TWMCC ⋆ Texas-Wisconsin Modeling and Control Consortium

Technical report number 2002-04

Closed-loop Behavior of Nonlinear Model Predictive Control

Matthew J. Tenny*and James B. Rawlings†
Department of Chemical Engineering
University of Wisconsin-Madison
Madison, WI 53706-1691

Stephen J. Wright‡
Department of Computer Sciences
University of Wisconsin-Madison
Madison, WI 53706-1691

October 10, 2002

*tenny@bevo.che.wisc.edu
†Author to whom correspondence should be addressed. jbraw@bevo.che.wisc.edu
‡swright@cs.wisc.edu

Abstract

Model predictive control (MPC) relies on real-time optimization to determine open-loop control profiles and state estimates given process measurements. When the underlying process model is nonlinear, the MPC system exhibits unique behavior not seen in linear MPC. In this article, we highlight some of the characteristics of nonlinear models in the context of closed-loop performance. We examine the effects of disturbance models on closed-loop performance and show necessary conditions for how and where steady states of the closed-loop system may be found. We demonstrate these conditions on a simple example to show that the input disturbance model can lead to failure of the control system, and that linear MPC is inadequate for controlling this class of systems.

Additionally, due to nonconvexity, the optimization problems solved in nonlinear MPC may have local optima. These local minima may lead to undesirable performance, particularly in the state estimator. We study the existence of these optima in the regulator and estimator and examine their potential effects on the performance of the closed-loop system. To avoid unwanted local minima, we advocate the use of constraints in the estimator and regulator formulations and show how a *shorter* prediction horizon in the regulator leads to better control profiles for some nonlinear models.

1 Introduction

Over the last twenty years, linear model predictive control has developed into a popular and effective advanced control strategy. Among its strengths are its ability to handle constraints naturally within its framework and the relative simplicity of obtaining a linear model as a basis for the controller design. Unfortunately, linear models often do not adequately describe the dynamics of the process to be controlled except near the point at which the model was identified. This fact causes difficulties for processes with multiple operating points, such as those requiring grade transitions or set point changes based on economic criteria. Also, for plants in which nonzero mean unmodeled disturbances occur, the system may be shifted to a region in which the linear model is no longer valid.

For these reasons, the models used as the basis for control and estimation must be nonlinear to capture the behavior of the plant accurately in all its probable operating

regimes. Nonlinearities are common in chemical processes. They may arise from kinetics, for instance in the relationship between reaction rate and temperature in the Arrhenius equation or from higher-order reactions. Nonlinear equations are also found in thermodynamics, especially those involving phase equilibria for nonideal solutions. Because of the common occurrence of nonlinear behavior in even simple systems, nonlinearity is expected in the majority of chemical processes.

Nonlinear model predictive control is gaining popularity in the industrial community. The formulations for these controllers vary widely, and almost the only common principle is to retain nonlinearities in the process model. Recently, Qin and Badgwell [14] reviewed the use of commercial MPC software in industry. In their survey, they list nearly one hundred nonlinear MPC applications, spanning several engineering disciplines. However, the number of reported nonlinear MPC applications is far fewer than those of linear MPC, mainly because the linear MPC problem is simpler and provides adequate performance on problems with simple control performance specifications.

Two major obstacles toward implementing a nonlinear (rather than linear) MPC controller are in the identification of suitable nonlinear models and design of computationally tractable online algorithms. In recent years, researchers have attempted to address both of these problems. Nonlinear models can be identified in myriad ways. The empirial approaches range from high order polynomials and neural networks to piecewise linear models. Deterministic models are also available from first principles in many cases, and, if desired, these models can be reduced to a lower dimensional space. See [13, 7] for an overview of nonlinear model identification.

Several researchers have made recent contributions toward obtaining computationally viable methods for computing local solutions to nonlinear control problems in real-time. The primary tool for solving the nonlinear programs associated with MPC is successive quadratic programming. It has been shown that tailoring this approach to the structure of the MPC problem yields an efficient framework for solving these problems on-line [5, 23, 1]. In this paper, we solve the optimization problems for all three components of the MPC controller successfully within each sampling time, delivering real-time performance.

While the behavior of the closed-loop linear MPC system is well-studied, the behavior of the corresponding nonlinear system is not well characterized. In this paper, we investigate some of the features that are unique to nonlinear MPC. We consider the nonlinear model predictive control problem with moving horizon state estimation and a nonlinear steady-state target calculation. We focus on the aspects of the resulting closed-loop system that are specific to nonlinear models. One such issue is the effect of choosing integrating disturbance models to account for unmeasured nonzero mean disturbances. We demonstrate that the choice of disturbance models is crucial for a certain class of nonlinearity, and illustrate our findings with

two simple examples. We compare our findings to the results of linear MPC on the same system.

The second part of this paper concentrates on the appearance of local minima in the optimization problems in the MPC framework. Due to nonconvexity of the nonlinear process model constraints, multiple optima may be present for the regulator, estimator, and target calculation problems. We present examples of such local minima for the regulator and estimator problems and comment on their impact on closed-loop performance. We also motivate methods for avoiding undesired local minima by constraining the optimization problems. Also, in the case of the regulator problem, we show that an unnecessarily long prediction horizon may lead to local minima that are not globally optimal.

The remainder of this paper is organized as follows. In Section 2, we present the framework for the nonlinear MPC closed-loop system. Section 3 highlights the behavior and potential pitfalls associated with choosing an integrating disturbance model. Section 4 investigates the effects of finding local minima of the nonlinear control and estimation problems of Section 2. We summarize our results and outline points of possible future investigation in Section 5.

2 Formulation

For the purposes of this paper, we consider a control system composed of three parts. The first part, the MPC regulator, is responsible for finding the best control profile given steady-state targets for the states and inputs. In some formulations, the regulator may instead be used to track a dynamic output trajectory. The second part, the state estimator, determines an approximate current state of the system, knowing the history of injected inputs and measured outputs. The state estimator is also used to estimate the integrating disturbance state. The final part is the steady-state target calculation, which adjusts the state and input targets to account for the integrated disturbance.

We begin by introducing the discrete time model

$$x_{k+1} = F(x_k, u_k + X_u p_k, w_k, t_k) \tag{1}$$

in which the index k represents the current sampling time, x_k is the state of the system, u_k is the input, w_k is the stochastic noise variable, and t_k is the time. We assume that w_k is normally distributed and has a zero mean. In our formulation, we enforce a zero-order hold (constant value between sampling times) on the u_k and w_k. The term $X_u p_k$ is the integrated input disturbance. In cases of plant/model mismatch or nonzero mean disturbances, this term is nonzero; however, in the nominal case in which the plant and model are identical, this term vanishes.

The discrete time model can be identified directly, or, more often, is the result of integrating a first-principles DAE model. In linear MPC, the function F is linear with respect to its arguments. In our study, however, F is any twice continuously differentiable function.

The outputs of the system are modeled as

$$y_k = g(x_k, t_k) + X_y p_k + v_k \tag{2}$$

in which y_k is the measurement at time k and v_k is a stochastic Gaussian zero-mean noise term. Again, g is assumed twice continuously differentiable. The term $X_y p_k$ is the integrating output disturbance. The integrating disturbance evolves as

$$p_{k+1} = p_k + \xi_k \tag{3}$$

in which ξ_k is a normally distributed zero-mean vector.

2.1 Regulation

To solve the regulation problem, we first assume that the stochastic variables w_k, v_k, and ξ_k take on their mean values. Since these means are zero, we note by (3) that the forecast of the integrating disturbance term p_k is constant at each sampling time. Suppose that the system is currently at time j.

Consider the following formulation of this N-step finite-horizon MPC problem:

$$\min_{x,u} \Gamma(\tilde{x}_{j+N}) + \sum_{k=j}^{j+N-1} \mathcal{L}(\tilde{x}_k, \tilde{u}_k) \tag{4a}$$

subject to:

$$x_{k+1} = F(x_k, u_k + X_u p_k, 0, t_k) \tag{4b}$$

$$Du_k \leq d, \qquad Gx_k \leq g \tag{4c}$$

in which x_0 is the current state estimate and x and u denote the sequences of vectors representing states and inputs, respectively; that is,

$$\begin{aligned} x &= (x_{j+1}, x_{j+2}, \ldots, x_{j+N}), \\ u &= (u_j, u_{j+1}, \ldots, u_{j+N-1}). \end{aligned}$$

We define the deviation variables

$$\tilde{x}_k = x_k - x_{t,j}, \qquad \tilde{u}_k = u_k - u_{t,j} \tag{5}$$

in which the values $x_{t,j}$ and $u_{t,j}$ are the steady-state targets for the states and inputs at time j. The inequality constraints on the inputs and states may not yield a feasible solution for all problems, and for this reason, we may consider soft constraints on the states; see [23] for more details.

The stage cost that we consider here is

$$\mathcal{L}(\tilde{x}_k, \tilde{u}_k) = \tilde{x}_k^T Q \tilde{x}_k + \tilde{u}_k^T R \tilde{u}_k + (\tilde{u}_k - \tilde{u}_{k-1})^T S (\tilde{u}_k - \tilde{u}_{k-1}) \tag{6}$$

in which Q is positive semidefinite, and one of R and S is positive definite and the other is positive semidefinite.

We use the terminal penalty

$$\Gamma(\tilde{x}_{j+N}) = \tilde{x}_{j+N}^T P \tilde{x}_{j+N} \tag{7}$$

in which P is the solution to the discrete-time linear quadratic regulator problem for the linearized system at the current state and input targets [4]. For the case in which S is nonzero in (6), the terminal penalty is on the augmented state $[\tilde{x}_{j+N}; \tilde{u}_{j+N-1}]$.

The terminal state x_{j+N} is usually required to be inside some invariant region of the state target [8, 19]. Enforcing such a constraint is unnatural and impractical, however, as it is difficult to determine how long a horizon length N is required such that the terminal point may reach this set [6]. In practice, the horizon length is often chosen to be suitably long such that the terminal state x_N lies within the invariant region without explicitly constraining the state [25].

It is straightforward to test whether x_N satisfies the required properties of the invariant region; namely that a linear control law stabilizes the system after time N. We perform the following check:

Define

$$\tilde{u}_{j+N} = K \tilde{x}_{j+N}. \tag{8}$$

in which K is the optimal gain associated with the penalty P in (7). Provided $\tilde{u}_{j+N}$ does not lie on an active constraint, we calculate $\tilde{x}_{j+N+1}$ using (4b). If this new state does not violate any state constraints, we calculate the terminal stage cost decrease ratio

$$\alpha = \frac{\Gamma(\tilde{x}_{j+N}) - \Gamma(\tilde{x}_{j+N+1})}{\mathcal{L}(\tilde{x}_{j+N}, \tilde{u}_{j+N})}. \tag{9}$$

For linear systems, α will be exactly one. For nonlinear systems, any positive α value is acceptable. In cases in which this ratio is negative, or when constraints are active on the final inputs or states, the horizon length is not long enough. By satisfying the positivity of α, we are guaranteed nominal closed-loop stability.

The value of x_j comes from the state estimator. The first input u_j of the optimal input trajectory that results from the optimization problem (4) is injected into the plant.

2.2 Estimation

The control system obtains measurements from the plant at each sampling time. The goal of the state estimator is to determine the optimal approximation to the state evolution based on current and past inputs and measurements. The extended Kalman filter (EKF) is a popular industrial choice for nonlinear models. Moving horizon estimation (MHE) has emerged as an alternative to the EKF [18, 8]. In recent work, moving horizon estimation has been shown to possess superior estimation properties compared to the EKF [24]. We consider the moving horizon estimation problem at time index j of

$$\min_{\rho,\rho_p,x,p,w,v,\xi} \quad \Gamma_e(\rho,\rho_p) + \sum_{k=j-N_e}^{j} \mathcal{L}_e(w_k, v_k, \xi_k) \tag{10a}$$

subject to:

$$x_{j-N_e} = \bar{x}_{j-N_e} + \rho \tag{10b}$$
$$p_{j-N_e} = \bar{p}_{j-N_e} + \rho_p \tag{10c}$$
$$x_{k+1} = F(x_k, u_k + X_u p_k, w_k, t_k) \tag{10d}$$
$$p_{k+1} = p_k + \xi_k \tag{10e}$$
$$y_k = g(x_k, t_k) + X_y p_k + v_k \tag{10f}$$
$$Hx_k \le h, \qquad Sw_k \le s, \qquad \Gamma v_k \le \gamma \tag{10g}$$

in which the current output is denoted as y_j. The *a priori* estimate $\bar{x}_{j-N_e}$ is given in this formulation. The current state estimate x_j is the desired result that is used as the given initial state in the regulator. Also, the current integrated disturbance p_j is estimated and used in the target calculation.

The estimator stage cost is defined as

$$\mathcal{L}_e(w_k, v_k, \xi_k) = w_k^T Q_w w_k + v_k^T R_v v_k + \xi_k^T Q_\xi \xi_k \tag{11}$$

in which the penalty matrices Q_w, R_v, and Q_ξ are the inverse of the covariances of w, v, and ξ, respectively.

A few strengths of the MHE formulation are that constraints are incorporated into the framework, the resulting estimates are optimal and the method possesses excellent stability properties [17]. In the full-information problem, the horizon length N_e increases at each sampling time, tending toward infinity. In MHE, however, N_e grows to a specified horizon length N_T, at which point $\bar{x}_0$ and $\Gamma_e(\rho,\rho_p)$ are updated to reflect previous estimates of x_1. By selecting the arrival cost penalties appropriately, MHE approximates the full-information problem.

The arrival cost is approximated as

$$\Gamma_e(\rho, \rho_p) = \begin{bmatrix} \rho \\ \rho_p \end{bmatrix}^T P_{e,k} \begin{bmatrix} \rho \\ \rho_p \end{bmatrix} + p_{e,k}^T \begin{bmatrix} \rho \\ \rho_p \end{bmatrix} \tag{12}$$

in which the penalty matrices $P_{e,k}$ and $p_{e,k}$ and the *a priori* estimate $\bar{x}_{j-N_e}$ are updated according to the nonlinear smoothing covariance update given in [24], which is based on approximating the nonlinear system as a linear time-varying system. The smoothing update is based on the corresponding work of Rao et al. [16] for linear time invariant systems.

The initial values for the penalty matrices are

$$P_{e,0} = \begin{bmatrix} \Pi_0 & 0 \\ 0 & \Pi_{p,0} \end{bmatrix}, \qquad p_{e,0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{13}$$

in which Π_0 represents the inverse of the initial covariance of the *a priori* state estimate and $\Pi_{p,0}$ is the inverse of the initial covariance of the *a priori* disturbance estimate. These matrices are not changed until the estimation horizon length reaches N_T, at which point the updating strategy is initialized.

2.3 Target Calculation

The goal of the target calculation is to find a steady state of the model that yields an output at the set point unless no such steady state exists. In circumstances where no steady-state targets can be found corresponding to the set point, we require the output target to be the closest output to the set point for which a steady state exists. If there are multiple steady-state inputs and states that satisfy the set point condition, then the state and input targets are selected as those that are nearest the previous input targets. At each time instant j, a new target must be calculated to account for the integrated disturbnce p_j. We formulate these notions as the following optimization problem, based on the linear case of Rao and Rawlings [15]:

$$\min_{x_{t,j}, u_{t,j}, \eta} \frac{1}{2}\eta^T \bar{Q}\eta + \bar{q}^T\eta + \frac{1}{2}\left(u_{t,j} - u_{t,j-1}\right)^T \bar{R}\left(u_{t,j} - u_{t,j-1}\right) \tag{14a}$$

subject to:

$$x_{t,j} = F(x_{t,j}, u_{t,j} + X_u p_j, 0, t_j) \tag{14b}$$

$$g(x_{t,j}, t_j) + X_y p_j - \eta \le y_{\text{set}} \le g(x_{t,j}, t_j) + X_y p_j + \eta \tag{14c}$$

$$Du_{t,j} \le d, \qquad Hx_{t,j} \le h, \qquad \eta \ge 0. \tag{14d}$$

In this nonlinear program, the variable η is a relaxation of the requirement that the state and input targets send the output to the set point when the set point is

not feasible. Infeasibility of the set point occurs due to constraints or the dynamics of the nonlinear system. The vector $u_{t,j-1}$ is the input target from the previous sampling time. In general, $\bar{q}$ is chosen to be relatively large and strictly positive, and both $\bar{Q}$ and $\bar{R}$ are positive definite.

By shifting the state and input targets, the target calculation accounts for modelling error and adjusts the model to remove offset from the closed-loop system. In nonlinear MPC, it is possible that (14) does not have a solution. This problem may occur due to large disturbances or difficult constraints, such as tight bounds on the inputs. Infeasible target problems can also arise from large mismatch between plant and model. While relatively rare, failure of the target calculation is a serious issue that requires further study.

3 Disturbance Models

Integrating disturbances, as described earlier, are useful for removing offset in the output caused by unmodeled nonzero mean disturbances and plant/model mismatch. Provided the appropriate number of integrating disturbances have been added, the addition of integrating disturbances either removes offset or does not allow the closed-loop system to reach a stable operating point (see [12, 11] for details on the linear MPC case).

The effects of adding integrating disturbances are specific to the application, and determining where to add the integrating disturbance term for nonlinear models is an open question. The distribution of the integrating disturbance term among the inputs and outputs in nonlinear model predictive control is important since it may determine whether the closed-loop system reaches its target.

For cases of plant/model mismatch in linear MPC, Muske and Badgwell [11] recommend the input disturbance model. This choice is also justified by Shinskey [20], who notes that pure output disturbances are unlikely to occur in industry, but rather enter a system upstream of a dominant time constant, and usually at the manipulated variables. Morari and Lee [10] further note that DMC, which uses output disturbance models, does not quickly reject slow disturbances. Despite the preponderance of literature advocating input disturbance models for linear MPC, a study of the steady-state behavior of the plant is necessary before choosing a disturbance model for nonlinear MPC. We begin by examining the effects of disturbance models on two different systems.

Example 3.1 *Consider a solar collector plant in which a fluid is used to absorb focussed solar energy along 790 meters of specially designed pipe [22, 21]. The fluid is circulated to a heat exchanger, where the energy is removed, and the cooled fluid is recirculated 790 meters back to the collector. The process is regulated by a pump that*

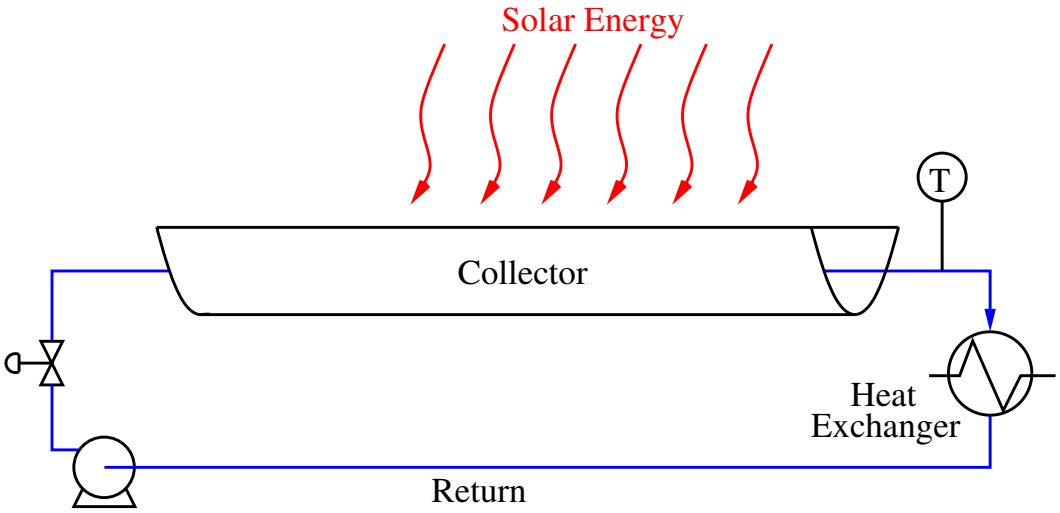

Figure 1: Schematic of solar collector

controls the flow rate of fluid through the system. A schematic of the solar collector is presented in Figure 1. Since the system is constructed to harness the sun's energy, the outlet temperature after the collector must be relatively high for the process to operate properly. In the case of this example, the desired outlet temperature is 543 K. Cloud cover, changes in ambient temperature, and other local weather effects act as unmeasured disturbances to the system, and must be rejected.

The system is modeled by discretizing the collector and recycle loop spatially along their respective lengths. For ease of modeling, the temperatures at these spatial discretization points are states in the model, and are located close enough to each other physically that to a good approximation

$$\frac{\partial T}{\partial z} \approx \frac{T_i - T_{i-1}}{\Delta z}$$

in which T is the temperature at location z at which state i is measured. In this example, the temperatures along the length of the collector are modeled as the first twenty states, the next state models the temperature in the exchanger, and the final twenty states represent the temperature along the length of the recycle tube. The governing equations for this system are

$$\dot{T}_i = \alpha_i(F)(T_{i-1} - T_i) + \beta_i(T_{amb,i} - T_i) + \gamma_i \tag{15}$$

for $i = 1, 2, \ldots, 41$. For the case of $i = 1$, $T_{i-1} = T_{41}$. In this equation, the $\alpha_i(F)$ term represents the energy entering location i of the system from the previous

Parameter	Collector $(1 \leq i \leq 20)$	Exchanger $(i = 21)$	Recycle $(22 \leq i \leq 41)$
$\alpha_i(F)$ (s^{-1})	$(8.22 \times 10^{-3} kg^{-1})$ F	1.0	$(8.22 \times 10^{-3} kg^{-1})$ F
β_i (s^{-1})	1.19×10^{-3}	-5.0	0
γ_i (K/s)	5.41×10^{-1}	0	0
$T_{amb,i}$ (K)	303.15	375.15	———

Table 1: Parameters for the solar collector model for Example 3.1.

location and leaves the current location by mass transport. In most cases, this term will depend on the mass flowrate F; however in the heat exchanger, we assume the exchange is fast enough that it does not depend on the flowrate of the circulating fluid. The term with β_i governs the energy exchanged between the system and the surroundings, such as the outside air for the collector, or the water in the heat exchanger. The γ_i term is the energy that enters the system from solar radiation. The values of the parameters for this model are listed in Table 1. Note that the input parameter F is included in $\alpha_i(F)$ for all states but the heat exchanger. Therefore, the $\alpha_i(F)$ terms are the source of nonlinearity in the system, since it multiplies the input with the states. Without disturbances, the steady-state input for the plant is 5.7 kg/s.

We examine the solar collector in Example 3.1 during an unmeasured 40% decrease in the solar power input (γ_i). Since this disturbance is unmodeled, an integrating disturbance model is added. We investigate the performance of the pure input disturbance model ($X_u = 0.02$, $X_y = 0$) and pure output disturbance model ($X_u = 0$, $X_y = 1$) cases. We have chosen the relative scaling of X_u and X_y for input and output disturbance models such that the magnitudes of the integrating disturbances for both cases are similar. By maintaining the same penalties in the estimator, this disturbance tuning yields a fair comparison between disturbance models. The system has the following tunings for the regulator:

$$Q = I_{41}, \qquad R = 0, \qquad S = 1 \times 10^{-6}.$$

The input F is constrained to be between 0.8 and 8 kg/s. The system is sampled once per minute, and the prediction horizon is ten minutes. The estimator matrices are

$$Q_w = 0, \qquad R_v = 1 \times 10^6, \qquad \Pi_0 = (1 \times 10^6) I_{41},$$
$$Q_\xi = 1, \qquad \Pi_{p,0} = 1.$$

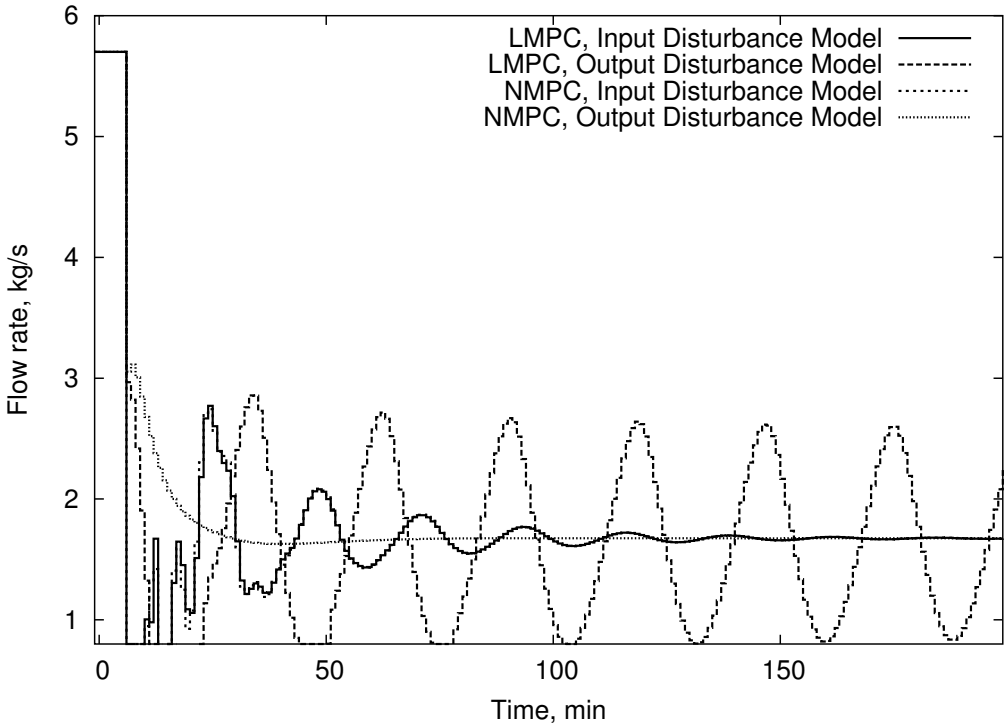

Figure 2: Disturbance model input performance for solar collector in Example 3.1

The system has forty-one states, and the output is the temperature at the outlet of the collector, which is the twentieth state.

The disturbance to the system occurs after five minutes, and since the solar input is unmeasured, the model and process no longer agree. The inputs and outputs of the system for the input and output disturbance models for both linear and nonlinear MPC are shown in Figures 2 and 3. The model for linear MPC is obtained by linearizing the model about the original set point. We choose this linearization point because the process is originally stable at this point and the disturbance is unmeasured; it would be impossible to linearize the system at the final target since it is not known in advance.

It is evident from Figure 3 that the input disturbance model handles the sudden decrease in solar power more effectively than the output disturbance model for both linear and nonlinear models. In fact, the output disturbance model for the linear model does not steer the system to the set point within the observed time frame. The input disturbance models for linear and nonlinear MPC are nearly identical for this example. One reason the input disturbance model is better is that the modelled

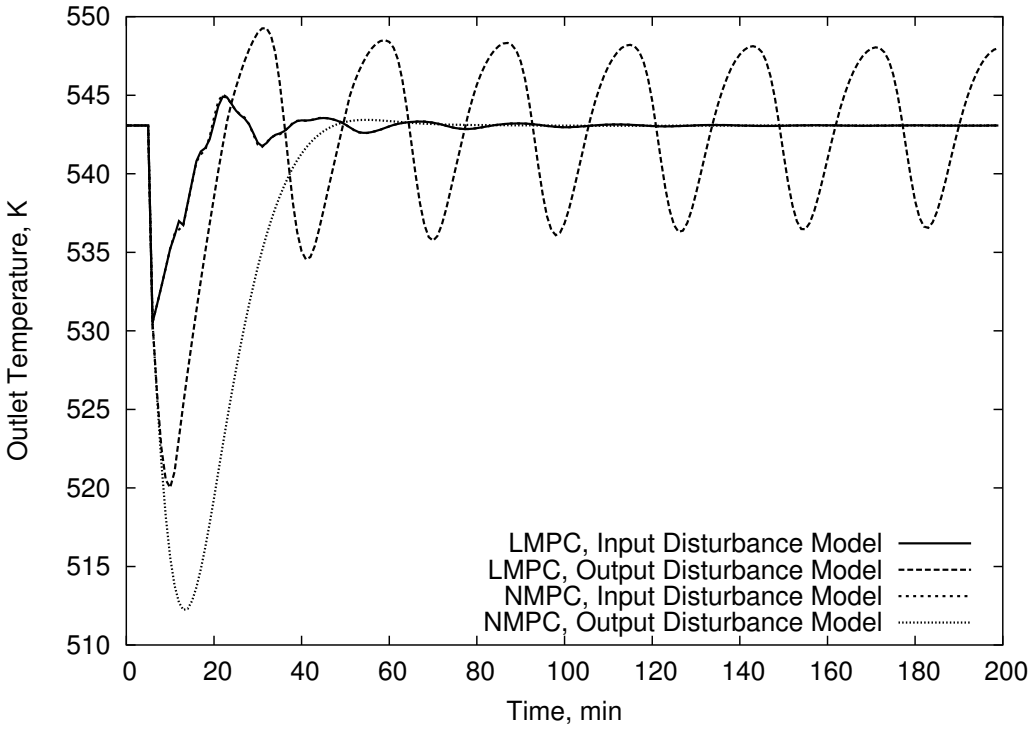

Figure 3: Disturbance model output performance for solar collector in Example 3.1

F	100 L/min	C_{Af}	1 mol/L
V	100 L	E_1/R	8750 K
k_1	7.2×10^{10} min^{-1}	E_2/R (plant)	9700 K
k_2	5.2×10^{10} min^{-1}	E_2/R (model)	9750 K

Table 2: Parameters for the CSTR Model for Example 3.2.

disturbance directly affects the states of the system, just as the true disturbance does. The output disturbance model, on the other hand, does not directly impact the states. Rather, it tries to find a new input and state target such that the original dynamics yield a system that accounts for the difference in observed and predicted outputs. In general, the input disturbance model is expected to perform better on systems with disturbances that alter the evolution of the states. We could conclude from this study that the input disturbance model is a better choice for mismodeled systems. However, such a conclusion would be premature, as we show in the following example.

Example 3.2 *Consider a continuously stirred tank reactor (CSTR) in which the irreversible reactions $A \rightarrow B \rightarrow C$ are taking place. The feed stream to the reactor is pure species A, and the maximum conversion to product B is desired. The concentration of the product B is measured and the process is regulated by adjusting the temperature of the reactor directly by a cascaded control system. The system is governed by the equations*

$$\dot{C}_A = \frac{F}{V}(C_{Af} - C_A) - k_1 C_A e^{-E_1/RT} \tag{16}$$

$$\dot{C}_B = k_1 C_A e^{-E_1/RT} - k_2 C_B e^{-E_2/RT} - \frac{F}{V} C_B \tag{17}$$

in which the concentrations C_A and C_B are the state variables and the temperature T is the manipulated variable. The values of the plant and model parameters are listed in Table 2.

In Example 3.2, the plant and the model do not agree since the activation energy in the plant is slightly lower than its model counterpart. This mismatch is realistic due to the difficulty in identifying accurate activation energies from experiments. In fact, the modeling error is relatively small; Figure 4 shows the locus of steady states for the model and the plant. These curves are similar in nature, with the exception that the maximum yield predicted by the model is higher than that of the plant. The model yield of species B is 0.670 mol/L, while the plant can achieve only 0.654 mol/L.

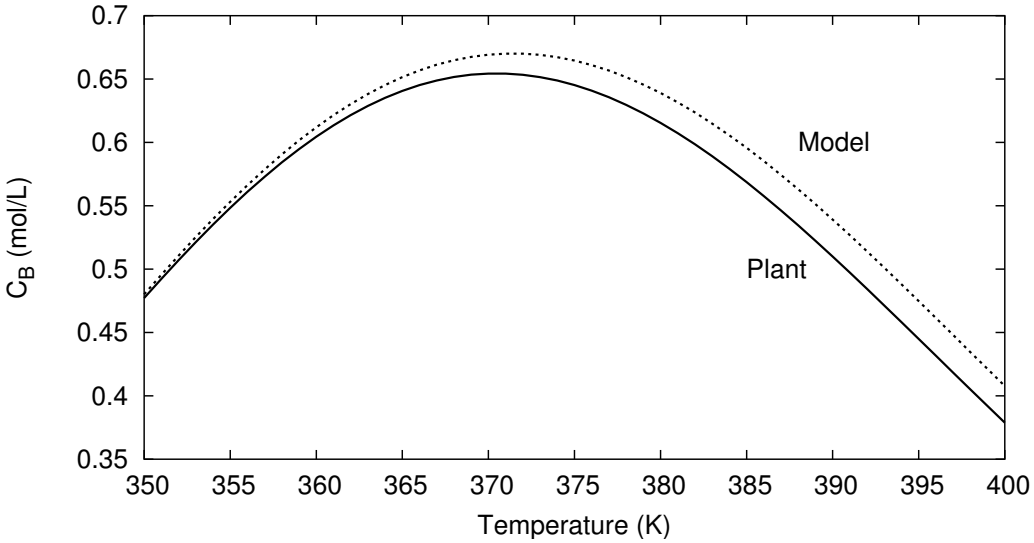

Figure 4: Locus of steady states of CSTR in Example 3.2

We now investigate the consequences of this small modelling discrepancy on closed-loop performance. We wish to operate the plant at its point of maximum yield. Therefore, the output set point in the target calculation is defined as the maximum yield of the model. Consider an MPC regulator with the penalty matrices

$$Q = \begin{bmatrix} 0 & 0 \\ 0 & 400 \end{bmatrix}, \qquad R = 2, \qquad S = 0.$$

a sampling time of 0.05 minutes, and a prediction horizon of 3 minutes. The estimator is tuned using

$$Q_w = 0, \qquad R_v = 1 \times 10^6, \qquad \Pi_0 = \begin{bmatrix} 1 \times 10^6 & 0 \\ 0 & 1 \times 10^6 \end{bmatrix},$$
$$Q_\xi = 1, \qquad \Pi_{p,0} = 1 \times 10^6,$$

and we begin by deciding what values of X_u and X_y to use. In the case of a pure input disturbance model, $X_u = 1$ and $X_y = 0$. For the output disturbance model $X_u = 0$ and $X_y = 1$. The closed-loop simulations for both cases are shown in Figures 5 and 6 for linear MPC and Figures 7 and 8 for nonlinear MPC. In these figures, the dashed lines represent the target values, while the solid lines represent the actual inputs and outputs of the plant. The linear model is obtained by linearizing the model at the original desired state and input targets.

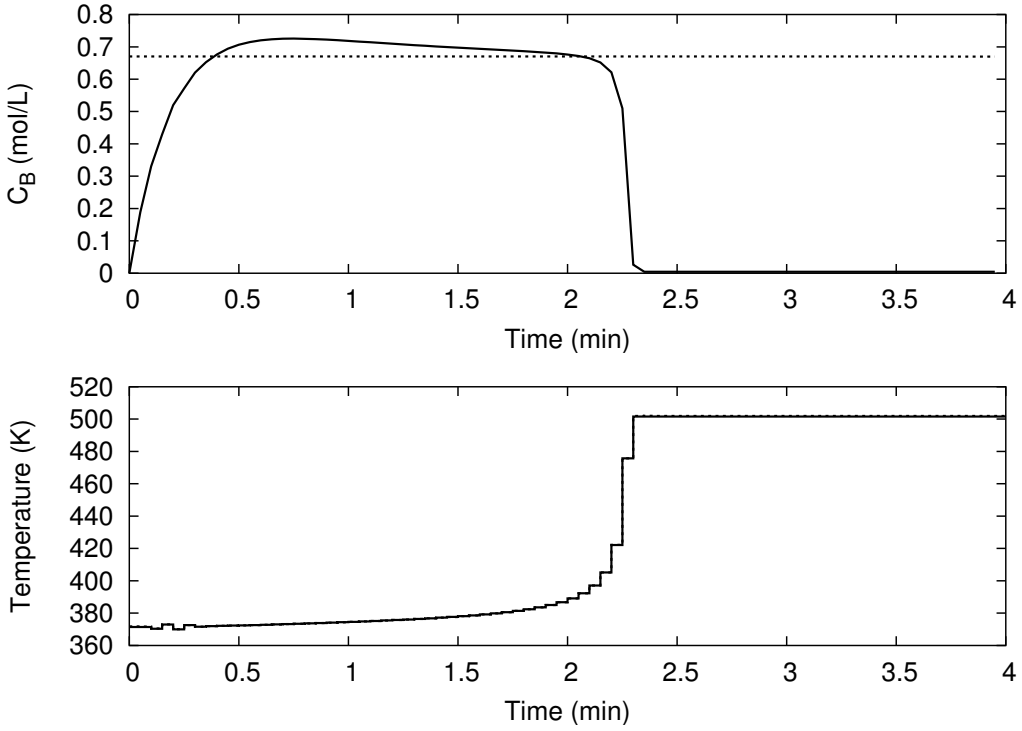

Figure 5: Closed-loop performance of input disturbance model on CSTR in Example 3.2 under linear MPC

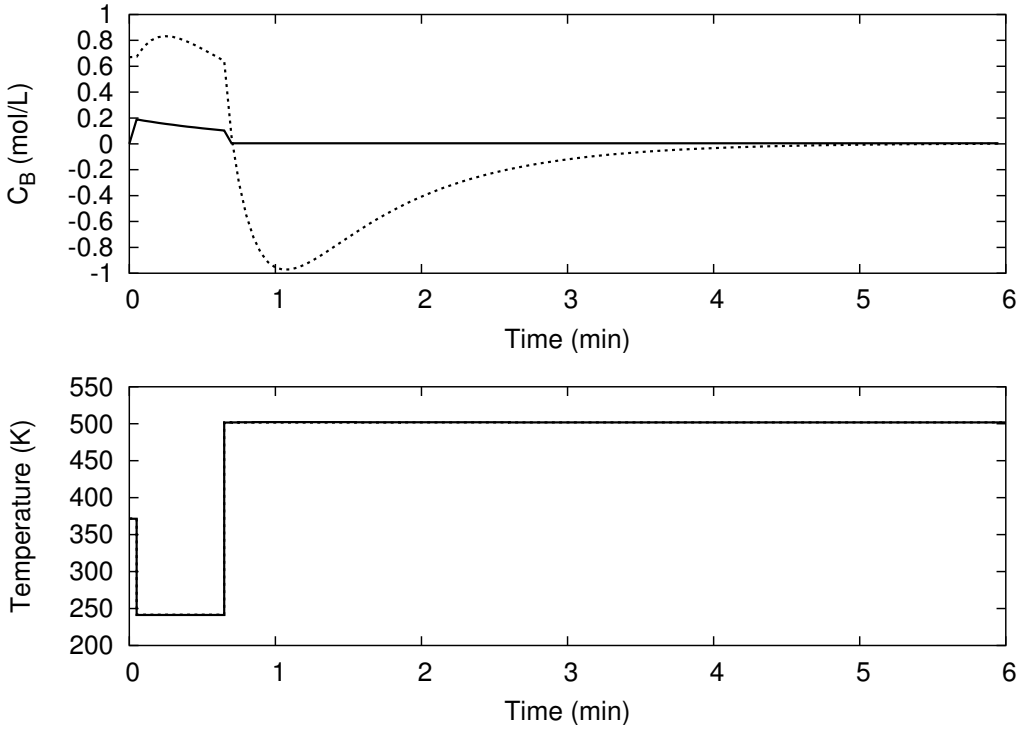

Figure 6: Closed-loop performance of output disturbance model on CSTR in Example 3.2 under linear MPC

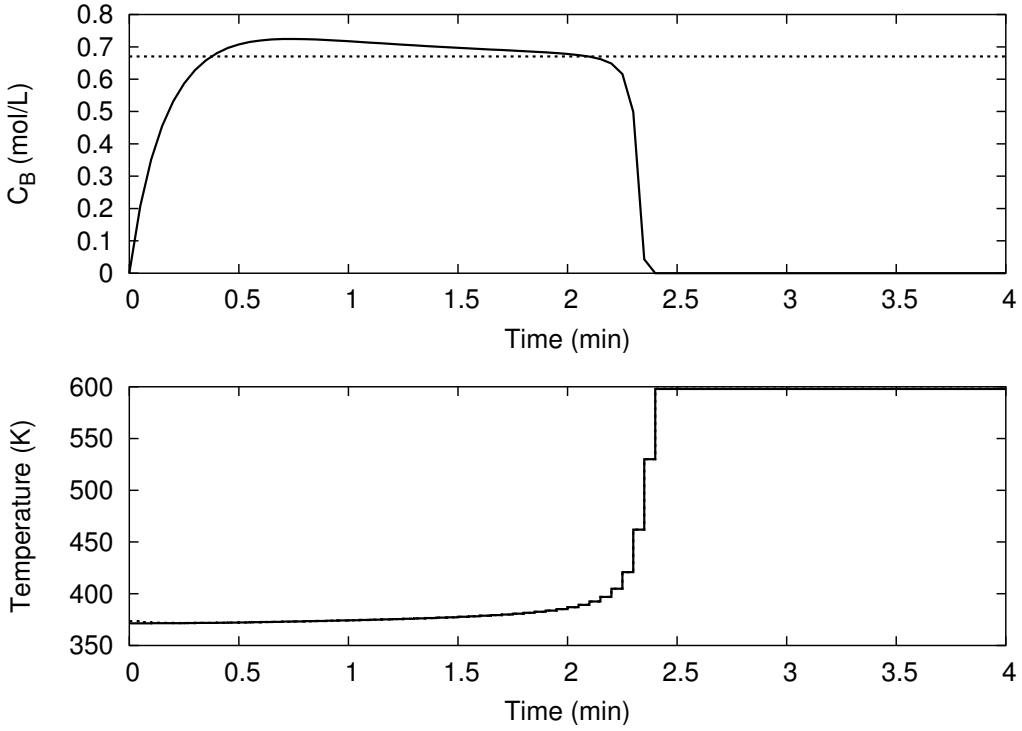

Figure 7: Closed-loop performance of input disturbance model on CSTR in Example 3.2 under nonlinear MPC

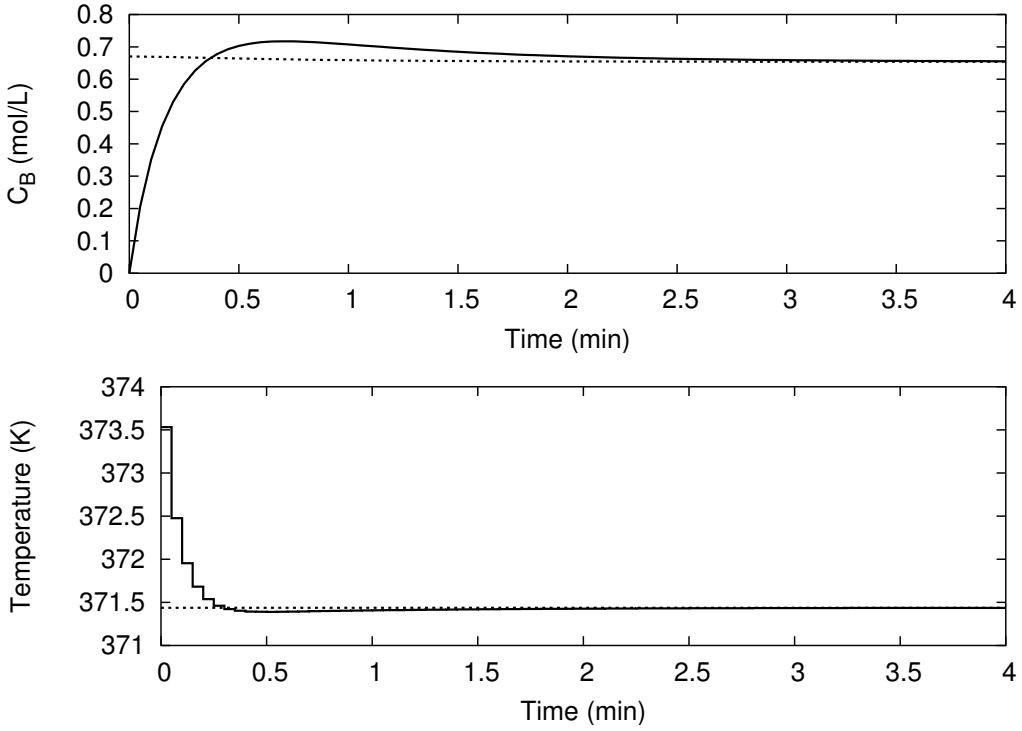

Figure 8: Closed-loop performance of output disturbance model on CSTR in Example 3.2 under nonlinear MPC

The performance of linear MPC on this example is poor. The closed-loop system with the input disturbance model converts the reactant completely to the unwanted product C. The output disturbance model cannot determine an output target that recovers any appreciable yield of product B. Linear MPC is inappropriate for this system.

From examining Figures 7 and 8, it is clear that the output disturbance model in nonlinear MPC performs quite well on this example, bringing the closed-loop system to a steady-state value. The yield of the product B is not the maximum yield of the plant, but it is close. The output target has shifted down from its original unreachable value to a lower value that can be attained by the plant.

In the case of the input disturbance model, as with linear MPC, the system cannot be stabilized. While the input tracks the input target almost exactly, the target itself diverges. As a result, the reactor converts the reactant A completely to the undesired product C by increasing the temperature.

In this example, it is evident that an input disturbance model is not sufficient to control the system with plant/model mismatch. The effects of disturbance models on the MPC system must be explored to explain the cause of this phenomenon. Figure 9 is a graphical representation of the steady-state effects of adding an integrating disturbance to a system. The curves in the figure are the loci of steady states for the model and the plant. In this example, our aim is to maximize the output, so the target always resides at the maximum of the model curve. The input disturbance model, since it adds a bias term to the input, shifts the steady-state curve of the model horizontally in the figure. Similarly, the output disturbance model may adjust the model curve only in the vertical direction. For a closed-loop steady state to occur, the model curve, the target, and the plant curve must all meet at one point. When the set point and the model do not cross, the output target shifts to the steady state of the model nearest the set point. In this example, the desired steady state is at the maximum point of the model, so the target and model move together, intersecting only at this point. It is obvious that this intersection point cannot lie on the plant curve solely by moving horizontally, which explains why the input disturbance model cannot stabilize Example 3.2. On the other hand, when the model curve is allowed to move down to intersect the plant at the maximum of the model, as happens with the output disturbance model, the system can reach a steady state. Finally, we note that the value of the steady state reached by using a pure output disturbance model is predicted by finding the output value of the plant at which the input value of the model attains its maximum output.

The ill effects of using an input disturbance model in this example are due to two factors. First, the plant has a maximum value for the steady-state output, and second, when an infeasible set point is specified, the target calculation is incapable of finding an attainable output target. A possible solution to this problem, while still

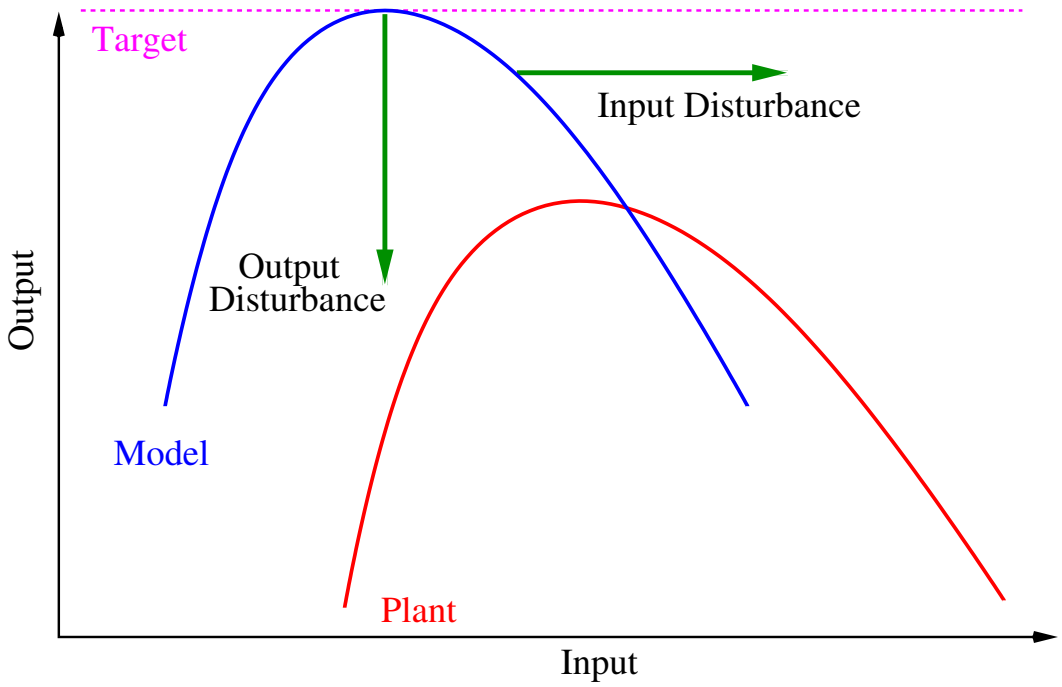

Figure 9: Effects of disturbance models on model steady states

using an input disturbance model, is to not specify that the process operate at the maximum value, but rather choose an appropriately conservative operating point that is known to be a steady state. The location of such an operating point may not be available *a priori*, and is a limiting weakness of the input disturbance model. In general, the steady-state behavior as shown in Figure 9 dictates whether closed-loop stability is possible; whether the closed-loop system is actually stabilized, however, is determined by the combined dynamics of the estimator, regulator, and target calculation.

4 Local Minima

Because the model equations are nonlinear, the equality constraints in the optimization problems of the estimator, regulator, and target calculation may be noncovex, yielding local minima that are not globally optimal. This phenomenon is not observed with linear models because the linear constraints that represent the process dynamics are always convex. The local optima of the nonlinear programs may be undesirable due to physical considerations, or they may be unacceptable from a performance standpoint. Finding the appropriate locally optimal solution can often be achieved through appropriate initial guesses to the optimizer and the use of realistic constraints. We begin by discussing how local optima arise in the NMPC framework.

In related work, we describe the appearance of local minima for an exothermic stirred tank reactor [23]. In that example, the local minima arise from varying the initial guesses to the optimization algorithm by adding noise to the initial guess for the inputs. At least seven local minima exist in the regulator optimization problem for the simple reactor example, all of which asymptotically reach the state and input targets. However, all but the globally optimal solution cause the process to ignite, converting nearly all the reactant to product at a high temperature.

Local minima are possible in the regulator, estimator, and target calculation problems when using nonlinear process models. The different solutions that result from a nonlinear problem are a function of the initial guess to the optimization algorithm. For NMPC strategies that rely on solutions from previous sample times to initialize the optimization problems, local solutions can appear in the regulator when the current state estimates are not close enough to the predicted states from the previous time step. This disagreement between the open-loop prediction of the regulator and the state estimate based on plant data can result from plant/model mismatch, inaccurate state estimates, unmeasured disturbances, or even system noise. In general, an educated guess for the initial starting point may often yield the global solution to the nonlinear program, but there are no guarantees. Occasionally,

additional steps can be taken to prevent convergence to an undesired local solution. We present the following example to illustrate how local minima arise and how to avoid such solutions in a nonlinear MPC regulator.

Example 4.1 *An electromagnetically actuated mass spring damper system is governed by the equations*

$$\begin{aligned} \dot{p} &= v \\ \dot{v} &= -\frac{k}{m}p - \frac{c}{m}v + \frac{\alpha}{m}\frac{C}{(d_0 - p)^\gamma} \end{aligned}$$

where the states p and v represent position and velocity, respectively, and the input C is a function of the current applied to the coil (see [9]). The physical parameters of this model are: $\alpha = 4.5 \times 10^{-5}$, $\gamma = 1.99$, $c = 0.6590$, $k = 38.94$, $d_0 = 0.0102$, and $m = 1.54$. We apply the constraint $0 \leq C \leq 3$ to the input.

The goal is to steer the system such that it is at rest ($v = 0$) at the position $p = .0074$. The corresponding steady-state input for this position is $C = .0532$. The initial state of the system is $p_0 = 0$, $v_0 = .012$.

In Example 4.1, we use the initial guess to the optimizer as described in [23]. The regulator has a sampling time of .01 and the following cost function weighting matrices:

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad R = 1, \quad S = 0.$$

We begin by examining the case in which the prediction horizon for the regulator is 200 time steps, or 2 time units. The initial guess to the optimizer and the resulting open-loop optimal control profile is shown in Figure 10, and the evolution of the position state is displayed in Figure 11. The locally optimal input profile oscillates once before settling; we show later that this sequence of inputs is not globally optimal. Even though the initial guess is relatively good, the global solution to this regulation problem cannot be found as currently formulated.

We now concentrate on ways to force the system to the globally optimal profiles. In the same figures, we present the effects of using a shorter prediction horizon of 100 time steps. The periodic behavior of the input profile is not present with the shorter horizon because the horizon length is not long enough to complete a full period. This result runs counter to the commonly held belief that increasing the prediction horizon length can only improve controller performance. Here, a long prediction horizon can lead to one – or several – periods before approaching the set point. Because we use a terminal penalty to approximate the infinite horizon cost to go, decreasing the prediction horizon does not prevent the controller from having an accurate forecast of the infinite horizon cost. The shorter prediction horizon does

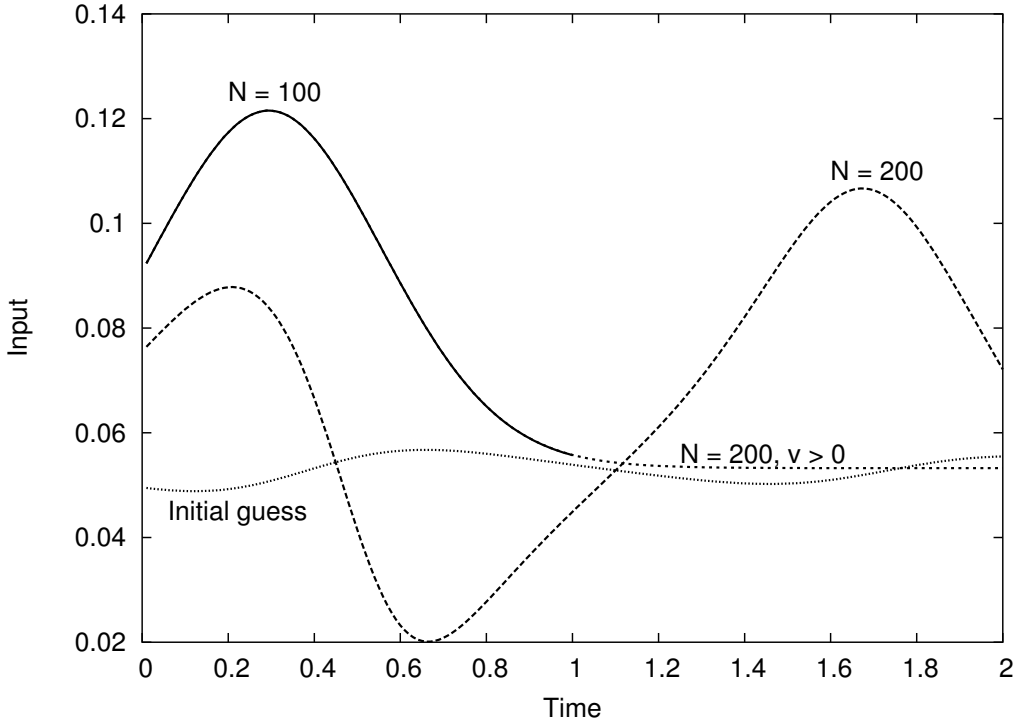

Figure 10: Optimal input profiles for spring system in Example 4.1

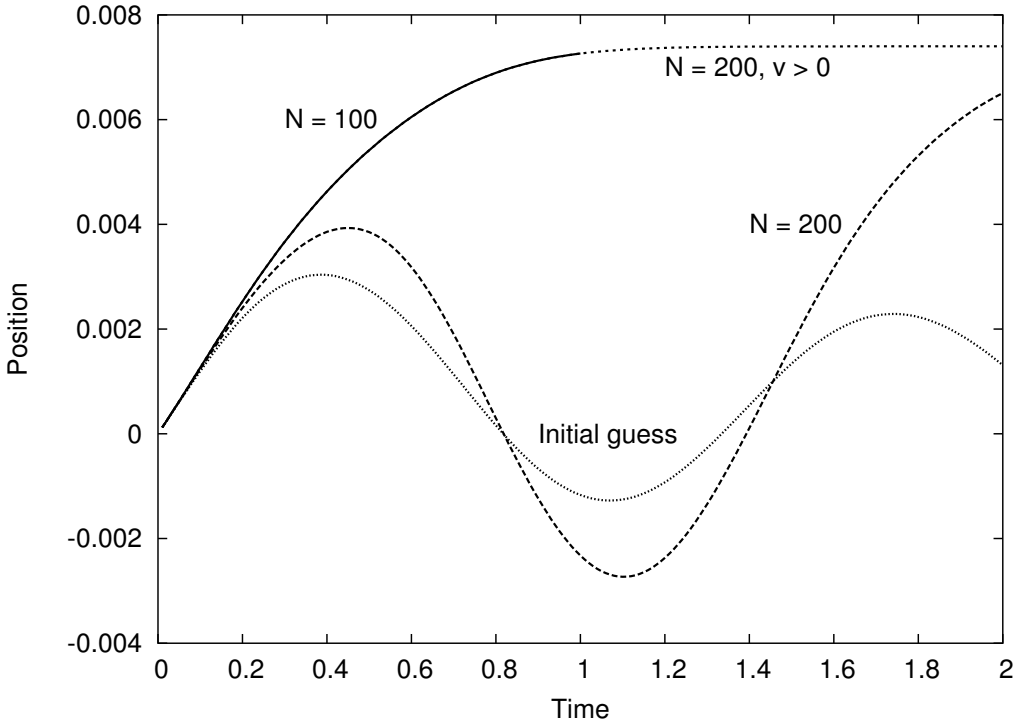

Figure 11: Optimal position profiles for spring system in Example 4.1

prevent the controller from approaching the steady state by a path with unnecessary periodic behavior.

We also present the results of using the original horizon length of 200 time steps but constraining the velocity of the system to be nonnegative. These results are also in Figures 10 and 11. Note that this case also removes the periodic result and that the additional constraint, although in the formulation, is not active at the optimal solution. Even though the constraint is present in the formulation, the resulting solution is unconstrained. This behavior is a key difference between the use of linear and nonlinear models. In linear MPC, if the control trajectory changes when a constraint is added, then that constraint must be active at the solution. However, constraints in the nonlinear MPC problem may change the optimal solution, but need not be active at the solution if a new local minimum to the cost function is found. This behavior is one of the main differences between linear and nonlinear MPC.

The value of the cost function for the locally optimal solution with a horizon length of 200 time steps is 0.21985, while the cost function values for both the shorter horizon solution and the constrained result are 0.21463. Based solely on the value of the cost function, both these solutions are better than the solution found by the nominal formulation of N=200 without constraints. In the closed-loop sense, the locally optimal solutions are nearly equivalent since they both achieve the main goal of the controller – namely to stabilize the system at its set point. However, in cases in which the local optima take unwanted excursions, for instance via ignition, constraints should be placed on the regulator to prevent poor performance.

In practice, it may not be possible or practical to apply a constraint in the regulator. For instance, in the above example, the constraint is not added out of consideration for the true dynamics of the process, and if a disturbance were to occur, this constraint may prevent the controller from properly stabilizing the system. On the other hand, shortening the prediction horizon in the regulator is a generic solution that is applicable regardless of the process. For instance, the unwanted local minima in the CSTR system studied in [23] can be removed by sufficiently shortening the prediction horizon, while applying a constraint is not successful on all initial guesses. If a longer horizon is desired for performance reasons, then the result for the shorter horizon as described above can be extrapolated by a linear control law and used as an initial guess for the full horizon problem.

For the estimation problem, we revisit the batch gas reactor presented in [24].

Example 4.2 *An isothermal gas-phase reactor is charged with an initial amount of A and B, and the species are allowed to react according to the reversible reaction $2A \rightleftharpoons B$. The goal is to reconstruct the partial pressure of each species in the reactor during the reaction based on measurements of the total pressure of the vessel as the*

reaction proceeds. The system is modeled by

$$\dot{x}_1 = -2k_1x_1^2 + 2k_2x_2, \qquad \dot{x}_2 = k_1x_1^2 - k_2x_2 \tag{18}$$

in which $k_1 = 0.16\ min^{-1}\ atm^{-1}$, $k_2 = 0.0064\ min^{-1}$. *The measured output is the total pressure described by*

$$y = x_1 + x_2. \tag{19}$$

We study the estimation problem presented in Example 4.2 with the initial conditions

$$x_0 = \begin{bmatrix} P_A & P_B \end{bmatrix}^T = \begin{bmatrix} 3 & 1 \end{bmatrix}^T, \qquad \bar{x}_0 = \begin{bmatrix} 0.1 & 4.5 \end{bmatrix}^T.$$

The estimator is tuned as

$$Q_w = \begin{bmatrix} 1 \times 10^5 & 0 \\ 0 & 100 \end{bmatrix}, \qquad R_v = 100, \qquad \text{and} \qquad \Pi_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Note that we do not require tunings on the integrating disturbance – we are using a perfect model, so no integrators are added.

The sampling time of the system is six seconds and the estimation horizon is 20 sampling times. We examine the results of the state estimator after twenty output measurements are taken, i.e. at 114 seconds, in Figure 12. Without state constraints, the optimal estimate of the partial pressure of species B is negative. Applying the constraint that the partial pressures must each be positive yields a new optimal estimate for which the constraint is not active. This constrained estimate is clearly better, since the estimated states converge to the true states of the system. The cost function of the constrained estimator is 19.76, compared to the unconstrained value of 40.35. We present the output values associated with the constrained and unconstrained estimator in Figure 13. The constrained estimate fits the data well, while the unconstrained estimate fits the data in a way that does not adequately represent the true dynamics of the system. It is clear that enforcing physical constraints in the estimation problem plays a crucial role in yielding accurate state estimates.

The effects of undesirable local minima in state estimation can be more severe to the closed-loop behavior of an MPC system [2] when compared to local minima in the regulator. One reason the estimator is more important is that both the target calculation and regulator rely on the results of the state estimator for good performance, while locally optimal behavior in the regulator does not affect the state estimates or steady-state targets. The stabilizing properties of local minima in the regulator were mentioned by Chen and Allgöwer [3]. These properties were proven in the context of stabilizing suboptimal control by Scokaert et al. [19]. We

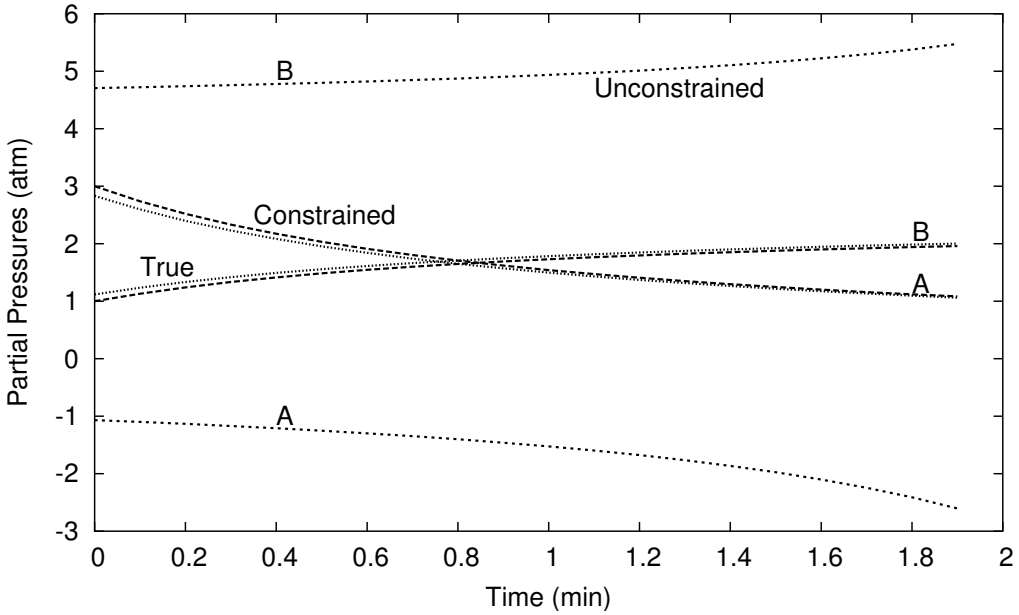

Figure 12: Optimal pressure estimates for Example 4.2

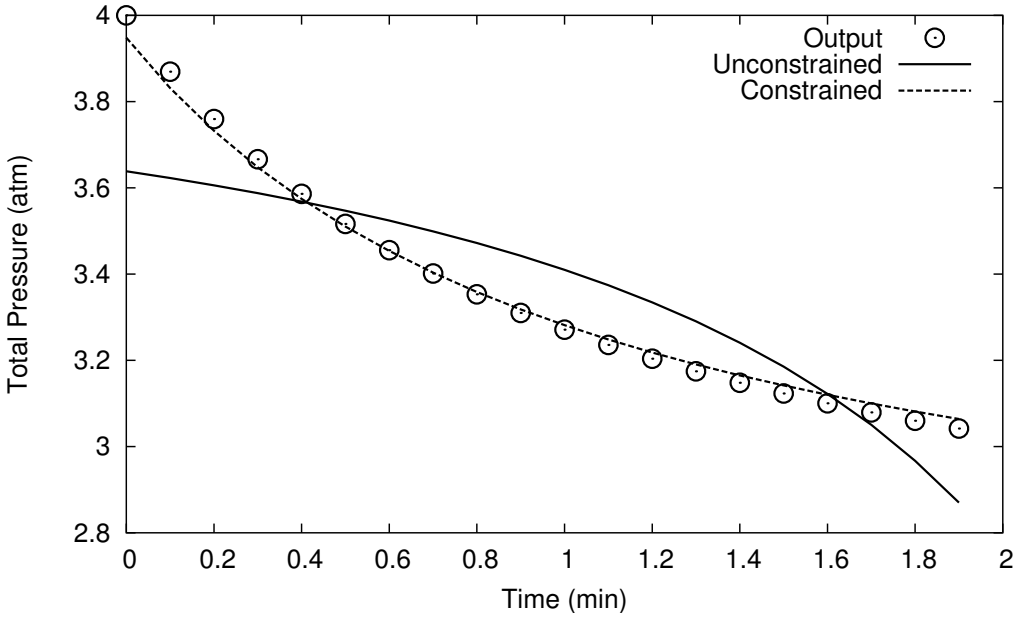

Figure 13: Fit of output estimates to measurements for Example 4.2

note that in the case of the regulator, the effect of local minima is to stabilize the system at set point, but to do so in an inefficient way. This behavior contrasts with the role of local minima in the estimation problem, since local minima do not necessarily converge to the true states of the system, as is evident in Example 4.2. It is possible that local minima in the estimator can lead to failure of the closed-loop system. We can conclude that local optima in the state estimation problem are an important concern for closed-loop operation when employing a nonlinear model. We show that through the application of physically motivated constraints, we can easily avoid these unwanted local minima for our examples.

5 Conclusions

Nonlinear models have significant advantages over linear models for control and estimation. At the same time, they also present new problems that must be addressed to guarantee their usefulness. This paper demonstrates both the features and caveats of nonlinear models in MPC, and presents simple and effective methods for analyzing and avoiding these difficulties.

In this paper, we show that the choice of disturbance models is an important issue for nonlinear plants. The choice of disturbance model can determine whether a system can be stabilized. In general, the input disturbance model is appropriate for cases of plant/model mismatch, although it is ineffective for a specific class of systems. We further demonstrate poor performance of linear MPC on this class of system. A future direction of research is to investigate a class of disturbance models in which the integrating term is added to model parameters. For instance, in the case of the solar collector, the disturbance is anticipated to come from the solar input, so adding the integrating disturbance to the solar input term is a natural choice.

We also investigate the occurence and role of local minima in the nonlinear regulator and estimator problems. While using linear models always results in convex optimization problems, regardless of horizon length, nonlinear models can introduce nonconvexities into these mathematical programs. In the case of the regulator, long prediction horizons can lead to locally optimal periodic solutions that can be removed by using a shorter prediction horizon. Also, constraints can be used to steer a system towards a more favorable local optimum. For the estimator, applying physical constraints to the system may avoid undesired local optima. Although in both cases, the constraints are not active at the solution, they aid the optimizer in finding worthwhile optimal solutions that may not be found otherwise. We recommend that even if a constraint is not expected to be encountered by the physical system, for instance negative absolute temperatures, it should be included in the controller and

estimator formulations to avoid unwanted local minima.

Acknowledgments

The authors thank Profs. Wiiliam A. Beckman and John W. Mitchell for introducing and describing the solar collector example, and Thorsten Stuetzle for providing the model of the solar collector. We are grateful to Eric Haseltine for supplying the batch gas example. This paper was also improved by valuable comments from Prof. David Q. Mayne.

The authors gratefully acknowledge the financial support of the industrial members of the Texas-Wisconsin Modeling and Control Consortium. Research of the first and second authors was supported in part by NSF grants #CTS-0105360, while research of the third author was supported in part by NSF grants #MTS-0086559, #ACI-0196485, and #EIA-0127857. All simulations were performed using the NMPC control tool (`www.che.wisc.edu/~tenny/nmpc/`) under Octave (`www.octave.org`). NMPC and Octave are freely distributed under the terms of the GNU General Public License.

References

[1] Joao Albuquerque, Vipin Gopal, George Staus, Lorenz T. Biegler, and Erik B. Ydstie. Interior point SQP strategies for large-scale, structured process optimization problems. *Comput. Chem. Eng.*, 23(4):543–554, 1999.

[2] B. Wayne Bequette. Nonlinear predictive control using multi-rate sampling. *Can. J. Chem. Eng.*, 69:136–143, February 1991.

[3] Hong Chen and Frank Allgöwer. A quasi-infinite horizon nonlinear model predictive control scheme with guaranteed stability. *Automatica*, 34(10):1205–1217, 1998.

[4] D. Chmielewski and V. Manousiouthakis. On constrained infinite-time linear quadratic optimal control. *Sys. Cont. Let.*, 29:121–129, 1996.

[5] Moritz Diehl, H. Georg Bock, Johannes P. Schlöder, Rolf Findeisen, Zoltan Nagy, and Frank Allgöwer. Real-time optimization and nonlinear model predictive control of processes governed by differential-algebraic equations. *J. Proc. Cont.*, 12(4):577–585, 2002.

[6] Ali Jadbabaie, Jie Yu, and John Hauser. Unconstrained receding horizon control of nonlinear systems. *IEEE Trans. Auto. Cont.*, 46(5):776–783, 2001.

[7] Jay H. Lee. Modeling and identification for nonlinear model predictive control: Requirements, current status and future needs. In *International Symposium on Nonlinear Model Predictive Control, Ascona, Switzerland*, 1998.

[8] Hannah Michalska and David Q. Mayne. Robust receding horizon control of constrained nonlinear systems. *IEEE Trans. Auto. Cont.*, 38(11):1623–1633, 1993.

[9] R. H. Miller, I. Kolmanovsky, E. G. Gilbert, and P. D. Washabaugh. Control of constrained nonlinear systems: a case study. *IEEE Control Systems Magazine*, 20(1):23–32, 2000.

[10] Manfred Morari and Jay H. Lee. Model predictive control: The good, the bad, and the ugly. In Y. Arkun and W. H. Ray, editors, *Chemical Process Control—CPC IV*, pages 419–444, Amsterdam, 1991. Fourth International Conference on Chemical Process Control, Elsevier.

[11] Kenneth R. Muske and Thomas A. Badgwell. Disturbance modeling for offset-free linear model predictive control. *J. Proc. Cont.*, 12(5):617–632, 2002.

[12] Gabriele Pannocchia and James B. Rawlings. Disturbance models for offset-free MPC control. Accepted for publication in *AIChE Journal*, 2002.

[13] Ronald K. Pearson and Babatunde A. Ogunnaike. Nonlinear process identification. In Michael A. Henson and Dale E. Seborg, editors, *Nonlinear Process Control*, pages 11–110. Prentice Hall, 1997.

[14] S. Joe Qin and Thomas A. Badgwell. A survey of industrial model predictive control technology. Accepted for publication in *Control Engineering Practice*, July 2002.

[15] Christopher V. Rao and James B. Rawlings. Steady states and constraints in model predictive control. *AIChE J.*, 45(6):1266–1278, 1999.

[16] Christopher V. Rao, James B. Rawlings, and Jay H. Lee. Constrained linear state estimation – a moving horizon approach. *Automatica*, 37(10):1619–1628, 2001.

[17] Christopher V. Rao, James B. Rawlings, and David Q. Mayne. Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations. Submitted for publication in *IEEE Transactions on Automatic Control*, 2000.

[18] Douglas G. Robertson, Jay H. Lee, and James B. Rawlings. A moving horizon-based approach for least-squares state estimation. *AIChE J.*, 42(8):2209–2224, August 1996.

[19] Pierre O. M. Scokaert, David Q. Mayne, and James B. Rawlings. Suboptimal model predictive control (feasibility implies stability). *IEEE Trans. Auto. Cont.*, 44(3):648–654, March 1999.

[20] F. G. Shinskey. *Feedback Controllers for the Process Industries*. McGraw-Hill, New York, 1994.

[21] Thorsten Stuetzle, Nathan Blair, William A. Beckman, and John W. Mitchell. Use of linear predictive control for a solar electric generating system. In *Proceedings of System Simulation in Buildings*, Liège, December 2002.

[22] Thorsten A. Stuetzle. Automatic control of the 30 MWe SEGS VI parabolic trough plant. Master's thesis, University of Wisconsin–Madison, 2002.

[23] M. J. Tenny, S. J. Wright, and J. B. Rawlings. Nonlinear model predictive control via feasibility-perturbed sequential quadratic programming. Optimization Technical Report 02-06, University of Wisconsin-Madison, Computer Sciences Departments, August 2002. Also Texas-Wisconsin Modeling and Control Consortium Report TWMCC-2002-02.

[24] Matthew J. Tenny and James B. Rawlings. Efficient moving horizon estimation and nonlinear model predictive control. In *Proceedings of the American Control Conference*, pages 4475–4480, Anchorage, Alaska, May 2002.

[25] Matthew J. Tenny, James B. Rawlings, and Rahul Bindlish. Feasible real-time nonlinear model predictive control. In James B. Rawlings, Babatunde A. Ogunnaike, and John W. Eaton, editors, *Chemical Process Control– VI: Sixth International Conference on Chemical Process Control*, volume 97, Tucson, Arizona, January 2001. AIChE Symposium Series.