$$

in which y_s, x_s, and u_s are the steady states satisfying the following relation:

$$\begin{aligned} x_s &= Ax_s + B(u_s + d) \\ y_s &= Cx_s + p. \end{aligned}$$

By using deviation variables, we treat separately the steady-state and the dynamic elements of the control problem, thereby simplifying the overall analysis of the controller.

The dynamic aspect of the control problem is to control (y, x, u) to the steady-state values (y_s, x_s, u_s) in the face of constraints, which are assumed not to be active at steady state, i.e., the origin is in the strict interior of regions $\mathcal{X}, \mathcal{U}$. See [20] for a preliminary treatment of the case in which the constraints are active at the steady-state operating point. This part of the problem is discussed in the "Receding Horizon Regulator" section. In particular, we determine the state feedback law $v_j = \rho(w_j)$ that minimizes (12). When there are no inequality constraints, the feedback law is the linear quadratic regulator. With the addition of inequality constraints, however, an analytic form for $\rho(w_j)$ may not exist. For cases in which an analytic solution is unavailable, the feedback law is obtained by repeatedly solving the open-loop optimal control problem. This strategy allows us to consider only the encountered sequence of measured states rather than the entire state space. For a further discussion, see Mayne [21].

If we consider only linear constraints on the input, input velocity, and outputs of the form

$$\begin{aligned} u_{\min} &\le Du_k \le u_{\max}, \\ -\Delta_u &\le \Delta u_k \le \Delta_u, \\ y_{\min} &\le Cx_k \le y_{\max} \end{aligned} \tag{13}$$

we formulate the regulator as the solution of the following infinite horizon optimal control problem:

$$\min_{\{w_k, v_k\}} \Phi(x_j) = \frac{1}{2}\sum_{k=0}^{\infty} z_k' Q z_k + v_k' R v_k + \Delta v_k' S \Delta v_k$$

subject to the constraints

$$\begin{aligned} & w_0 = x_j - x_s, \;\; v_{j-1} = u_{j-1} - u_s \\ & w_{k+1} = Aw_k + Bv_k, \;\; z_k = Cw_k \\ & u_{\min} - u_s \le Dv_k \le u_{\max} - u_s \\ & -\Delta_u \le \Delta v_k \le \Delta_u \\ & y_{\min} - y_s \le Cw_k \le y_{\max} - y_s. \end{aligned}$$

If we denote

$$\left\{ w_{k+1}^*(x_j), v_k^*(x_j) \right\}_{k=0}^{\infty} = \arg\min \Phi(x_j),$$

then the control law is

$$\rho(x_j) = v_0^*(x_j).$$

We address the regulation problem in the "Receding Horizon Regulator" section.

Combining the solution of the target tracking problem and the constrained regulator, we define the MPC algorithm as follows:

1. Obtain an estimate of the state and disturbances $\Rightarrow (x_j, p, d)$.

2. Determine the steady-state target $\Rightarrow (y_s, x_s, u_s)$.

3. Solve the regulation problem $\Rightarrow v_j$.

4. Let $u_j = v_j + u_s$.

5. Repeat for $j \leftarrow j+1$.

Target Calculation

When the number of the inputs equals the number of outputs, the solution to the unconstrained target problem is obtained using the steady-state gain matrix, assuming such a matrix exists (i.e., the system has no integrators). For systems with unequal numbers of inputs and outputs, integrators, or inequality constraints, however, the target calculation is formulated as a mathematical program [22], [23]. When there are at least as many inputs as outputs, multiple combinations of inputs may yield the desired output target at steady state. For such systems, a mathematical program with a least-squares objective is formulated to determine the best combinations of inputs. When the number of outputs is greater than the number of inputs, situations exist in which no combination of inputs satisfies the output target at steady state. For such cases, we formulate a mathematical program that determines the steady-state output $y_s \neq \bar{y}$ that is closest to $\bar{y}$ in a least-squares sense.

Instead of solving separate problems to establish the target, we prefer to solve one problem for both situations. Through the use of an exact penalty [24], we formulate the target tracking problem as a single quadratic program that achieves the output target, if possible, and relaxes the problem in an l_1/l_2^2 optimal sense if the target is infeasible. We formulate the soft constraint

$$\bar{y} - Cx_s - p \le \eta,$$
$$\bar{y} - Cx_s - p \ge -\eta,$$
$$\eta \ge 0$$

by relaxing the constraint $Cx_s + p = \bar{y}$ using the slack variable η. By suitably penalizing η, we guarantee that the relaxed constraint is binding when it is feasible. We formulate the exact soft constraint by adding an l_1/l_2^2 penalty to the objective function. The l_1/l_2^2 penalty is simply the combination of a linear penalty $q_s'\eta$ and a quadratic penalty $\eta' Q_s \eta$, in which the elements of q_s are strictly nonnegative and Q_s is a symmetric positive definite matrix. By choosing the linear penalty sufficiently large, the soft constraint is guaranteed to be exact. A lower bound on the elements of q_s to ensure that the original hard constraints are satisfied by the solution cannot be calculated explicitly without knowing the solution to the original problem, because the lower bound depends on the optimal Lagrange multipliers for the original problem. In theory, a conservative state-dependent upper bound for these multipliers may be obtained by exploiting the Lipschitz continuity of the quadratic program [25]. In practice, however, we rarely need to guarantee that the l_1/l_2^2 penalty is exact. Rather, we use approximate values for q_s obtained by computational experience. When constructing an exact penalty, the quadratic term is superfluous. The quadratic term adds an extra degree of freedom for tuning, however, and is necessary to guarantee uniqueness.

We now formulate the target tracking optimization as the following quadratic program:

$$\min_{x_s, u_s, \eta} \frac{1}{2}\left(\eta' Q_s \eta + (u_s - \bar{u})' R_s (u_s - \bar{u})\right) + q_s' \eta \tag{14}$$

subject to the constraints

$$\begin{bmatrix} I-A & -B & 0 \\ C & 0 & I \\ C & 0 & -I \end{bmatrix} \begin{bmatrix} x_s \\ u_s \\ \eta \end{bmatrix} \begin{Bmatrix} = \\ \ge \\ \le \end{Bmatrix} \begin{bmatrix} Bd \\ \bar{y} - p \\ \bar{y} - p \end{bmatrix} \tag{15a}$$

$$\eta \ge 0 \tag{15b}$$

$$u_{\min} \le Du_s \le u_{\max}, \; y_{\min} \le Cx_s + p \le y_{\max} \tag{15c}$$

in which R_s and Q_s are assumed to be symmetric positive definite.

Because x_s is not explicitly in the objective function, the question arises as to whether the solution to (14) is unique. If the feasible region is nonempty, the solution exists because the quadratic program is bounded below on the feasible region. If Q_s and R_s are symmetric positive definite, y_s and u_s are uniquely determined by the solution of the quadratic program. Without a quadratic penalty on x_s, however, there is no guarantee that the resulting solution for x_s is unique. Nonuniqueness in the steady-state value of x_s presents potential problems for the controller because the origin of the regulator is not fixed at each sample time. Consider, for example, a tank in which the level is unmeasured (i.e., an unobservable integrator). The steady-state solution is to set $u_s = 0$ (i.e., balance the flows). Any level x_s, within bounds, however, is an optimal alternative. Likewise, at the next time instant, a different level would be a suitably optimal steady-state target. The resulting closed-loop performance for the system could be erratic, because the controller may constantly adjust the level of the tank, never letting the system settle to a steady state.

To avoid such situations, we restrict our discussion to detectable systems and recommend redesign if a system does not meet this assumption. For detectable systems, the solution to the quadratic program is unique, assuming the feasible region is nonempty. The details of the proof are given in [20]. Uniqueness is also guaranteed when only the integrators are observable. For the practitioner, this condition translates into the requirement that all levels are measured. The reason we choose the stronger condition of detectability is that if good control is desired, then the unstable modes of the system should be observable. Detectability is also required to guarantee the stability of the regulator.

Empty feasible regions are a result of the inequality constraints (15c). Without the inequality constraints (15c) the feasible region is nonempty, thereby guaranteeing the existence of a feasible and unique solution under the condition of detectability. For example, the solution $(u_s, x_s, \eta) = (-d, 0, |\bar{y} - p|)$ is feasible. The addition of the inequality constraints (15c), however, presents the possibility of infeasibility. Even with well-defined constraints, $u_{min} < u_{max}$ and $y_{min} < y_{max}$, disturbances may render the feasible region empty. Since the constraints on the input usually result from physical limitations such as valve saturation, relaxing only the output constraints is one possibility to circumvent infeasibilities. Assuming that $u_{min} \le -d \le u_{max}$, the feasible region is always nonempty. We contend, however, that the output constraints should not be relaxed in the target calculation. Rather, an infeasible solution, readily determined during the initial phase in the solution of the quadratic program, should be used as an indicator of a process exception. While relaxing the output constraints in the dynamic regulator is common practice [26]-[30], the output constraint violations are transient. On the other hand, by relaxing output constraints in the target calculation, the controller seeks a steady-state target that continuously violates the output constraints. The steady violation indicates that the controller is unable to compensate adequately for the disturbance and, therefore, should indicate a process exception.

The difficulty that MPC introduces into the robustness question is the open-loop nature of the optimal control problem and the implicit feedback produced by the receding horizon implementation.

Receding Horizon Regulator

Given the calculated steady state, we formulate the regulator as the following infinite horizon optimal control problem:

$$\min_{\{w_k, v_k\}} \Phi(x_j) = \frac{1}{2}\sum_{k=0}^{\infty} w_k' C' Q C w_k + v_k' R v_k + \Delta v_k' S \Delta v_k \tag{16}$$

subject to the constraints:

$$w_0 = x_j - x_s, \quad v_{j-1} = u_{j-1} - u_s \tag{17a}$$

$$w_{k+1} = A w_k + B v_k \tag{17b}$$

$$u_{min} - u_s \le D v_k \le u_{max} - u_s \tag{17c}$$

$$-\Delta_u \le \Delta v_k \le \Delta_u \tag{17d}$$

$$\underline{y} - y_s \le C w_k \le \bar{y} - y_s. \tag{17e}$$

We assume that Q and R are symmetric positive definite matrices. We also assume that the origin $(w_j, v_j) = (0,0)$ is an element of the feasible region $\mathbb{W} \times \mathbb{V}$ (where $\mathbb{W} = \{w \mid y_{min} - y_s \le Cw \le y_{max} - y_s\}$ and $\mathbb{V} = \{v \mid u_{min} \le Dv \le d_{max}, -\Delta_u - u_s \le \Delta v \le \Delta_u - u_s\}$. If the pair (A,B) is constrained stabilizable and the pair $(A, Q^{1/2}C)$ is detectable, then $x_j = 0$ is an exponentially stable fixed point of the closed-loop system. For unstable state transition matrices, the optimization problem may be ill-conditioned because the system dynamics are propagated through the unstable A matrix. To improve the conditioning of the optimization, one can reparametrize the input as $v_k = K w_k + r_k$, in which K is a linear stabilizing feedback gain for (A,B) [31], [32]. The system model becomes

$$w_{k+1} = (A + BK) w_k + B r_k \tag{18}$$

in which r_k is the new input. By initially specifying a stabilizing, potentially infeasible, trajectory, we can improve the numerical conditioning of the optimization by propagating the system dynamics through the stable $(A + BK)$ matrix.

This reparametrization of input is highly recommended if one chooses to solve the state equations explicitly and remove the w_k decision variables in (16). If one instead solves for the state and input simultaneously, the conditioning issue for unstable A largely disappears, because pivoting in the linear algebra subproblems required to solve the optimization provides good conditioning even if A is unstable. For nonlinear problems, simultaneous solution of state and input is also recommended if the plant state trajectory is potentially unstable or exhibits high sensitivity. Techniques for applying the simultaneous approach and producing a well-conditioned discrete-time representation of the continuous-time differential equation models are known as multiple shooting methods in the optimization literature. Biegler and Bock provide excellent further reading on this topic [33], [34].

By expanding Δv_k and substituting in for v_k, we transform (16)-(17) into the following form:

$$\min_{\{w_k, v_k\}} \Phi(x_j) = \frac{1}{2}\sum_{k=0}^{\infty} w_k' Q w_k + v_k' R v_k + 2 w_k' M v_k \tag{19}$$

subject to the following constraints:

$$w_0 = x_j \qquad w_{k+1} = A w_k + B v_k \tag{20a}$$

$$d_{\min} \le D v_k - G w_k \le d_{\max} \tag{20b}$$

$$y_{\min} - y_s \le C w_k \le y_{\max} - y_s. \tag{20c}$$

The original formulation (16)-(17) can be recovered from (19)-(20) by making the following substitutions into the second formulation:

$$\begin{aligned}
&x_j \leftarrow \begin{bmatrix} x_j - x_s \\ u_{j-1} - u_s \end{bmatrix}, \quad w_k \leftarrow \begin{bmatrix} w_k \\ v_{k-1} \end{bmatrix}, \quad v_k \leftarrow r_k \\
&A \leftarrow \begin{bmatrix} A + BK & 0 \\ K & 0 \end{bmatrix}, \quad B \leftarrow \begin{bmatrix} B \\ I \end{bmatrix} \\
&C \leftarrow \begin{bmatrix} C \\ 0 \end{bmatrix}, \quad M \leftarrow \begin{bmatrix} K'(R+S) \\ -S \end{bmatrix} \\
&Q \leftarrow \begin{bmatrix} C'QC + K'(R+S)K & -K'S \\ -SK & S \end{bmatrix} \\
&R \leftarrow R + S, \quad D \leftarrow \begin{bmatrix} D \\ I \end{bmatrix}, \quad G \leftarrow \begin{bmatrix} -DK & 0 \\ -K & I \end{bmatrix} \\
&d_{\max} \leftarrow \begin{bmatrix} u_{\max} - u_s \\ \Delta_u \end{bmatrix}, \quad d_{\min} \leftarrow \begin{bmatrix} u_{\min} - u_s \\ -\Delta_u \end{bmatrix}.
\end{aligned}$$

While formulation (19)-(20) is theoretically appealing, the solution is intractable in its current form, because it is necessary to consider an infinite number of decision variables. To obtain a computationally tractable formulation, we reformulate the optimization in a finite-dimensional decision space.

Several authors have considered this problem in various forms. We concentrate on the constrained linear quadratic methods proposed in the literature [31], [35]-[37]. The key concept behind these methods is to recognize that the inequality constraints remain active only for a finite number of sample steps along the prediction horizon. We demonstrate informally this concept as follows: if we assume that there exists a feasible solution to (19), (20), then the state and input trajectories $\{w_k, v_k\}_{k=0}^{\infty}$ approach the origin exponentially. Furthermore, if we assume the origin is contained in the interior of the feasible region $\mathbb{W} \times \mathbb{V}$, then there exists a positively invariant convex set [38]

$$\mathcal{O}_\infty = \left\{ w \mid (A + BK)^j w \in \mathbb{W}_K, \quad \forall j \ge 0 \right\}$$

such that the optimal unconstrained feedback law $v = Kw$ is feasible for all future time. The set $\mathbb{W}_K$ is the feasible region projected onto the state space by the linear control K (i.e., $\mathbb{W}_K = \{w \mid (w, Kw) \in \mathbb{W} \times \mathbb{V}\}$). Because the state and input trajectories approach the origin exponentially, there exists a finite N^* such that the state trajectory $\{w_k\}_{k=N^*}^{\infty}$ is contained in $\mathcal{O}_\infty$.

To guarantee that the inequality constraints (20b) are satisfied on the infinite horizon, N^* must be chosen such that $w_{N^*} \in \mathcal{O}_\infty$. Since the value of N^* depends on x_j, we need to account for the variable decision horizon length in the optimization. We formulate the variable horizon length regulator as the following optimization:

$$\min_{\{w_k, v_k, N\}} \Phi(x_j) = \frac{1}{2}\sum_{k=0}^{N-1} \left[w_k' Q w_k + v_k' R v_k + 2 w_k' M v_k \right] + (1/2) w_N' \Pi w_N \tag{21}$$

subject to the constraints

$$w_0 = x_j, \qquad w_{k+1} = A w_k + B v_k, \qquad w_N \in \mathcal{O}_\infty \tag{22a}$$

$$d_{\min} \le D v_k - G w_k \le d_{\max}, \tag{22b}$$

$$y_{\min} - y_s \le C w_k \le y_{\max} - y_s. \tag{22c}$$

The cost to go Π is determined from the discrete-time algebraic Riccati equation

$$\Pi = A'\Pi A + Q - (A'\Pi B + M)(R + B'\Pi B)^{-1}(B'\Pi A + M'), \tag{23}$$

for which many reliable solution algorithms exist. The variable horizon formulation is similar to the dual-mode receding horizon controller [39] for nonlinear systems with the linear quadratic regulator chosen as the stabilizing linear controller.

While the problem (21)-(22) is formulated on a finite horizon, the solution cannot be obtained, in general, in real time since the problem is a mixed-integer program. Rather than try to solve (21)-(22) directly, we address the problem of determining N^* from a variety of semi-implicit schemes while maintaining the quadratic programming structure in the subsequent optimizations.

Gilbert and Tan [38] show that there exists a finite number t^* such that $\mathcal{O}_{t^*}$ is equivalent to the maximal $\mathcal{O}_\infty$, in which

$$\mathcal{O}_t = \left\{ w \mid (A + BK)^j w \in \mathbb{W}_K, \quad \text{for } j = 0, \ldots, t \right\}. \tag{24}$$

They also present an algorithm for determining t^* that is formulated efficiently as a finite number of linear programs. Their method provides an easy check whether, for a fixed N, the solution to (21)-(22) is feasible (i.e., $w_N \in \mathcal{O}_\infty$). The check consists of determining whether state and input trajectories generated by unconstrained control law $v_k = Kw_k$ from the initial condition w_N are feasible with respect to inequality constraints for t^* time steps in the future. If the check fails, then the optimization (21)-(22) needs to be resolved with a longer control horizon $N' > N$ since $w_N \notin \mathcal{O}_\infty$. The process is repeated until $w_{N'} \in \mathcal{O}_\infty$.

When the set of initial conditions $\{w_0\}$ is compact, Chmielewski and Manousiouthakis [36] present a method for calculating an upper bound $\bar{N}$ on N^* using bounding arguments on the optimal cost function Φ^*. Given a set $\mathbb{P} = \{x^1, \ldots, x^m\}$ of initial conditions, the optimal cost function $\Phi^*(x)$ is a convex function defined on the convex hull (co) of $\mathbb{P}$. An upper bound $\bar{\Phi}(x)$ on the optimal cost $\Phi^*(x)$ for $x \in \text{co}(\mathbb{P})$ is obtained by the corresponding convex combinations of optimal cost functions $\Phi^*(x^j)$ for $x^j \in \mathbb{P}$. The upper bound on N^* is obtained by recognizing that the state trajectory w_j only remains outside of $\mathcal{O}_\infty$ for a finite number of stages. A lower bound q on the cost of $w_j'Qw_j$ can be generated for $x_j \notin \mathcal{O}_\infty$ (see [36] for explicit details). It then follows that $N^* \le \bar{\Phi}(x)/q$. Further refinement of the upper bound can be obtained by including the terminal stage penalty Π in the analysis.

Finally, an efficient solution of the quadratic program generated by the MPC regulator is discussed in [40] and [41].

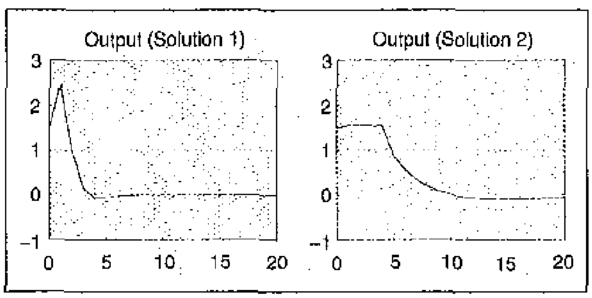

Figure 3. *Two controllers' resolution of output infeasibility: output versus time. Solution (1) minimizes duration of constraint violation; Solution (2) minimizes peak size of constraint violation.*

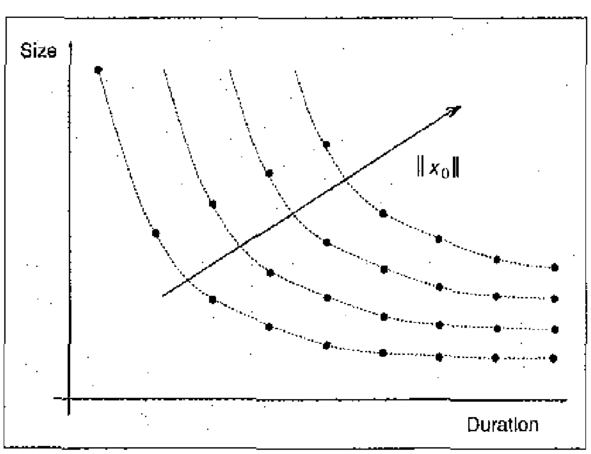

Figure 4. *Pareto optimal curves for size versus duration of constraint violation as a function of initial condition x_0.*

Feasibility

In the implementation of MPC, process conditions arise where there is no solution to the optimization problem (21) that satisfies the constraints (22). Rather than declaring such situations process exceptions, we sometimes prefer a solution that enforces some of the inequality constraints while relaxing others to retain feasibility. Often the input constraints represent physical limitations such as valve saturation that cannot be violated. Output constraints, however, frequently do not represent hard physical bounds. Rather, they often represent desired ranges of operations that can be violated if necessary. To avoid infeasibilities, we relax the output constraints by treating them as "soft" constraints.

Various authors have considered formulating output constraints as soft constraints to avoid potential infeasibilities [26]-[30]. We focus on the l_1/l_2^2 exact soft constraint strategy first advocated by de Oliveira and Biegler [28]. The attractive feature of the l_1/l_2^2 formulation is that the quadratic programming structure is retained and the resulting solution is exact if a feasible solution exists.

Multiobjective Nature of Infeasibility Problems

In many plants, the simultaneous minimization of the size and duration of the state constraint violations is not a conflicting objective. The optimal way to handle infeasibility is simply to minimize both size and duration; regulator performance may then be optimized, subject to the "optimally" relaxed state constraints. Unfortunately, not all infeasibilities are as easily resolved. In some cases, such as nonminimum-phase plants, a reduction in size of violation can only be obtained at the cost of a large increase in duration of violation, and vice versa. The optimization of constraint violations then becomes a multiobjective problem. In Fig. 3 we show two different controllers' resolution of an infeasibility problem.

The two-state, single-input/single-output (SISO) system model is

$$x_{k+1} = \begin{bmatrix} 1.6 & -0.64 \\ 1 & 0 \end{bmatrix} x_k + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u_k$$
$$y_k = [-1 \quad 2] x_k$$

with constraints and initial condition

$$|y_k| \le 1, \quad x_0 = [1.5 \quad 1.5]'.$$

Solution (1) corresponds to a controller minimizing the duration of constraint violation, which leads to a large peak violation, and solution (2) corresponds to a controller minimizing the peak constraint violation, which leads to a long duration of violation. This behavior is a system property caused by the unstable zero and cannot be avoided by clever controller design. For a given system and horizon N, the Pareto optimal size/duration curves can be plotted for different initial conditions, as in Fig. 4. The user must then decide where in the size/duration plane the plant should operate at times of infeasibility. Desired operation may lie on the Pareto optimal curve, because points below this curve cannot be attained and points above it are inferior, in the sense that they correspond to larger sizes and/or durations than are required.

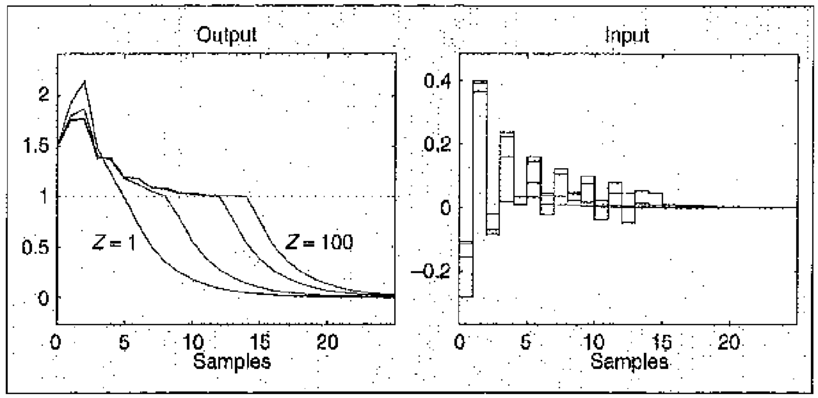

Figure 5. *Least-squares soft constraint solution. $Z = 1, 10, 50$ and 100. Solid lines: closed-loop; dashed lines: open-loop predictions at time 0; dotted line: output upper constraint.*

We next construct soft output inequality constraints by introducing the slack variable ε_k into the optimization. We reformulate the variable horizon regulator with soft constraints as the following optimization:

$$\min_{\{w_k, v_k, N\}} \Phi(x_j) = \frac{1}{2}\sum_{k=0}^{N-1}\left[w_k' Q w_k + v_k' R v_k + 2w_k' M v_k + \varepsilon_k' Z \varepsilon_k + z'\varepsilon_k\right] + (1/2) w_N' \Pi w_N$$

subject to the constraints

$$\begin{aligned}
&w_0 = x_j,\ w_{k+1} = Aw_k + Bv_k,\ w_N \in \mathcal{O}_\infty \\
&d_{\min} \le Dv_k - Gw_k \le d_{\max} \\
&(y_{\min} - y_s) - \varepsilon_k \le Cw_k \\
&Cw_k \le (y_{\max} - y_s) + \varepsilon_k \\
&\varepsilon_k \ge 0.
\end{aligned}$$

We assume Z is a symmetric positive definite matrix and z is a vector with positive elements chosen such that output constraints can be made exact if desired.

As a second example, consider the third-order nonminimum phase system

$$A = \begin{bmatrix} 2 & -1.45 & 0.35 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \tag{25}$$

$$C = [-1 \ \ 0 \ \ 2] \tag{26}$$

for which the output displays inverse response. The controller tuning parameters are $Q = C'C$, $R = 1$, and $N = 20$. The input is unconstrained, the output is constrained between ± 1, and we perform simulations from the initial condition $x_0 = [1.5 \ \ 1.5 \ \ 1.5]'$. Fig. 5 shows the possible trade-offs that can be achieved by adjusting the quadratic soft-constraint penalty, Z. We also see that open-loop predictions and nominal closed-loop responses are in close agreement for all choices of tuning parameter.

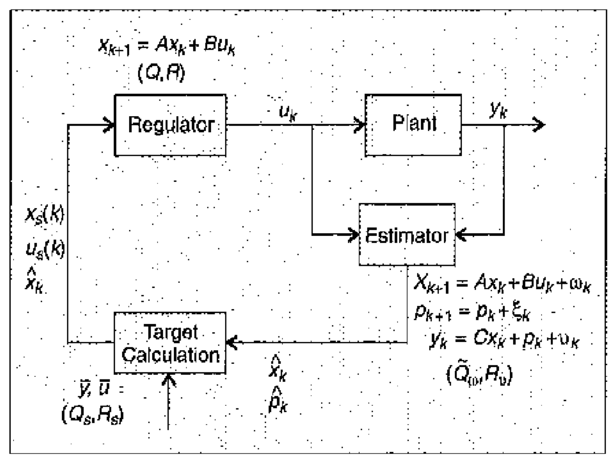

Figure 6. *MPC controller consisting of receding horizon regulator, state estimator, and target calculator.*

State Estimation

We now turn to reconstruction of the state from output measurements. In the model of (10), the nonzero disturbances d and p are employed to give offset-free control in the face of nonzero disturbances. The original industrial MPC formulations [GPC, quadratic dynamic matrix control (QDMC), identification command (IDCOM)] were designed for offset-free control by using an integrating output disturbance model. The integrating output disturbance model is a standard device in LQ design [42], [43]. Similarly, to track nonzero targets or desired trajectories that asymptotically approach nonzero values, one augments the plant model

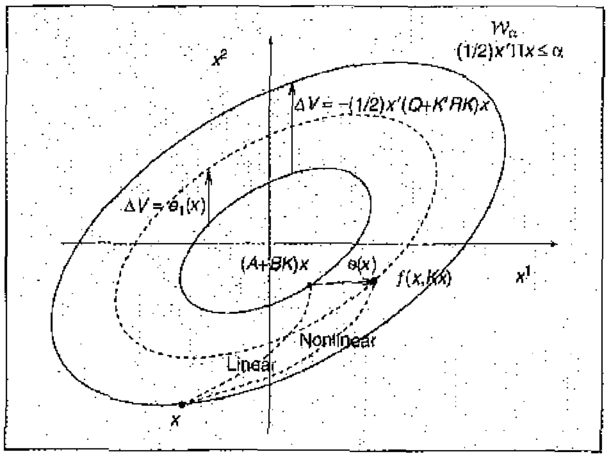

Figure 7. *Region W level sets of $(1/2)x'\Pi x$, and the effect of nonlinearity in a neighborhood of the setpoint.*

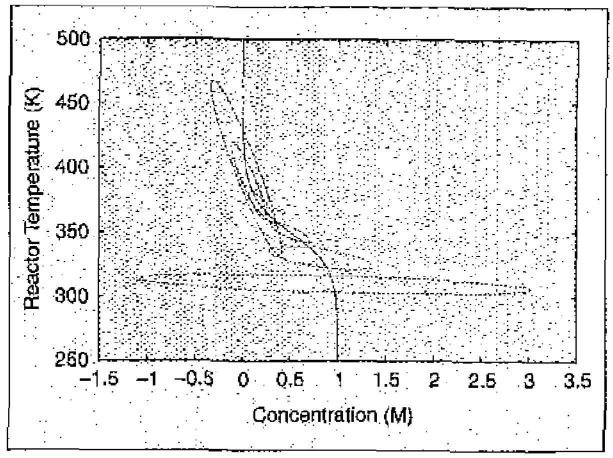

Figure 8. *W regions surrounding the locus of steady-state operating points. Solid line: steady states; dashed lines: W regions.*

dynamics with integrators. The disturbances may be modeled at the input, output, or some combination. These disturbance models are not used in the regulator; the disturbances are obviously uncontrollable and are required only in the state estimator. The effects of the disturbance estimates is to shift the steady-state target of the regulator. Bitmead et al. [10] provide a nice discussion of the disturbance models popular in GPC. Lee et al. [44] discuss the equivalence between the original industrial MPC algorithms and different disturbance model choices. Shinskey [45] provides a good discussion of the disadvantages of output disturbance models, in the original DMC formulation, compared to input disturbance models.

We set $d = 0$ and for simplicity focus on the output disturbance model. We augment the state of the system so the estimator produces estimates of both state, $\hat{x}$, and modeled disturbance, $\hat{p}$, with the standard Kalman filtering equations. The disturbance may be modeled by passing white noise, ξ_k, through an integrator, or by passing white noise through some other stable linear system (filter) and then through an integrator. The disturbance-shaping filter enables the designer to attenuate disturbances with selected frequency content. Bitmead et al. [10] provide a tutorial discussion of these issues in the unconstrained predictive control context.

In the simplest case, the state estimator model takes the form

$$x_{j+1} = Ax_j + Bu_j + \omega_j \tag{27}$$

$$p_{j+1} = p_j + \xi_j \tag{28}$$

$$y_j = Cx_j + p_j + \nu_j \tag{29}$$

in which ω_j, ξ_j, ν_j are the noises driving the process, integrated disturbance, and output measurement, respectively. As shown in Fig. 6, we specify $\tilde{Q}_\omega = \mathrm{diag}(Q_\omega, Q_\xi), R_\nu$, which are the covariances of the zero mean, normally distributed noise terms. The optimal state estimate for this model is given by the classic Kalman filter equations [46]. As in standard LQG design, one can tune the estimator by choosing the relative magnitudes of the noises driving the state, integrated disturbance, and measured output. Practitioners certainly would prefer tuning parameters more closely tied to closed-loop performance objectives, and more guidance on MPC tuning, in general, remains a valid research objective.

Assembling the components of the previous sections produces the structure shown in Fig. 6.

This structure is certainly not the simplest that accounts for output feedback, nonzero setpoints and disturbances, and offset-free control, nor is it the structure found in the dominant commercial vendor products. It is presented here mainly as a prototype to display a reasonably flexible means of handling these critical issues. Something similar to this structure has been implemented by industrial practitioners with success, however [47].

MPC with Nonlinear Models

What Is Desirable and What Is Possible

From the practical side, industrial implementation of MPC ith nonlinear models has already been reported, so it is certainly *possible*. Bequette reviews both the industrial and academic predictive control literature up to 1990 [19]. A nice early industrial application is reported by Garcia, in which he uses repeated local linearization of a nonlinear model to control a semibatch polymerization reactor [48]. Qin and Badgwell provide an excellent summary of the emerging vendor products for nonlinear MPC [18]. Controller objectives also vary widely. In batch operations, output trajectories are often considered known, or determined at a higher level, and the controller's objective is to track the specified dynamic trajectory. In continuous operations, steady-state targets may be considered known, and the controller is to find the optimal trajectory to the steady state. Sometimes the state targets change abruptly and the controller's objective is to perform the grade transition smoothly. Sometimes a reference output trajectory is provided for continuous operations, and, as in batch operations, the controller seeks to follow the given output trajectory.

Representing or approximating a nonlinear model's dynamic response with some form of linear dynamics is a recurring theme in much of the literature. The motivation is clearly to obtain a more easily solved online optimization. One issue that seems relatively neglected, however, is that obtaining these updated linearized models requires knowledge of the state, which may be either the current state or the desired target steady state. The most popular technique for estimating the state in the early literature is to mimic the linear case, solve the state equations in an open-loop fashion, and describe the difference between measured output and model forecast as an integrating disturbance. This method works well in the linear case where the model dynamics are independent of the state. It is unlikely that integrating the state equations in open loop is a general approach for applications requiring nonlinear models. As the current model state deviates from the plant due to open-loop integration of model errors, the current model's linearization loses any connection to the true dynamics. Bequette concludes that the *most* important issue in implementing nonlinear MPC is obtaining good state estimates [19], and more attention is being focused on the state estimation part of the nonlinear control problem. It is interesting to note that Qin and Badgwell report that two recent nonlinear MPC vendor products provide state estimation functionality in the form of the extended Kalman filter (EKF) as well as the standard MPC regulation functionality [18].

The industrial nonlinear MPC implementations are largely without any established closed-loop properties, even nominal closed-loop properties. A lack of supporting theory should not and does not, according to historical record, discourage experiments in practice with promising new technologies. But if nonlinear MPC is to become widespread in the harsh environment of applications, it must eventually become reasonably reliable, predictable, efficient, and robust against online failure.

From the theoretical side, it would be desirable to solve in real time infinite horizon nonlinear optimal control problems of the type

$$\min_{\{x_k, u_k\}} \Phi(x_j) = \sum_{k=0}^{\infty} L(x_k, u_k) \qquad (30)$$

subject to the constraints

$$x_{k+1} = f(x_k, u_k), \;\; x_0 = x_j \qquad (31)$$

$$u_k \in \mathcal{U}, \quad x_k \in \mathcal{X}. \qquad (32)$$

Nonlinear MPC based on this optimal control problem would have the strongest provable closed-loop properties. The concomitant theoretical and computational difficulties associated with this optimal control problem, either offline, but especially online, are well known and formidable [3]. The current view of problem (30) is: *desirable*, but *not possible*. In the next two sections, we evince one viewpoint of the current status of bringing these two sides closer together.

State Feedback

As an attempt to approximately solve (30), it is natural to try to extend to the nonlinear case the ideas of the linear receding horizon regulator. In the linear case, we define a region in state space, $\mathcal{W}$, with the following properties:

$$\mathcal{W} \subset \mathcal{X}, \quad K\mathcal{W} \subset \mathcal{U}$$
$$x \in \mathcal{W} \Rightarrow (A + BK)x \in \mathcal{W}$$

which tells us that in $\mathcal{W}$ the state constraints are satisfied, the input constraints are satisfied under the unconstrained,

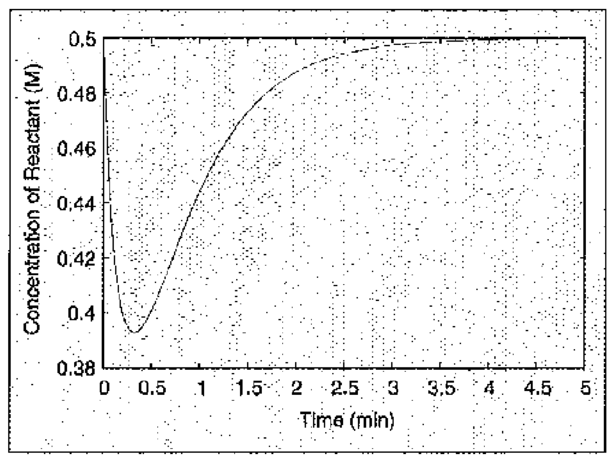

Figure 9. *Closed-loop response: C_A versus t.*

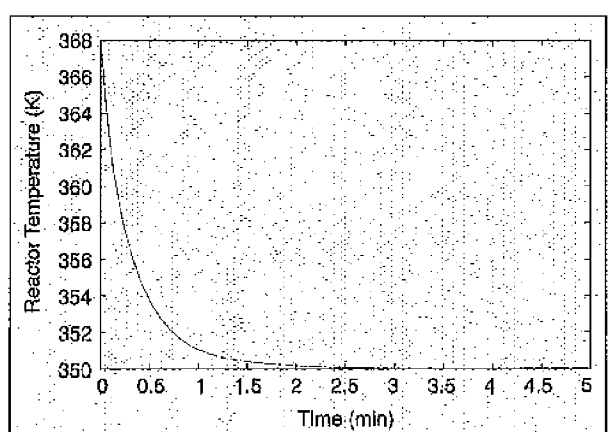

Figure 10. *Closed-loop response: T versus t.*

linear feedback law $u = Kx$, and once a state enters $\mathcal{W}$, it remains in $\mathcal{W}$ under this control law. We can compute the cost to go for $x \in \mathcal{W}$; it is $(1/2)x'\Pi x$ in which Π is given in (23). For the linear case, the Gilbert and Tan algorithm provides in many cases the largest set $\mathcal{W}$ with these properties, $\mathcal{O}_\infty$.

Ingredients of the Open-Loop, Optimal Control Problem

In the simplest extension to the nonlinear case, consider a region $\mathcal{W}$ with the analogous properties

$$\mathcal{W} \subset \mathcal{X}, \quad K\mathcal{W} \subset \mathcal{U}$$
$$x \in \mathcal{W} \Rightarrow f(x, Kx) \in \mathcal{W}.$$

The essential difference is that we must, under the nonlinear model, ensure that the state remains in $\mathcal{W}$ with the linear control law. Again, for this simplest version, we determine the linear control law by considering the linearization of f at the setpoint:

$$A = \frac{\partial f}{\partial x}(0,0), \quad B = \frac{\partial f}{\partial u}(0,0), \quad C = \frac{\partial g}{\partial x}(0).$$

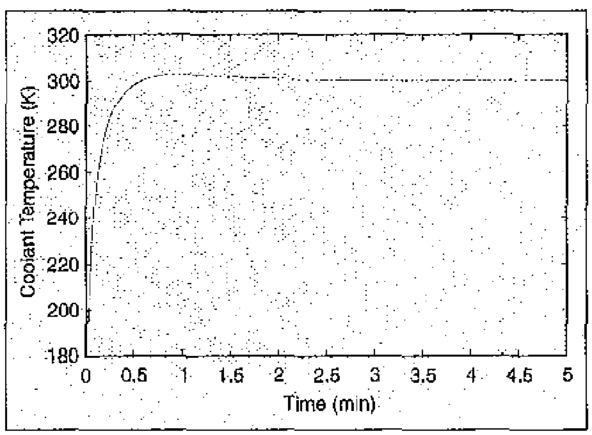

Figure 11. *Manipulated variable: T_c versus t.*

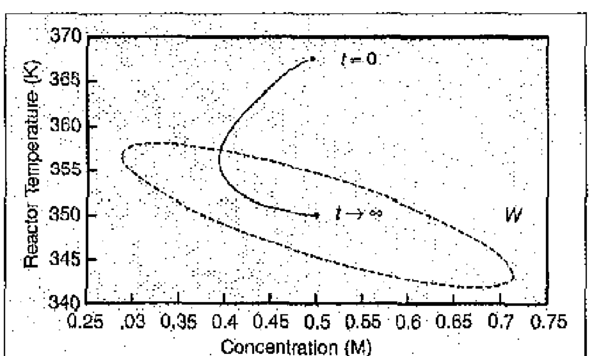

Figure 12. *Closed-loop phase portrait under nonlinear MPC: Region $\mathcal{W}$ and closed-loop response T versus C_A.*

For the nonlinear case, we cannot easily compute the largest region with these properties, but we can find a finite-sized region with these properties. Chen and Allgöwer, therefore, refer to this approach with nonlinear systems as "quasi-infinite horizon" [49]. The main assumption required is that f's partial derivatives are Lipschitz continuous [50] to bound the size of the nonlinear effects and that the linearized system is controllable. Most chemical processes satisfy these assumptions, and with them we can construct region $\mathcal{W}$ as shown in Fig. 7.

Let Q and R represent the quadratic approximation of the stage cost function at the origin

$$Q = L_{xx}(0,0), \quad R = L_{uu}(0,0).$$

Consider the quadratic function, V, and level sets of this function, $\mathcal{W}_\alpha$

$$V(x) = (1/2)x'\Pi x, \quad \mathcal{W}_\alpha = \{x | V(x) \le \alpha\}$$

for α a positive scalar. Define $e(x)$ to be the difference between the state propagation under the nonlinear model and linearized model, $e(x) = f(x, Kx) - (A + BK)x$ and $e_1(x)$, to be the difference in V at these two states, $e_1(x) = V(f(x, Kx)) - V((A+BK)x)$. We can show that near the setpoint (origin)

$$\|e(x)\| \le c\|x\|^2, \quad |e_1(x)| \le d\|x\|^3$$

which bounds the effect of the nonlinearity. We can, therefore, find an α such that the finite horizon control law with terminal constraint and approximate cost to go penalty is stabilizing

$$\min_{\{x_k, u_k, N\}} \Phi(x_j) = \sum_{k=0}^{N-1} L(x_k, u_k) + (1/2)x_N'\Pi x_N \tag{33}$$

subject to the constraints

$$x_{k+1} = f(x_k, u_k), \; x_0 = x_j$$
$$u_k \in \mathcal{U}, \quad x_k \in \mathcal{X}, \quad x_N \in \mathcal{W}_\alpha.$$

We choose α such that

$$\max_{x \in \mathcal{W}_\alpha} \{V(f(x, Kx)) - V(x) + (1/4)x'(Q + K'RK)x\} \le 0. \tag{34}$$

It has also been established that global optimality in (33) is not required for closed-loop stability [49], [50]. Calculation of the $\mathcal{W}_\alpha$ region in (34) remains a challenge, particularly when the target calculation and state estimation and disturbance models are added to the problem as described earlier. Under those circumstances, $\mathcal{W}_\alpha$, which depends on the current steady target, changes at each sample. It may be possible that some of this computation can be performed offline, but resolving this computational issue remains a research challenge.

We present a brief example to illustrate these ideas. Consider the simple model presented by Henson and Seborg [51] for a continuously stirred tank reactor (CSTR) undergoing reaction A → B at an unstable steady state:

$$\frac{dC_A}{dt} = \frac{q}{V}(C_{Af} - C_A) - kC_A$$
$$\frac{dT}{dt} = \frac{q}{V}(T_f - T) + \frac{(-\Delta H)}{\rho C_p} kC_A + \frac{UA}{V\rho C_p}(T_c - T)$$
$$k = k_0 \exp(-E/RT).$$

Fig. 8 displays the W regions computed by solving (34) along the locus of steady-state operating points.

For the steady-state operating point

$$T_s = 350K, \quad C_{As} = 0.5M, \quad T_{cs} = 300K,$$

the closed-loop behavior of the states with MPC control law (33) is shown in Figs. 9-10. The manipulated variable is shown in Fig. 11.

Fig. 12 displays a phase-portrait of the two states converging to the setpoint and the terminal region W.

Future Developments

Although this article is intended as a tutorial, brief consideration of areas of future development may prove useful. The theory for nominal MPC with linear models and constraints is reasonably mature in that nominal properties are established, and efficient computational procedures are available. The role of constraints is reasonably well understood. Applications in the process industries are ubiquitous.

MPC with Nonlinear Models

In MPC for nonlinear models, the territory is much less explored. The nonconvexity of the optimal control problems presents theoretical and computational difficulties. The research covered in this tutorial on quasi-infinite horizons and suboptimal MPC provide one avenue for future development [52], [49]. Contractive MPC [53]-[55] and exact linearization MPC [57], [58] are two other alternatives that show promise. Mayne et al. [59] and De Nicolao et al. [60] provide recent reviews of this field for further reading. It is expected, as in the case of linear MPC of the 1970s and 1980s, that these theoretical hurdles will not impede practitioners from evaluating nonlinear MPC.

Indeed, as summarized by Qin and Badgwell, vendors are actively developing new nonlinear MPC products [18], and many new industrial applications are appearing [61]. A variety of different nonlinear model forms are being pursued, including NARMAX and neural network models.

Robustness

Robustness to various types of uncertainty and model error is, of course, an active research area in MPC as well as in other areas of automatic control. The difficulty that MPC introduces into the robustness question is the open-loop nature of the optimal control problem and the implicit feedback produced by the receding horizon implementation. Several robust versions of MPC have been introduced that address this issue [62], [27], [63]. Lee and Yu [64] define a dynamic programming problem for the worst-case cost. Badgwell [65] appends a set of robustness constraints to the open-loop problem, which ensures robustness for a finite set of plants. Kothare et al. [66] address the feedback issue by optimization over the state feedback gain rather than the open-loop control sequence subject to constraints.

The rapid development of time domain worst-case controller design problems as dynamic games (see [67] for an excellent summary) has led to further proposals for robust MPC exploiting this connection to H_∞ theory [68], [69], [60]. At this early juncture, online computation of many of the robust MPC control laws appears to be a major hurdle for practical application.

Moving Horizon Estimation

The use of optimization subject to a dynamic model is the underpinning for much of state estimation theory. A moving horizon approximation to a full infinite horizon state estimation problem has been proposed by several researchers [70]-[72]. The theoretical properties of this framework are only now emerging [73], [74]. Again, attention should be focused on what key issues of practice that are out of reach with previous approaches can be addressed in this framework. Because moving horizon estimation with linear models produces simple, positive definite quadratic programs, online implementation is possible today for many process applications. The use of constraints on states or state disturbances presents intriguing opportunities, but it is not clear what applications benefit from using the extra physical knowledge in the form of constraints. Nonlinear, fundamental models coupled with moving horizon state estimation may start to play a larger role in process operations. State estimation for unmeasured product properties based on fundamental, nonlinear models may have more impact in the short term than closed-loop regulation with these models. State estimation using empirical, nonlinear models is already being used in commercial process monitoring software. Moreover, state estimation is a wide-ranging technique for addressing many issues of process operations besides feedback control, such as process monitoring, fault detection, and diagnosis.

MPC for Hybrid Systems

Essentially all processes contain discrete as well as continuous components: on/off valves, switches, logical overrides, and so forth. Slupphaug and Foss [75], [76] and Bemporad and Morari [77]-[79] have considered application of MPC in this environment. This problem class is rich with possibilities: for example, the rank ordering rather than softening of output constraints to handle infeasibility. Some of the intriguing questions at this early juncture are: how far can this framework be pushed, are implementation bottlenecks expected from the system modeling or online computations, what benefits can be obtained compared to traditional heuristics, and what new problem types can be tackled?

Acknowledgments

I wish to acknowledge Dr. Thomas A. Badgwell for numerous helpful discussions of this material during our co-teaching of industrial MPC short courses and Matthew J. Tenny and Christopher V. Rao for their help in preparing the manuscript and figures.

The financial support of the National Science Foundation, through grant CTS-9708497, and the industrial members of the Texas-Wisconsin Modeling and Control Consortium is gratefully acknowledged.

References

[1] J.B. Rawlings, "Tutorial: Model predictive control technology," in *Proc. American Control Conf.*, San Diego, CA, 1999, pp. 662-676.

[2] J.H. Lee and B. Cooley, "Recent advances in model predictive control and other related areas," in *Proc. Chemical Process Control - V*, J.C. Kantor, C.E. Garcia, and B. Carnahan, Eds., CACHE, AIChE, 1997, pp. 201-216.

[3] D.Q. Mayne, "Nonlinear model predictive control: An assessment," in *Proc. Chemical Process Control - V*, J.C. Kantor, C.E. Garcia, and B. Carnahan, Eds., CACHE, AIChE, 1997, pp. 217-231.

[4] S.J. Qin and T.A. Badgwell, "An overview of industrial model predictive control technology," in *Proc. Chemical Process Control-V*, J.C. Kantor, C.E. Garcia, and B. Carnahan, Eds., CACHE, AIChE, 1997, pp. 232-256.

[5] H. Chen and F. Allgöwer, "Nonlinear model predictive control schemes with guaranteed stability," in *NATO AM on Nonlinear Model Based Process Control*, R. Berber and C. Kravaris, Eds. Norwell, MA: Kluwer, 1998, pp. 465-494.

[6] M. Morari and J.H. Lee, "Model predictive control: past, present and future," in *Proc. Joint 6th Int. Symp. Process Systems Engineering* (PSE '97) *and 30th European Symp. Computer Aided Process Systems Engineering* (ESCAPE 7), Trondheim, Norway, 1997.

[7] W.H. Kwon, "Advances in predictive control: theory and application," in *Proc. 1st Asian Control Conf.*, Tokyo, 1994.

[8] E. Mosca, *Optimal, Predictive and Adaptive Control*. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[9] R. Soeterboek, *Predictive Control—A Unified Approach*. Englewood Cliffs, NJ: Prentice-Hall, 1992.

[10] R.R. Bitmead, M. Gevers, and V. Wertz, *Adaptive Optimal Control, The Thinking Man's GPC*. Englewood Cliffs, NJ: Prentice-Hall, 1990.

[11] F. Allgöwer and A. Zheng, Eds., *Int. Symp. Nonlinear Model Predictive Control*, Ascona, Switzerland, 1998.

[12] F. Allgöwer, T.A. Badgwell, S.J. Qin, J.B. Rawlings, and S.J. Wright, "Nonlinear predictive control and moving horizon estimation—An introductory overview," in *Advances in Control: Highlights of ECC'99*, P.M. Frank, Ed., London: Springer, 1999, pp. 391-449.

[13] J. Richalet, A. Rault, J.L. Testud, and J. Papon, "Model predictive heuristic control: Applications to industrial processes," *Automatica*, vol. 14, pp. 413-428, 1978.

[14] C.R. Cutler and B.L. Ramaker, "Dynamic matrix control—A computer control algorithm," in *Proc. Joint Automatic Control Conf.*, 1980.

[15] D.M. Prett and R.D. Gillette, "Optimization and constrained multivariable control of a catalytic cracking unit," in *Proc. Joint Automatic Control Conf.*, San Francisco, CA, 1980, pp. WP5-C.

[16] R.K. Pearson and B.A. Ogunnaike, "Nonlinear process identification," in *Nonlinear Process Control*, M.A. Henson and D.E. Seborg, Eds. Englewood Cliffs, NJ: Prentice-Hall, 1997, pp. 11-110.

[17] J.H. Lee, "Modeling and identification for nonlinear model predictive control: Requirements, current status and future needs," in *Proc. Int. Symp. Nonlinear Model Predictive Control*, Ascona, Switzerland, 1998.

[18] S.J. Qin and T.A. Badgwell, "An overview of nonlinear model predictive control applications," in *Proc. Int. Symp. Nonlinear Model Predictive Control*, Ascona, Switzerland, 1998.

[19] B.W. Bequette, "Nonlinear predictive control using multi-rate sampling," *Can. J. Chem. Eng.*, vol. 69, pp. 136-143, Feb. 1991.

[20] C.V. Rao and J.B. Rawlings, "Steady states and constraints in model predictive control," *AIChE J.*, vol. 45, no. 6, pp. 1266-1278, 1999.

[21] D.Q. Mayne, "Optimization in model based control," in *Proc. IFAC Symp. Dynamics and Control of Chemical Reactors, Distillation Columns and Batch Processes*, Helsingor, Denmark, June 1995, pp. 229-242.

[22] K.R. Muske and J.B. Rawlings, "Model predictive control with linear models," *AIChE J.*, vol. 39, no. 2, pp. 262-287, 1993.

[23] K.R. Muske, "Steady-state target optimizaton in linear model predictive control," in *Proc. American Control Conf.*, Albuquerque, NM, June 1997, pp. 3597-3601.

[24] R. Fletcher, *Practical Methods of Optimization*. New York: Wiley, 1987.

[25] W.W. Hager, "Lipschitz continuity for constrained processes," *SIAM J. Cont. Opt.*, vol. 17, no. 3, pp. 321-338, 1979.

[26] N.L. Ricker, T. Subrahmanian, and T. Sim, "Case studies of model-predictive control in pulp and paper production," in *Proc. 1988 IFAC Workshop Model Based Process Control*, T.J. McAvoy, Y. Arkun, and E. Zafiriou, Eds. New York: Pergamon, 1988, pp. 13-22.

[27] H. Genceli and M. Nikolaou, "Robust stability analysis of constrained l1-norm model predictive control," *AIChE J.*, vol. 39, no. 12, pp. 1954-1965, 1993.

[28] N.M.C. de Oliveira and L.T. Biegler, Constraint handling and stability properties of model-predictive control," *AIChE J.*, vol. 40, no. 7, pp. 1138-1155, 1994.

[29] A. Zheng and M. Morari, "Stability of model predictive control with mixed constraints," *IEEE Trans. Automat. Contr.*, vol. 40, pp. 1818-1823, Oct. 1995.

[30] P.O. Scokaert and J.B. Rawlings, "Feasibility issues in linear model predictive control," *AIChE J.*, vol. 45, pp. 1649-1659, Aug. 1999.

[31] S.S. Keerthi, "Optimal feedback control of discrete-time systems with state-control constraints and general cost functions," Ph.D. dissertation, Univ. Michigan, 1986.

[32] J.A. Rossiter, M.J. Rice, and B. Kouvaritakis, "A robust stable state-space approach to stable predictive control stategies," in *Proc. American Control Conf.*, Albuquerque, NM, June 1997, pp. 1640-1641.

[33] L.T. Biegler, "Efficient solution of dynamic optimization and NMPC problems," in *Proc. Int. Symp. Nonlinear Model Predictive Control*, Ascona, Switzerland, 1998.

[34] H.G. Bock, M. Diehl, D.B. Leineweber, and J.P. Schlösser, "A direct multiple shooting method for real-time optimization of nonlinear DAE processes," in *Int. Symp. Nonlinear Model Predictive Control*, Ascona, Switzerland, 1998.

[35] M. Sznaier and M.J. Damborg, "Suboptimal control of linear systems with state and control inequality constraints," in *Proc. 26th Conf. Decision and Control*, Los Angeles, CA, 1987, pp. 761-762.

[36] D. Chmielewski and V. Manousiouthakis, "On constrained infinite-time linear quadratic optimal control," *Syst. Contr. Lett.*, vol. 29, pp. 121-129, 1996.

[37] P.O. Scokaert and J.B. Rawlings, "Constrained linear quadratic regulation," *IEEE Trans. Automat. Contr.*, vol. 43, pp. 1163-1169, Aug. 1998.

[38] E.G. Gilbert and K.T. Tan, "Linear systems with state and control constraints: The theory and application of maximal output admissible sets," *IEEE Trans. Automat. Contr.*, vol. 36, pp. 1008-1020, Sept. 1991.

[39] H. Michalska and D.Q. Mayne, "Robust receding horizon control of constrained nonlinear systems," *IEEE Trans. Automat. Contr.*, vol. 38, pp. 1623-1633, Nov. 1993.

[40] S.J. Wright, "Applying new optimization algorithms to model predictive control," in *Chemical Process Control-V*, J.C. Kantor, C.E. Garcia, and B. Carnahan, Eds., CACHE, AIChE, 1997, pp. 147-155.

[41] C.V. Rao, S.J. Wright, and J.B. Rawlings, "On the application of interior point methods to model predictive control," *J. Optim. Theory Applicat.*, vol. 99, pp. 723-757, 1998.

[42] E.J. Davison and H.W. Smith, "Pole assignment in linear time-invariant multivariable systems with constant disturbances," *Automatica*, vol. 7, pp. 489-498, 1971.

[43] H. Kwakernaak and R. Sivan, *Linear Optimal Control Systems*. New York: Wiley, 1972.

[44] J.H. Lee, M. Morari, and C.E. Garcia, "State-space interpretation of model predictive control," *Automatica*, vol. 30, no. 4, pp. 707-717, 1994.

[45] F.G. Shinskey, *Feedback Controllers for the Process Industries*. New York: McGraw-Hill, 1994.
[46] B.D.O. Anderson and J.B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.
[47] J.J. Downs, "Control strategy design using model predictive control," in *Proc. American Control Conf.*, San Diego, CA, 1999.
[48] C.E. Garcia, "Quadratic dynamic matrix control (QDMC) of nonlinear processes: An application to a batch reaction process," presented at AIChE National Meeting, San Francisco, CA, Nov. 1984.
[49] H. Chen and F. Allgöwer, "A quasi-infinite horizon nonlinear model predictive control scheme with guaranteed stability," *Automatica*, vol. 34, no. 10, pp. 1205-1217, 1998.
[50] P.O.M. Scokaert, D.Q. Mayne, and J.B. Rawlings, "Suboptimal model predictive control (feasibility implies stability)," *IEEE Trans. Automat. Contr.*, vol. 44, pp. 648-654, Mar. 1999.
[51] M.A. Henson and D.E. Seborg, *Nonlinear Process Control*. Upper Saddle River, NJ: Prentice-Hall PTR, 1997.
[52] T. Parisini and R. Zoppoli, "A receding-horizon regulator for nonlinear systems and a neural approximation," *Automatica*, vol. 31, no. 10, pp. 1443-1451, 1995.
[53] E. Polak and T.H. Yang, "Moving horizon control of linear systems with input saturation and plant uncertainty—Part 1: Robustness," *Int. J. Contr.*, vol. 58, no. 3, pp. 613-638, 1993.
[54] E. Polak and T.H. Yang, "Moving horizon control of linear systems with input saturation and plant uncertainty—Part 2: Disturbance rejection and tracking," *Int. J. Contr.*, vol. 58, no. 3, pp. 639-663, 1993.
[55] Z. Yang and E. Polak, "Moving horizon control of nonlinear systems with input saturation, disturbances and plant uncertainty," *Int. J. Contr.*, vol. 58, pp. 875-903, 1993.
[56] M. Morari and S.L. de Oliveira, "Contractive model predictive control for constrained nonlinear systems," *IEEE Trans. Automat. Contr.*, 2000, to be published.
[57] S.L. de Oliveira, V. Nevistic, and M. Morari, "Control of nonlinear systems subject to input constraints," in *Proc. IFAC Symp. Nonlinear Control System Design*, Tahoe City, CA, 1995, pp. 15-20.
[58] M.J. Kurtz and M.A. Henson, "Input-output linearizing control of constrained nonlinear processes," *J. Proc. Cont.*, vol. 7, no. 1, pp. 3-17, 1997.
[59] D.Q. Mayne, J.B. Rawlings, C.V. Rao, and P.O.M. Scokaert, "Model predictive control: a review," *Automatica*, 2000, to be published.
[60] G. De Nicolao, L. Magni, and R. Scattolini, "Stability and robustness of nonlinear receding horizon control," in *Proc. Int. Symp. Nonlinear Model Predictive Control*, Ascona, Switzerland, 1998.
[61] G. Martin, "Nonlinear model predictive control," in *Proc. American Control Conf.*, San Diego, CA, 1999.
[62] Z.Q. Zheng and M. Morari, "Robust stability of constrained model predictive control," in *Proc. 1993 American Control Conf.*, 1993, pp. 379-383.
[63] G. De Nicolao, L. Magni, and R. Scattolini, "Robust predictive control of systems with uncertain impulse response," *Automatica*, vol. 32, no. 10, pp. 1475-1479, 1996.
[64] J.H. Lee and Z. Yu, "Worst-case formulations of model predictive control for systems with bounded parameters," *Automatica*, vol. 33, no. 5, pp. 763-781, 1997.
[65] T.A. Badgwell, "Robust model predictive control of stable linear systems," *Int. J. Contr.*, vol. 68, no. 4, pp. 797-818, 1997.
[66] M.V. Kothare, V. Balakrishnan, and M. Morari, "Robust constrained model predictive control using linear matrix inequalities," *Automatica*, vol. 32, no. 10, pp. 1361-1379, 1996.
[67] T. Basar and P. Bernhard, *H^∞-Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*. Boston, MA: Birkauser, 1995.
[68] H. Chen, "Stability and robustness considerations in nonlinear model predictive control," Ph.D. dissertation, Univ. Stuttgart, 1997.
[69] H. Chen, C.W. Scherer, and F. Allgöwer, "A game theoretic approach to nonlinear robust receding horizon control of constrained systems," in *Proc. American Control Conf.*, Albuquerque, NM, 1997, pp. 3073-3077.
[70] K.R. Muske, J.B. Rawlings, and J.H. Lee, "Receding horizon recursive state estimation," in *Proc. 1993 American Control Conf.*, June 1993, pp. 900-904.
[71] D.G. Robertson, J.H. Lee, and J.B. Rawlings, "A moving horizon-based approach for least-squares state estimation," *AIChE J.*, vol. 42, pp. 2209-2224, Aug. 1996.
[72] M.L. Tyler and M. Morari, "Stability of constrained moving horizon estimation schemes," Preprint AUT96-1 8, Automatic Control Laboratory, Swiss Federal Institute of Technology, 1996.
[73] C.V. Rao and J.B. Rawlings, "Nonlinear moving horizon estimation," in *Proc. Int. Symp. Nonlinear Model Predictive Control*, Ascona, Switzerland, 1998.
[74] C.V. Rao, J.B. Rawlings, and J.H. Lee, "Stability of constrained linear moving horizon estimation," in *Proc. American Control Conf.*, San Diego, CA, 1999.
[75] O. Slupphaug and B.A. Foss, "Model predictive control for a class of hybrid systems," in *Proc. European Control Conf.*, Brussels, Belgium, 1997.
[76] O. Slupphaug, "On robust constrained nonlinear control and hybrid control BMI and MPC based state-feedback schemes," Ph.D. dissertation, Norwegian Univ. Science Technol., 1998
[77] A. Bemporad and M. Morari, "Predictive control of constrained hybrid systems," in *Proc. Int. Symp. Nonlinear Model Predictive Control*, Ascona, Switzerland, 1998.
[78] A. Bemporad and M. Morari, "Control of systems integrating logic, dynamics, and constraints," *Automatica*, vol. 35, pp. 407-427, 1999.
[79] M. Morari, A. Bemporad, and D. Mignone, "A framework for control, state estimation and verification of hybrid systems," *Automatisisierungstechnik*, vol. 47, no. 8, pp. 374-381, 1999.

James B. Rawlings received the B.S. degree from the University of Texas in 1979 and the Ph.D. degree from the University of Wisconsin in 1985, both in chemical engineering. He spent one year at the University of Stuttgart as a NATO Postdoctoral Fellow and then joined the faculty at the University of Texas. He moved to the University of Wisconsin in 1995 and is currently the Paul A. Elfers Professor of Chemical Engineering and the Codirector of the Texas-Wisconsin Modeling and Control Consortium (TWMCC). His research interests are in the areas of chemical process modeling, monitoring and control, nonlinear model predictive control, moving horizon state estimation, particulate systems modeling, and crystallization.