"Model Predictive Control: Multivariable Control Technique of Choice in the 1990s?"
M. Morari

Control and Dynamical Systems
California Institute of Technology
Pasadena, CA 91125
Model Predictive Control: Multivariable Control Technique of Choice in the 1990s?

Manfred Morari

Abstract

The state space and input/output formulations of model predictive control are compared and preference is given to the former because of the industrial interest in multivariable constrained problems. Recently, by abandoning the assumption of a finite output horizon several researchers have derived powerful stability results for linear and nonlinear systems with and without constraints, for the nominal case and in the presence of model uncertainty. Some of these results are reviewed. Optimistic speculations about the future of MPC conclude the paper.

1 Introduction

The objective of this paper is to review some major trends in model predictive control (MPC) research with emphasis on recent developments in North America. We will focus on the spirit rather than the details, i.e. we do not attempt to provide a complete list of all the relevant papers published during the last few years.1 We will try to contrast the motivations driving the research in the different camps.

There is little doubt that most of the research on model predictive control in the U.S. and much of the research elsewhere started with the publication of the seminal paper by Cutler and Ramaker (1980) from Shell. This is not to suggest that they invented model predictive control, but they did convince a generation of control consultants, application engineers, managers, and researchers of the merits and the potential of this type of tool for industrial applications. Early joint work by Amoco and IBM (Crowther, Pitrak and Ply 1961, Kuehn and Davidson 1961, Pendleton 1961) contains some of the essential features, but does not take into account process dynamics. There is also the theoretical work on “open-loop optimal feedback” with references going back to 1962 which is reviewed in the thesis by Gutman (1982). We do not wish to get entangled in historical disputes, but rather side with Richalet (1993b) who recalls the following discussion taking place at the IFAC Congress in Munich in 1987:

- Who invented predictive control?
- God...
- Predictive control is a discovery, not an invention...but God needs prophets.

1A simple database search for “predictive control” generated 128 references for the years 1991-1993 alone.
Model predictive control is now widely used by industry. The various implementations of MPC preferred by the different vendors and users are identical in their main structure, but differ in details. These details are largely proprietary and are often critical for the success of the algorithm in an application. The general structure is shown in Figure 1. An observer utilizes knowledge of the plant inputs $u$ and measurements $y$ to arrive at a state estimate $\hat{x}$. Starting from the current state estimate $\hat{x}$, one can employ classic prediction algorithms to predict the behavior of the process outputs over some output horizon $H_p$ when the manipulated inputs $u$ are changed over some input horizon $H_c$ (Figure 2).

The task of the optimizer is to compute the present and future manipulated variable moves $u(k), \ldots, u(k + H_c - 1)$ such that the predicted outputs follow the reference in a desirable manner. The optimizer takes into account constraints on the inputs and outputs which may be present. For linear process models, depending on the objective function, either a linear or a quadratic program results which is solved on-line in real-time at each time step. For commercial applications, various vendors have developed short-cut optimization procedures.

Only $u(k)$, the first one of the sequence of optimal control moves is implemented on the real plant. At time $k + 1$ another measurement $y(k + 1)$ and another state estimate $\hat{x}(k + 1)$ is obtained, the horizons are shifted forward by one step, and another optimization is carried out. This procedure results in a moving horizon or receding horizon strategy. A key feature of the technique is that the input and output horizons ($H_c$ and $H_p$) are generally finite. Often the values chosen for $H_c$ and $H_p$ are different. Furthermore, in some of the algorithms, there is the option not to include the control error during the first few time steps in the objective function. The problem definition as presented allows one to treat with equal ease multivariable problems with an unequal number of inputs and outputs, non-minimum phase systems and systems subject to constraints.

Many applications are reported in the literature and even more in sales publications. Some of them are mentioned in the review by García, Prett and Morari (1989) and in the more recent summary article by Richalet (1993a). MPC also enjoys widespread use in the
Japanese process industries, as one can learn from the survey published by Yamamoto and Hashimoto (1991). It is most significant that in a similar survey ten years prior (Hashimoto and Takamatsu 1982), MPC can not even be found in the list of control techniques.

2 Problem definition and assumptions

We will assume throughout most of this paper that the system to be controlled is linear time invariant discrete-time and that the system model and its parameters are known. A discussion of the advantages and disadvantages of the MPC formulation for adaptive control is beyond the scope of this paper. Moreover, most works on adaptive MPC simply combine some type of parameter estimator with MPC without any analysis of the resulting nonlinear system.

In North America a state space formulation has been dominant. The details are available in the paper by Lee, Morari and García (1994). As demonstrated in that paper, the popular step response models used, for example, in Dynamic Matrix Control and other algorithms are just a special realization of a state space model. In Europe there has been a preference for input/output descriptions; Soeterboek (1991) provides an excellent exposition of the formulation and the assumptions.

2.1 State space formulation

Assume that the system is described by

\[ x(k) = Ax(k-1) + Bu(k-1) + w(k-1) \]  \hspace{1cm} (1)  
\[ y_m(k) = Cx(k) + v(k) \] \hspace{1cm} (2)

where the customary nomenclature has been employed. The vector of manipulated variables is \( u \), \( y_m \) is the vector of process measurements, \( w \) is the state disturbance and \( v \) the mea-
surement noise. The disturbance and the noise could be of a deterministic or a stochastic nature. In the latter case, this model can describe signals of arbitrary spectral density. For a good discussion of disturbance modeling, the reader is referred to chapter 6 of the book by Åström and Wittenmark (1990). The theory for output prediction is well developed (see for example the books by Åström and Wittenmark (1990), and Goodwin and Sin (1984)). It is summarized in the following:

\[
\begin{align*}
x(k|k - 1) &= Ax(k - 1|k - 1) + Bu(k - 1) \\
y(k|k - 1) &= Cx(k|k - 1)
\end{align*}
\]

Correction based on measurements:

\[
x(k|k) = x(k|k - 1) + K(y_m(k) - y(k|k - 1))
\]

Prediction:

\[
\begin{align*}
x(k + 1|k) &= Ax(k|k) + Bu(k) \\
y(k + 1|k) &= Cx(k + 1|k)
\end{align*}
\]

The filter gain \( K \) is determined from the solution of a Riccati equation. Prediction for more than one step ahead is obtained by applying the prediction equations recursively. The present and future control actions are found from the solution of the following optimization problem.

\[
\min_{u(k), u(k+1), \ldots} \sum_{i=1}^{H_p} x(k + i|k)^T R_i x(k + i|k) + \sum_{i=1}^{H_o} u(k + i - 1)^T S_i u(k + i - 1)
\]

Through the appropriate definition of the weighting matrices \( R_i \) and \( S_i \) a range of objectives can be expressed. For example, outputs at the end of the horizon can be emphasized more than at the beginning. One can further generalize the objective and penalize changes in the manipulated variables \( \Delta u \) as well.

### 2.2 Input/Output formulation

The model has the form

\[
y_m(k) = \frac{q^{-d}B}{A} u(k - 1) + \frac{C}{D} e(k)
\]

where \( A, B, C, \) and \( D \) are polynomials in the forward shift operator \( q \), and \( e \) are the disturbance inputs which can be either deterministic or stochastic (white noise). When comparing equations 1 and 2 with equation 9, we note that the two external inputs \( w \) and \( v \) have been replaced by a single input \( e \). It can be shown, that without loss of generality it is indeed possible to represent the effect of several stochastic inputs on the measured output \( y_m \) by a single input \( e \). The following algorithm
provides the $i$ step prediction. The polynomial $F$ is found from the solution of a Diophantine equation.

A limitation inherent in the input/output approach is that it does not provide an estimate and a predicted value for the true output $y$ but only a predicted value of the measured output $y_m$ (including all the measurement noise) which is not of direct interest for control. The trade-off between close tracking of the true output $y$ in the presence of disturbances and the rejection of measurement noise is determined indirectly by the designer through the specification of the observer polynomial $C$. In principle, for a single-input single-output system, the same prediction can be obtained via a state space or an input/output approach, but it is arguably simpler for the designer to achieve the desired trade-offs by specifying the disturbance ($w$) and noise ($v$) parameters than the observer polynomial $C$.

The present and future control actions are found by solving the optimization problem

$$
\min_{u(k), u(k+1), \ldots} \sum_{i=H_m}^{H_p} \left[ P y_m(k+i|k) - P(1)r(k+i) \right]^2 + \rho \sum_{i=1}^{H_c} \left[ \frac{Q_n}{Q_d} u(k+i-1) \right]^2
$$

where $P$, $Q_n$ and $Q_d$ are polynomials in $q$. It is well known that for any arbitrary $P$, $Q_n$ and $Q_d$ an identical objective can be expressed in the state space formulation.

While the input/output approach can, in principle, be generalized to multi-input multi-output systems, this is quite awkward notationally. More importantly, the numerical problems which have to be solved are inherently sensitive so that it is impossible to develop reliable solution procedures except for some very specific simple problems. Also, even in the case that there is only one manipulated variable and one output of interest, there is often more than one disturbance or noise process. Trade-offs, which are required to take into account these multiple noise processes in the prediction algorithm, are often expressed more directly in the state space framework than in the input/output formulation.

### 2.3 The finite horizon assumption

If one selects $H_p = H_c = \infty$, the well studied linear quadratic Gaussian (LQG) optimal control problem results, which has been studied extensively for decades. It has some nice properties, most importantly that the resulting controller is a constant gain acting either on the states, if available, or the state estimates, and that closed-loop stability is guaranteed under rather general assumptions. So why did the model predictive control researchers in the last decade decide to adopt a finite receding horizon formulation? Three reasons have been mentioned.
- **Simpler computation:** In certain situations, it may be simpler to use the MPC approach to find the controller gain matrix via a least squares problem, rather than by solving a Riccati equation which is necessary in the infinite horizon case.

- **Constraints:** It is not immediately clear how a problem involving constraints on both manipulated variables and process outputs can be addressed in an infinite horizon setting.

- **More tuning flexibility:** The variable horizon length may offer another tuning parameter to achieve improved performance and robustness.

Unfortunately in retrospect there is little merit to these and other arguments in favor of a finite horizon approach.

- **Simpler computation:** With today’s computer power at our disposal, the computational issue is largely irrelevant.

- **Constraints:** We can argue that the constrained case can be handled in an infinite horizon setting \((H_p = H_e = \infty)\) as well. Let us assume for simplicity that we are regulating the state from some initial state \(x_0\) to the origin and that the optimization problem is feasible, i.e. there exists a solution \(u(k), u(k+1), \ldots\) which satisfies all the constraints and brings the state back to the origin. Clearly, the steady state solution \(u_{ss} = 0, x_{ss} = 0\) is feasible and inside the constraint set. Thus, the problem is only constrained initially when the state is far from the origin and becomes unconstrained after sufficiently long time. This time can be estimated from some simple norm arguments. Therefore, we can solve the constrained problem over an infinite horizon by appropriately splicing together the solution for a constrained finite horizon and an unconstrained infinite horizon problem. The details are given by Rawlings and Muske (1993) and are summarized below.

- **More tuning flexibility:** Tuning of control systems based on a finite horizon approach is often exceedingly difficult. First and foremost, there are no stability guarantees, i.e. it is not known a priori what sets of tuning parameters will give rise to a system which is closed-loop stable. Moreover, the effect of the available parameters is often non-monotonic as demonstrated by Soeterboek (1991). For example, increasing a particular parameter like the input weight \(\rho\) which one would expect to suppress control action and stabilize the system, can actually destabilize a system. Upon further increase of the parameter, stable behavior is found. This is shown in Figure 3. This behavior is not observed with \(H_p = \infty\) (Figure 4).

The stability results which have been obtained for finite horizon formulations are all very weak (see for example the early results by García and Morari (1982), Clarke, Mohtadi and Tufts (1987), Clarke and Mohtadi (1987).) They are either of an asymptotic nature, utilizing the well known results for the infinite horizon problem, or apply to very particular situations only (a specific class of systems, deadbeat control, etc.). After more than a decade of research, it appears safe to assume that there are no generally useful stability results for finite horizon controllers and that it is time to revise the problem formulation to obtain more
powerful results. Indeed, this is exactly what has been initiated by several research groups independently during the last couple of years and a wealth of exciting results have appeared, establishing for the first time a solid theoretical base for model predictive control.

3 Infinite horizon MPC

3.1 The basic idea

Representative for the approach taken by various groups we reproduce a result by Rawlings and Muske (1993) which is particularly enticing because of its simplicity. Let the receding horizon problem at time $k$ be defined through

\[
x(k+1) = Ax(k) + Bu(k),
\]

\[
\Phi_k = \min_{u(k), u(k+1), \ldots, u(k+H_e-1)} \sum_{i=k}^{\infty} (x^T(i)Rx(i) + u^T(i)Su(i)); \quad R, S > 0
\]

and be subject to:

\[
Du(i) \leq d, \quad i = k, k+1, \ldots, k + H_e - 1
\]
Figure 4: Same system as in Fig 3. $H_p = \infty$. For any input horizon $H_c$ the system behavior is "monotonic" as the input weight $\rho$ penalizing $\Delta u$ is increased ($\rho = 0$ solid; $\rho = 0.1$ dash; $\rho = 1$ dot)

$$H x(i) \leq h, \quad i = k_1, k_1 + 1, \ldots$$

(16)

The problem statement assumes that there are no disturbances or noise and that the states are measurable. This allows us to replace the predicted states $x(i|k)$ by $x(i)$ in the objective function. Then we have the following theorem.

**Theorem 1** Assume that the optimization problem stated above is feasible. For stabilizable \{A,B\} with $r$ unstable modes and $H_c \geq r$, $x = 0$ is an asymptotically stable solution of the closed-loop receding horizon controller for the quadratic program, defined above.

If the problem as stated is not feasible then the solution $u(k), u(k + 1), \ldots$ is not defined and an alternate problem has to be defined which is feasible, for example by relaxing the constraints. Because it is often impossible to enforce state constraints at all times, it may be necessary to select $k_1 > k$. If the open-loop system is unstable, then it is either necessary to choose $H_c = \infty$ or to force the unstable modes to be zero at $k = H_c$ (i.e. immediately after the control action stops), otherwise the objective is unbounded. We sketch the proof of the theorem because the ideas are simple and instructive.

**Proof:** If the problem is feasible, then $\Phi_k$ is finite. For arbitrary $k > 0$, $\Phi_{k+1} \leq \Phi_k - (x^T(k)R x(k) + u^T(k)S u(k))$ because the optimal control action $u(k), u(k + 1), \ldots$ computed at time $k$ is feasible at time $k + 1$. 

8
It follows that $\Phi_k$ is a Lyapunov function which is nonincreasing. Indeed, it can be shown that $\Phi_k$ is monotonically decreasing, establishing asymptotic closed loop stability.

If the problem is unconstrained, it can be solved in the standard manner through a combination of Lyapunov and Riccati equations depending on if $H_c$ is finite or not. If the problem is constrained we have to determine first from simple bounding arguments the time point $k'$ after which the solution is guaranteed to be unconstrained as discussed above. The matrix $P$ defining the optimal value of the objective function, $x_{k'}^TPx_{k'}$, after this point is found from the solution of an infinite horizon unconstrained problem (Lyapunov or Riccati equation). We then formulate a finite horizon (up to time $k'$) constrained problem with $x_{k'}^TPx_{k'}$ added to the finite horizon objective function. A Quadratic Program results which can be solved readily.

The feedback control algorithm established through this theorem is noteworthy in that it guarantees asymptotic stability for linear systems in the presence of general constraints whenever such a stabilization is possible at all. Therefore, as we will show below, it can solve some exceedingly difficult stabilization problems which have defied traditional approaches.

While the work by Rawlings and Muske (1993) is exemplary in its clarity, it must be mentioned that other authors (e.g., Mayne and Michalska (1990), Kouvaritakis, Rossiter and Chang (1992)) have suggested independently to prove stability via the Lyapunov function $\Phi_k$. An alternate but essentially equivalent approach is to enforce a state constraint $x_{H_c} = 0$ at the end of a finite output horizon. (Some of the early work is due to Kwon and Pearson (1977) but the ideas have seen a revival recently (Clarke and Scattolini 1991a, Clarke and Scattolini 1991b, Leva and Scattolini 1993).) Translated into the infinite output horizon framework the terminal state constraint forces the objective to be identically equal to zero after $H_c$. This approach is identical to setting the output horizon to infinity when the system is FIR and when the output horizon $H_c$ has been chosen long enough for the system to settle.

Until the publication of the recent work just described, establishing stability for receding horizon control systems in the presence of constraints seemed an exceedingly difficult task (Zafiriou 1990), not only because the problem is nonlinear but because there is no explicit functional description of the control algorithm as is required for most stability analyses. The recent work removed this technical and - to some extent - psychological barrier (people did not even try) and started widespread efforts to tackle extensions of this basic problem with the new tools. We will describe some of these developments in the following.

Much more will undoubtedly follow in the next few years. For example, all the results established so far assume state feedback. Because of the nonlinearities introduced by the constraints the separation principle does not hold and the stability properties under output feedback are not obvious. Application of the MPC concept to sampled-data nonlinear systems (Mayne and Michalska 1990) usually leads to involved optimization problems. The available numerical tools may not be reliable enough to permit the unsupervised on-line use of these techniques. It may be preferable to use a type of gain-scheduled linear MPC controller based on a local linear approximation of the nonlinear system. de Oliveira and Morari (1994) have developed the prototype of such an algorithm and have proven some stability properties.
3.2 Stabilization of linear discrete-time systems with actuator constraints

Consider the plant equation 13 and assume that the system is stabilizable, that all the eigenvalues of $A$ are in the closed unit disk and that each component of $u$ is magnitude bounded ($|u_i| \leq \varepsilon$). For discrete systems, Sontag (1984) proved the existence of a feedback controller which globally asymptotically stabilizes this system. However, the construction of a stabilizing controller is difficult (Sontag and Yang 1991, Sussman, Sontag and Yang 1991). For example, Teel (1992) showed that for a system with more than two integrators the stabilizing controller has to be nonlinear. It turns out to be quite easy to prove that MPC globally asymptotically stabilizes such a system (Balakrishnan, Zheng and Morari 1993). Moreover, the tuning parameters (weights) available in MPC can be used in a transparent manner to obtain excellent performance.

**Example 1** (Tsirakis and Morari 1992) Consider the following system from Sontag and Yang (1991)

$$
\begin{align*}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= -x_1 + x_3 \\
\dot{x}_3 &= x_4 \\
\dot{x}_4 &= -x_3 + u
\end{align*}
$$

where $u$ must satisfy the constraint $|u| \leq 1$. The system has four poles on the imaginary axis $(-j, -j, j, j)$.

The system was discretized with a sampling time of 0.1 to apply the MPC algorithm. The initial condition is $x_0 = [1 0.5 0.5 1]^T$. The weights are $R = I$ and $S = 10$. The input horizon is $H_c = 50$. Figure 5 depicts the time-evolution of state $x_1$ for the controller from Sontag and Yang (1991) and the MPC controller. The behavior of the other three states is similar. The corresponding control actions are shown in Figure 6. Although both controllers stabilize the system, the difference in performance is striking. In all fairness, we should point out that the controller was designed by Sontag and Yang (1991) to ensure stability and that they made no attempt to achieve good performance.

3.3 Asymptotic stabilization of nonlinear discrete-time systems

Consider the system

$$
x_{k+1} = f(x_k, u_k)
$$

where $f$ is assumed to be continuous and $f(0, 0) = 0$. Meadows and Rawlings (1993) showed that MPC can globally asymptotically stabilize such a system, assuming that the implied optimization problem is feasible and that some other mild conditions are satisfied. This is particularly remarkable if one looks at the following simple example (a discretized version of the system studied by Hermes (1991)) with the two states $x$ and $y$:

$$
\begin{align*}
x_{k+1} &= x_k + u_k \\
y_{k+1} &= y_k + u_k^3
\end{align*}
$$
Implicitly, MPC generates the state feedback law

\[ u_k = g(x_k, y_k) \]  

which according to theory globally asymptotically stabilizes the system. It can be shown (Meadows, Henson, Eaton and Rawlings 1993a) that any state feedback law \( g(x, y) \) which globally asymptotically stabilizes this system must be discontinuous. MPC automatically generates such a discontinuous function \( g(x, y) \). Needless to say, it would be exceedingly difficult to design such a feedback control system with other tools. The reader is referred to Meadows, Muske and Rawlings (1993b) for further discussions and another example.

### 3.4 Robust control of linear systems with input constraints

In the robust control context the model which describes the dynamical behavior of the real plant is not known exactly, but assumed to lie in a family of models \( \mathcal{P} \) which is parameterized in some fashion. Because there is no single model we cannot generate a single prediction but only a set of predictions corresponding to the family \( \mathcal{P} \). Accordingly, the control objective assumes a set of values and it is reasonable to define as the optimal control strategy the one which minimizes the largest (worst) value in the set:

\[ \Psi'_k = \min_{u(k), u(k+1), \ldots} \max_{\mathcal{P}} \sum_{i=k}^{\infty} x^T(i) R x(i) + u^T(i) S u(i); \ R, S > 0 \]  

(21)
In general, the family $\mathcal{P}$ is such that it is not possible to find a control strategy $u(k), u(k+1), \ldots$ which drives the predicted state $x$ to zero for all models in the family $\mathcal{P}$. To make the problem meaningful one can use different norms in the objective function (Campo and Morari 1986, Campo and Morari 1987), for example, the $1$-norm spatially and the $\infty$-norm temporally, resulting in the modified objective:

$$\Psi_{k} = \min_{u(k), u(k+1), \ldots} \max_{\mathcal{P}, i = k, \ldots} \left[ \|Rx(i)\|_{1} + \sum_{i=k}^{k+H_c} \|Su(i)\|_{1} \right]$$

(22)

which is always finite when the system is open loop stable. In general, this min-max optimization problem is very involved but in the special case, when the system is FIR and when the family $\mathcal{P}$ is parameterized in terms of "uncertain" impulse response coefficients which can vary between some upper and some lower bound, the min-max problem leads to a linear program of modest size. (See the original formulation by Campo and Morari (1987) and the simplifications by Allwright and Papavasiliou (1992).) The problem statement is appealing but unfortunately Zheng and Morari (1993) showed that the resulting feedback control law is not robustly stabilizing. After a slight generalization they were able to prove several important results which demonstrate that this new MPC controller can robustly stabilize a large class of FIR systems.

Robust control has taken center stage in control research during the last decade. It appears, however, that the conventional approaches not based on MPC have not led to a synthesis tool for controllers which provide global asymptotic robust stability guarantees in
the presence of actuator saturation. This further demonstrates the promise of the MPC approach to controller synthesis.

4 Conclusions

In the first part of the paper we compared the state space formulation of MPC with the input/output formulation. We concluded that, while in the single-input single-output case the differences between the two approaches are a matter of taste, the latter does not generalize well to multivariable systems which are of interest in industrial applications. The input/output approach may be preferable for adaptive formulations which were not discussed in this paper.

We argued that under the still popular assumption of a finite output horizon it appears to be impossible to provide stability guarantees which are general enough to be of practical value. In the last few years several researchers have modified this assumption in various similar ways (infinite horizon, terminal state constraint) and several theoretical results have emerged which demonstrate that MPC can provide powerful solutions to problems which have defied conventional approaches, like the asymptotic stabilization of nonlinear discrete-time systems and the robust control of systems with actuator constraints. We take this as an indication that in the next few years MPC will make inroads in new areas and is likely to emerge as a versatile tool with many desirable properties and with a solid theoretical foundation.

Acknowledgment Financial support from the National Science Foundation is gratefully acknowledged. I would also like to thank Alex Zheng for his help in preparing this paper.

References


