Proceedings of the American Control Conference
Chicago, Illinois • June 2000

Efficient Nonlinear Model Predictive Control

M. Cannon, B. Kouvaritakis, A.C. Brooms
Dept. of Engineering Science
Oxford University
Oxford, OX1 3PJ, UK
mark.cannon@eng.ox.ac.uk

Y.I. Lee
Dept. of Control and Instrumentation
Gyeongsang National University
Jinju Gyeong-Nam 660-701, Korea
ciman@nongae.gsnu.ac.kr

Abstract

For large numbers of degrees of freedom and/or high dimensional systems, nonlinear model predictive control algorithms based on dual mode control can become intractable. This paper proposes an alternative which deploys the closed-loop paradigm that has proved effective for the case of linear time-varying or uncertain systems. The various attributes and computational advantages of the approach are shown to carry over to the nonlinear case.

1 Introduction

Dual mode control [6, 1, 2] is now a well-established paradigm for use in nonlinear model predictive control (NMPC). According to this, a number of free control moves are deployed in order to steer the predicted state into a target set which is positively invariant and feasible with respect to a state feedback law $u = -Kx$ and system constraints. The free control moves are then chosen within their allowable limits so as to minimize an upper bound on an infinite horizon quadratic cost. It is now recognized [1] that this approach can become intractable for high order systems and/or for a large number of free control moves.

A similar problem of intractability applies to the case of linear time-varying or uncertain plants (even for the case of no free control moves [3]) because of the constraint that the final predicted state lies in a given target set. An effective remedy to this problem is provided by the so-called closed-loop paradigm [4]. The idea here is to shift as much of the on-line computational burden to off-line computations as possible, and the device for achieving this is remarkably simple. The predicted control moves are centred about a state feedback control law $u = -K(x)$, which is optimal (in a suitable sense) for the unconstrained case, and then modified over n_c samples as $u = -K(x) + c$, where the sole purpose of the perturbation c is to guarantee feasibility. To ensure a return to the unconstrained optimal law, the norm $\|c\|$ of the c trajectory minimized at successive sampling instants, subject to a guarantee of feasibility and a constraint that $\|c\|$ be monotonically decreasing. Both these requirements are met by an augmentation of state and through the use of feasible invariant ellipsoidal sets [5]. The computation of these sets, though systematic, could be demanding, however both this and the computation of K can be performed offline. More significantly it turns out that due to the use of ellipsoidal sets and the simple objective function defined by $\|c\|$, the online optimization reduces to the solution of a simple polynomial equation.

The purpose of the present paper is to show that the approach above can be extended to the case of NMPC. The key to this development is the use of a quadratic form of the triangle inequality in conjunction with a quadratic Lipschitz condition. The former is used to extract as much information as possible from linear approximations, whereas the latter is used in bounding the error of approximation. The use of the closed-loop paradigm together with ellipsoidal invariant sets implies that the computational benefits of the earlier algorithm carry over to the nonlinear case. The results of the paper are illustrated by means of a numerical example which for the purposes of transparency is chosen to be particularly simple, rather than challenging.

2 Autonomous formulation

Let the model equation be

$$x' = f(x,u), \quad |u| \leq \bar{u} \tag{1}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $|\cdot|$ denotes elementwise absolute value, and for convenience we use $(\cdot)'$ to indicate values at the next time instant (ie. $x' = x_{k+1}$ if $x = x_k$). Use any preferred method (eg. global linearization, extended linearization) to determine the unconstrained control law

$$u = -K(x) \tag{2}$$

which, in the absence of constraints is stabilizing and gives good performance for all initial conditions x_0 lying in a given set:

$$X_0(s) = \{x;\ |e_i x| \leq s\bar{x}_i,\ i = 1, \ldots, n\} \quad \text{for } s = 1 \tag{3}$$

where e_i denotes the ith row of the identity matrix (of conformal dimensions). The definition of $X_0(1)$ will be dictated by the particular application. For example, in the coupled tanks system considered in section 7 below, the maximum range for the states is determined by the physical dimensions of the tanks and the limits on the water pump.

The presence of input constraints may cause performance degradation and even instability. It therefore becomes necessary to perturb (2) as:

$$u = -K(x) + Ec, \quad E = \begin{bmatrix} I_m & 0 & \cdots & 0 \end{bmatrix} \tag{4}$$

where c is a block vector with n_c blocks. Substitution of (4) into (1) gives the prediction dynamics

$$x' = F(x,c), \quad |-K(x)+Ec| \le \bar{u}. \tag{5}$$

The n_c degrees of freedom in c can be used to steer x, over n_c steps, into a target set X_{n_c} which is invariant under the control law of (2) and for which this control law is feasible with respect to constraints. Moreover, since c represents a perturbation on the desirable control law of (2), a sensible objective is to make $||c||_2$ as small as possible. This suggests the following optimization problem

$$\min_c ||c||_2 \quad \text{subject to} \quad \begin{cases} (5) \\ x_{k+n_c|k} \in X_{n_c} \end{cases} \tag{6}$$

where $x_{k+n_c|k}$ is the n_c-step ahead prediction of x at time k.

Given the nonlinear dynamics of (5), this optimization is nonconvex and becomes computationally intractable for large n and n_c. Instead it is possible to define an augmented state z which includes c, and recast (5) as

$$z' = \Omega z + \begin{bmatrix} q \\ 0 \end{bmatrix}, \; z = \begin{bmatrix} x \\ c \end{bmatrix}, \; \Omega = \begin{bmatrix} \Phi & HE \\ 0 & T \end{bmatrix} \tag{7}$$

where T is the operator which assigns $e_i Tc = e_{i+1}c$, $i = 1,\dots,n_c-1$, and $e_{n_c}Tc = 0$. Thus Φ and HE represent the state- and input-maps of a linear approximation to the dynamics of (5), and q denotes the approximation error. The importance of considering the effects of Tc is twofold. If at the current time c defines (through (4)) a predicted input trajectory which is feasible, then Tc will define a feasible predicted input trajectory at the next time instant. Also the objective in (6) associated with Tc is necessarily less than or equal to the cost incurred at the previous time instant. The combination of these two properties provides a mechanism for guaranteeing the feasibility and stability of the receding horizon law developed below.

3 Invariant ellipsoidal sets

To guarantee future feasibility we make use of ellipsoidal sets $S_z = \{z;\ z^T Pz \le 1\}$ which have the following invariance and feasibility properties.

$$z \in S_z \Rightarrow \begin{cases} z' \in S_z \\ |-K(x)+Ec| \le \bar{u} \end{cases}$$

To ensure invariance we make use of a quadratic Lipschitz condition on the dynamics of (7),

$$||q||_2^2 \le z^T \Gamma z. \tag{8}$$

Define the projection of S_z onto the $c=0$ subspace as S_x (so that $z \in S_z \Rightarrow [I_n \;\; 0]z \in S_x$). Condition (8) will only be invoked inside $X_0(s)$, where s is such that $S_x \subset X_0(s)$. Hence the values of Φ, H, and Γ are all dependent on the choice of s.

In general the computation of suitable Φ, H, Γ is non-trivial but can be performed numerically offline. In some special cases, such as the one illustrated in section 7, the computation of the above constants is straightforward. Using (8) in conjunction with the lemma below it is possible to derive appropriate sufficient conditions for invariance.

Lemma 1. *For $\mu > 1$ define $\eta = 1+(\mu-1)^{-1}$. Then $\eta > 0$ and for any positive definite P we have*

$$z'^T Pz' \le \mu z^T \Omega^T P \Omega z + \eta q^T E_x P E_x^T q \tag{9}$$

where $E_x = [I_n \;\; 0]$.

Proof. Substituting (7) into (9), bringing all terms to the RHS and dividing by $\mu - 1$ yields

$$0 \le \left(\Omega z - \frac{1}{\mu-1}\begin{bmatrix} q \\ 0 \end{bmatrix}\right)^T P \left(\Omega z - \frac{1}{\mu-1}\begin{bmatrix} q \\ 0 \end{bmatrix}\right)$$

which is clearly true for $P > 0$. □

Theorem 1. *An ellipsoidal set $S_z = \{z;\ z^T Pz \le 1\}$ is invariant under (7) if*

$$E_x P E_x^T \le \gamma_x I_n \tag{10a}$$

$$\mu \Omega^T P \Omega + \eta \gamma_x \Gamma \le P \tag{10b}$$

where Γ satisfies (8).

Proof. By lemma 1 and (10a) we have

$$\begin{aligned} z'^T Pz' &\le \mu z^T \Omega^T P \Omega z + \eta q^T E_x P E_x^T q \\ &\le \mu z^T \Omega^T P \Omega z + \eta \gamma_x ||q||_2^2 \end{aligned}$$

so that invoking (8) and (10b) we get

$$z'^T Pz' \le \mu z^T \Omega^T P \Omega z + \eta \gamma_x z^T \Gamma z \le z^T Pz.$$

From the definition of S_z it follows that $z \in S_z \implies z' \in S_z$. □

4 The feasibility condition

For feasibility of S_z with respect to constraints we require that $|-K(x)+Ec| \le \bar{u}$ for all $z \in S_z$. Satisfaction of this condition can be ensured using an approach similar to that employed above to guarantee invariance. The offline derivation of the Lipschitz bound (8) on q can also be invoked to compute Lipschitz bounds on the feedback law (2) for $x \in X_0(s)$:

$$||k_i(x) - k_i x||_2^2 \le x^T \Gamma_i x, \; i = 1,\dots,m \tag{11}$$

where $K(x) = [k_1(x) \;\; \cdots \;\; k_m(x)]^T$ and $k_i x$ is a suitable linear approximation to $k_i(x)$. It is then easy to state conditions under which S_z is feasible.

Theorem 2. *For $\mu_i > 1$ and $\eta_i = 1 + (\mu_i - 1)^{-1}$, the feasibility of S_z is ensured by the following conditions for $i = 1, \dots, m$.*

$$Q_i = \mu_i \begin{bmatrix} -k_i^T \\ e_i^T \end{bmatrix} \begin{bmatrix} -k_i & e_i \end{bmatrix} + \eta_i \begin{bmatrix} \Gamma_i & 0 \\ 0 & 0 \end{bmatrix} \tag{12a}$$

$$Q_i \leq \bar{u}_i^2 P \tag{12b}$$

Proof. For feasibility we require $|-k_i(x) + e_i c| \leq \bar{u}_i$, $i = 1, \dots, m$, for all $z \in S_z$. But by arguments similar to those used in the proof of lemma 1 we have

$$\begin{aligned} |-k_i(x) + e_i c|^2 &= \left|-k_i(x) + k_i^T x + \begin{bmatrix} -k_i & e_i \end{bmatrix} z\right|^2 \\ &\leq \mu_i z^T \begin{bmatrix} -k_i^T \\ e_i^T \end{bmatrix} \begin{bmatrix} -k_i & e_i \end{bmatrix} z + \eta_i x^T \Gamma_i x \end{aligned}$$

and feasibility is therefore ensured by (12b). □

5 Offline computation

In the absence of the perturbation c, $z \in S_z$ implies $x \in \tilde{S}_x$ where $\tilde{S}_x = \{x;\ x^T E_x P E_x{}^T x \leq 1\}$. Clearly $\tilde{S}_x$ lies within S_x, and an obvious benefit of the use of the perturbation is the enlargement of the invariant ellipsoid. To enable the approach to be applied over as large a set of initial conditions as possible, it is important to choose P so as to maximize S_x in some sense. The lemma below provides a measure of the size of this ellipsoid and suggests an optimization procedure for the selection of P.

Lemma 2. *The volume of S_x is proportional to $\det(E_x P^{-1} E_x{}^T)$. Furthermore the linear matrix inequality*

$$P \leq \begin{bmatrix} \gamma_x I_n & 0 \\ 0 & \gamma_c I_{n_c} \end{bmatrix}$$

ensures condition (10a) as well as implying that the volume of S_x is greater than that of a spheroid of radius $1/\sqrt{\gamma_x}$. Similarly it ensures that the projection S_c of S_z onto the subspace has volume greater than that of a spheroid of radius $1/\sqrt{\gamma_c}$.

Proof. This follows from the partitioned inverse of P. □

The lemma above helps identify possible objective functions J_P which can be minimized in a procedure for the selection of P:

$$J_P = -\log\det(E_x P^{-1} E_x{}^T), \tag{13a}$$

$$J_P = \gamma_x, \tag{13b}$$

$$J_P = a_1 \gamma_x + a_2 \gamma_c, \tag{13c}$$

each meeting a different objective. For example (13a) addresses the problem of maximizing the volume of S_x; this would normally result in highly directional ellipsoids in that for most system models, small decreases in some semi-axes would be followed by significant increases of other semi-axes. Alternatively, minimization of (13b) would maximize the minimum semi-axis of S_x, thereby forcing S_x to be as near-spheroidal as possible. The third choice (13c) offers a compromise between directional and near spheroidal S_x; this can be achieved by suitable adjustment of the weights a_1, a_2.

Theorem 3. *If $u = -K(x)$ asymptotically stabilizes the origin $x = 0$ of (1) in the absence of constraints, then there exist Φ, H, s and $\mu > 1$ such that the minimization*

$$\min_{P, \gamma_x, \gamma_c} J_P \quad \text{subject to:}$$

$$\begin{bmatrix} P & P\sqrt{\mu}\Omega \\ \sqrt{\mu}\Omega^T P & P - \eta\gamma_x \Gamma \end{bmatrix} \geq 0 \tag{14a}$$

$$\begin{bmatrix} \gamma_x I_n & 0 \\ 0 & \gamma_c I_{n_c} \end{bmatrix} - P \geq 0 \tag{14b}$$

$$\begin{bmatrix} \bar{u}_i^2 P & Q_i^{1/2} \\ Q_i^{1/2} & I_{n_c} \end{bmatrix} \geq 0, \quad i = 1, \dots, m \tag{14c}$$

$$\begin{bmatrix} P & e_i^T \\ e_i & s^2 \bar{x}_i^2 \end{bmatrix} \geq 0, \quad i = 1, \dots, n \tag{14d}$$

is feasible, and for which the optimal solution P defines an ellipsoid S_z which is invariant and feasible.

Proof. It is easily shown that (14a) is equivalent to (10b) and that (14b) implies (10a); thus (14a) and (14b) together ensure invariance. Also (14c) is equivalent to (12b) and therefore ensures feasibility; whereas (14d) implies $S_x \subset X_0(s)$ and hence guarantees that (8) and (11) are valid for all $z \in S_z$. It therefore remains to show that the minimization is feasible. However, if $\Phi = \partial F/\partial x|_{(0,0)}$ and $H = \partial F/\partial c|_{(0,0)}$, then the stability of the prediction dynamics (5) in the absence of constraints implies that Ω is contractive by Lyapunov's linearization method. Consequently there exists $P > 0$ satisfying (14a) for any $\gamma_x > 0$ and $\Gamma > 0$ if $1 < \mu < 1/(\rho(\Omega))^2$, where $\rho(\Omega)$ denotes the spectral radius of Ω. This choice for Φ, H also ensures that $q = O(s^2)$ by Taylor's theorem, and from (8) we therefore have $\Gamma = O(s^2)$. It follows that, if P satisfies (14a) for $s = s^0$ and $\gamma_x = \gamma_x^0$, then for small s^0 and $\varepsilon < 1$, P also satisfies (14a) for $s = \varepsilon s^0$ and $\gamma_x = \gamma_x^0/\varepsilon^2$. Thus it is clear that all of the constraints of the minimization can be met simultaneously through choice of sufficiently small s and sufficiently large γ_x, γ_c. □

Remark 1. *The values of Φ, H assumed in the proof of theorem 3 were chosen simply to demonstrate existence of feasible solutions to (14a–14d). In practice other choices may be preferable (eg. in order to minimize conservativeness in the bound (8)).*

Remark 2. *For each of the objective functions J_P of (13a), (13b) and (13c), the optimization of theorem 3 is convex, and can be performed efficiently using semi-definite programming techniques.*

Theorem 3 above suggests the following offline procedure for determining P.

Algorithm 1 (Offline computation of P). *Set $s = 1$ and choose $\mu \in (1, 1/(\rho(\Omega))^2)$ to be just less than $1/(\rho(\Omega))^2$.*

1. *Compute Φ, H, Γ, and perform the minimization of theorem 3. If this is feasible, increase s and repeat until s can be increased no further. Otherwise decrease s and repeat until a feasible value for s is found.*
2. *For the computed P, if $S_x \not\subset X_0(1)$ then detune $K(x)$ and repeat step 1. If, after detuning, it is still not possible to find P such that $S_x \subset X_0(1)$, then redefine the control objectives (ie. choose a smaller set of desired initial conditions).*

6 Predictive control law

In the absence of constraints, the control law $u = -K(x)$ is assumed to provide good performance for all initial conditions in $X_0(1)$, and the purpose of the perturbation term Ec in the control law of (4) is simply to ensure stability when constraints are present. The NMPC strategy described in this section is therefore based on minimizing the objective of (6), but subject to the constraint that z lies in the feasible invariant ellipsoid S_z described in sections 3 and 4. This affords an enormous reduction in online computational burden over the optimization (6) since the online computation then reduces to solving for the (unique) negative real root of a polynomial function of a single variable, as we show below.

Algorithm 2 (Predictive control law). *At each sampling instant, compute*

$$\min_c \|c\|_2 \quad \text{subject to} \quad z \in S_z \tag{15}$$

and implement $u = -K(x) + Ec$.

Theorem 4. *The control law of algorithm 2 asymptotically stabilizes the equilibrium $x = 0$ of (1) with region of attraction S_x, and converges asymptotically to the control law $u = -K(x)$.*

Proof. The feasibility property of S_z implies that the control law of algorithm 2 satisfies constraints at all times, and by invariance of S_z it also follows that the optimization (15) is feasible at all times if $x_0 \in S_x$. The feasibility of $c' = Tc$ implies that the difference between the optimal values of the objective in (15) at consecutive sample times satisfies the bound

$$\|c'\|_2 - \|c\|_2 \leq -\|Ec\|_2.$$

It follows that the sequence of perturbations Ec is square-summable, and therefore that $Ec \to 0$ and $u \to -K(x)$. Given that the origin of (1) under $u = -K(x)$ is asymptotically stable in the absence of constraints, this implies that algorithm 2 is asymptotically stabilizing for the constrained plant. □

7 Illustrative example

The model of water pump connected to two coupled tanks, discretized for a sampling period of 4 seconds, is given by

$$x' = x + M\phi(x) + bu, \quad |u| \leq 0.5, \tag{16}$$

$$M = \begin{bmatrix} -1.0361 & 0.3684 \\ 0.5181 & -0.3684 \end{bmatrix}, \; b = \begin{bmatrix} 2.5029 \\ 0 \end{bmatrix},$$

where $\phi(x) = \begin{bmatrix} \phi_1(x_1) & \phi_2(x_2) \end{bmatrix}^T$ and

$$\phi_i(x_i) = |x_i + x_i^0|^{1/2}\text{sign}(x_i + x_i^0) - (x_i^0)^{1/2}, \; i = 1,2.$$

In terms of the water levels h_1, h_2 in the two tanks and the pump input signal U, we have

$$x_1 = h_1 - h_2 - x_1^0, \quad x_2 = h_2 - x_2^0, \quad u = U - U^0,$$

with $x_1^0 = 5.84$, $x_2^0 = 11.54$, $U^0 = 0.5$.

For simplicity we choose a linear state feedback law:

$$K(x) = Kx, \quad K = \begin{bmatrix} 0.2985 & 0.2155 \end{bmatrix}, \tag{17}$$

which, in the absence of constraints, is optimal with respect to an LQ cost (with state and control weights I and 1 respectively) for (16) linearized about $x = 0$. To decouple the dependence of q on the elements of x, we take $\Phi = I - MD - bK$, where $D = \text{diag}\{0.207, 0.1472\}$, and $H = bE$ in (7); thus $q = M(\phi(x) - Dx)$. Defining $X_0(1)$ to cover $\pm 100\%$ of the steady state values of x_1,x_2 (so that $\bar{x}_1 = x_1^0$, $\bar{x}_2 = x_2^0$ in (3)), we have

$$|\phi(x) - Dx| \leq \begin{bmatrix} 0.0355 & 0 \\ 0 & 0.0253 \end{bmatrix} |x|, \; \forall x \in X_0(1),$$

and since both elements of $\phi(x) - Dx$ are negative for all $x \in X_0(1)$, a suitable Γ in the Lipschitz condition (8) is given by $\Gamma = \text{diag}\{\Gamma_x, \Gamma_c\}$,

$$\Gamma_x = \begin{bmatrix} 0.0017 & 0 \\ 0 & 0.0002 \end{bmatrix}, \; \Gamma_c = 0.$$

The critical value for μ is $1/(\rho(\Omega))^2 = 1.2789$, and the choice $\mu = 1.25$ results in $\eta = 5$.

The invariant sets S_x, computed using algorithm 1 with objective (13a), are shown in figure 1 for $n_c = 0$, and $n_c = 5$. The dashed lines lie on the boundary of the largest possible admissible set for $u = -K(x)$, and the dotted lines show the boundary of $X_0(1)$. Clearly the area of S_x increases with n_c, thereby showing the benefits of the degrees of freedom in c. Although the increase in area for $n_c = 5$ appears modest, this is primarily due to the constraint $S_x \in X_0(1)$, which is active for $n_c \geq 5$.

The simulation results for two initial conditions, $x_0 = [1,3]^T$ (run 1) and $x_0 = [-5.5,3]^T$ (run 2), both of which lie outside the feasible region for the unconstrained optimal control law $u = -K(x)$ of (17), are shown in figures 2 and 3. As expected, the constrained responses for algorithm 2 are worse than the unconstrained LQ responses (which violate constraints). However the closed-loop costs defined by

$$J_{\text{run}} = \sum_{k=0}^{\infty} (x_k^T Q x_k + u_k^T R u_k); \quad Q = I,\ R = 1$$

are comparable, as shown in table 1.

Table 1: Closed-loop runtime costs J_{run}

	run 1	run 2
Unconstrained LQ	223.67	214.43
Algorithm 2	236.14	219.63

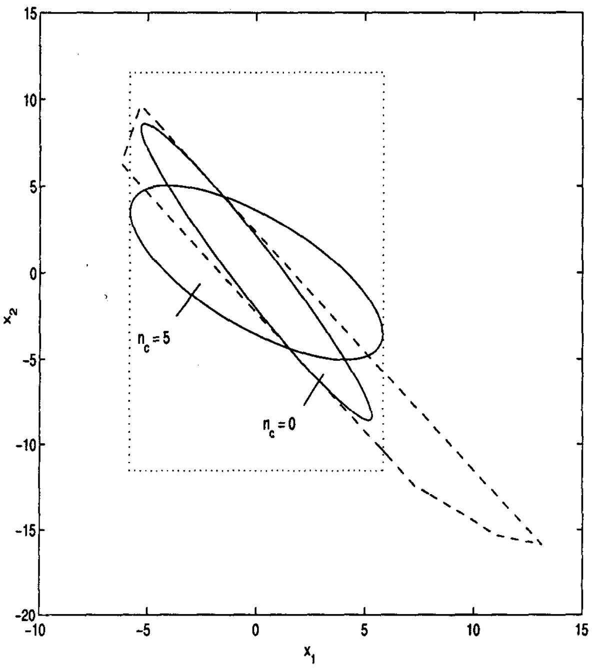

Figure 1: Invariant ellipsoidal sets of maximum area.

Acknowledgments The authors thank EPSRC for financial support.

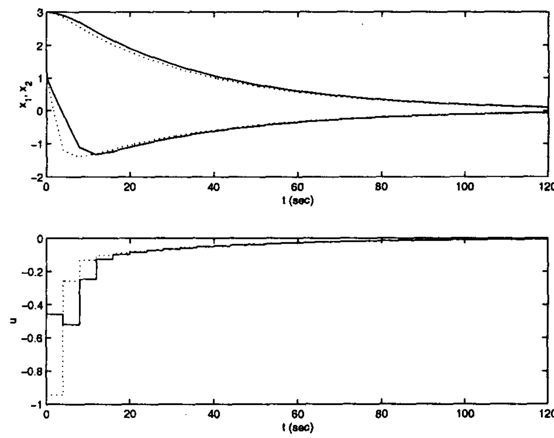

Figure 2: State (above) and input responses (below) for run 1: constrained (solid line) and unconstrained (dotted line).

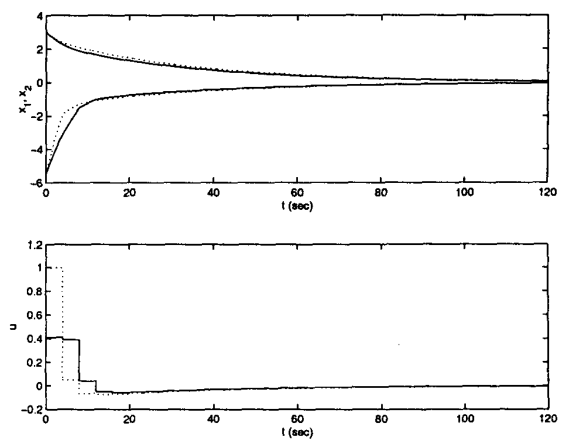

Figure 3: State (above) and input responses (below) for run 2: constrained (solid line) and unconstrained (dotted line).

References

[1] H. Chen and F. Allgöwer. Quasi-infinite horizon nonlinear model predictive control scheme with guaranteed stability. *Automatica*, 34(10):1205–1217, 1998.

[2] G. De Nicolao, L. Magni, and R. Scattolini. Stabilizing receding horizon control of nonlinear time-varying systems. *IEEE Trans. Autom. Control*, 43(10):1030–1036, 1998.

[3] M. V. Kothare, V. Balakrishnan, and M. Morari. Robust constrained model predictive control using linear matrix inequalities. *Automatica*, 32(10):1361–1379, 1996.

[4] B. Kouvaritakis, J. A. Rossiter, and M. Cannon. Stability, feasibility, optimality and the degrees of freedom in constrained predictive control. In *Nonlinear Model Predictive Control: Assessment and Future Directions for Research*. Birkhäuser-Verlag, 1998.

[5] B. Kouvaritakis, J. A. Rossiter, and J. Schuurmans. Efficient robust predictive control. In *Proc. ACC*, 1999.

[6] D. Q. Mayne and H. Michalska. Receding horizon control of nonlinear systems. *IEEE Trans. Autom. Control*, 35(7):814–824, 1990.