Putting Nonlinear Model Predictive Control into Use

Bjarne A. Foss[1] and Tor S. Schei[2]

[1] Department of Engineering Cybernetics, Norwegian University of Science and Technology, Trondheim, Norway
   Bjarne.Foss@itk.ntnu.no
[2] Cybernetica AS, Norway
   Tor.S.Schei@cybernetica.no

Summary. We will in this paper highlight our experience with NMPC. In our context NMPC shall mean the use of a nonlinear mechanistic model, state estimation, and the solution of an online constrained nonlinear optimisation problem. Our reference base is a number of applications of NMPC in a variety of processes.

We discuss the use of mechanistic models in NMPC applications and in particular the merits and drawbacks of applying such models in online applications. Further, we focus on state estimation, and the use of Kalman filters and moving horizon estimation. Finally, we consider the design of the optimization problem itself and implementation issues.

1 Introduction

Nonlinear model predictive control (NMPC) opens for the use of MPC in more demanding applications than has normally been the case for linear MPC. In particular NMPC lends itself to nonlinear systems which exhibit large variations in operating conditions and which are critically dependent on the use of a dynamic nonlinear model to gain sufficient performance. A nice overview of NMPC can be found in [9].

NMPC is not a well defined term in the sense that NMPC may be used for controllers ranging from a slight variation of linear MPC to the online solution of a constrained nonlinear optimisation problem. One example of a slight modification to account for nonlinearities is the use of multiple linear models in such a way that the current working point defines which model should be active at a given time instant. Hence, the QP-problem frequently encountered in linear MPC will change as the active model changes. In our context NMPC shall mean the use of a nonlinear mechanistic model, state estimation, and the solution of an online constrained nonlinear optimisation problem.

The scope of this paper is to pinpoint critical issues when applying NMPC by drawing on our experience within the process industries since 2000. To ensure a sound level of credibility we first present our most important application areas and accompanying control challenges in some detail. Thereafter we address four

critical areas: modelling, state estimation, formulation of the control problem itself and implementation issues. The paper ends with some conclusions.

2   Reference Base

The reference base contains a number of industrial applications of NMPC in a variety of processes. Some examples are:

- A system for optimization of suspension PVC polymerization processes has been implemented on two large (140 m3) autoclaves located at Hydro Polymers' plant in Porsgrunn, Norway. The system is characterized as follows: It contains a rather detailed nonlinear model of the polymerization reactor. The reactor model includes reaction kinetics, thermodynamic calculations for the four-phase suspension process, quality parameters and energy balances for the suspension and cooling systems. The application optimizes the temperature reference trajectory and the amount of initiators charged to the reactor in order to initiate the polymerization process. The optimization is based on an economic criterion which includes the batch time as well as cost of initiators. The process is highly exothermic, and the purpose of the optimization is to minimize the batch time without exceeding available cooling capacity and without using an inhibitor to slow down the polymerization process. The optimization is performed once for each batch cycle, and the temperature profile, which consists of approximately 80 "straight line segments", is optimised for the entire batch under a number of constraints imposed by the quality specifications of the various PVC products. Based on logged data a few model parameters are estimated as functions of conversion. The system is implemented using Cybernetica's NMPC and batch optimisation platform.
- Three NMPC applications for stabilization and quality control of the Borealis polypropolyne plant in Schwechat, Austria have been developed and implemented in cooperation with Borealis' personnel. The implementations are based on nonlinear first-principles models of the polyolefine plant (including three different polymerization reactors), and on Borealis in-house system for model predictive control (BorAPC). The system is characterized as follows: The three MPC applications comprise a multivariable control system with all together 11 control inputs and 19 controlled outputs. The nonlinear model consists of 77 states which are estimated on-line together with a few model parameters.
- NMPC of a base-catalyzed phenol-formaldehyde batch polymerization process has been implemented. The system is based on a rigorous model of the polymerization reactor. The system is implemented at Dynea's plant in Springfield, Oregon, USA. The model for this condensation polymerization process includes reaction kinetics, thermodynamics, population balances for functional groups and energy balance for the reactor and cooling system. Safety is an important issue and one driving force for implementing NMPC.
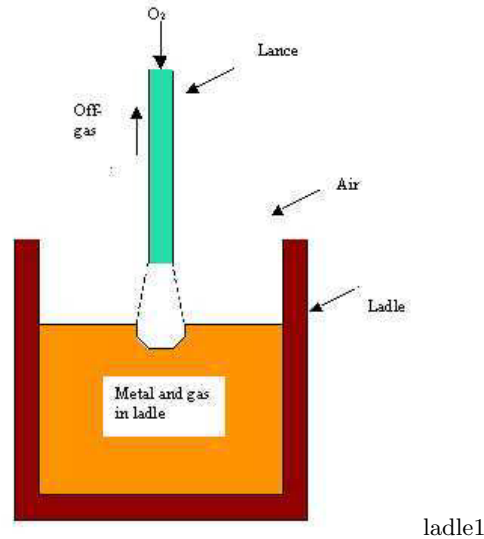
ladle1

Fig. 1. A sketch of the metal refining process

Another important consideration is reduced batch time. A second application on a twin reactor is presently being developed. The system is implemented using Cybernetica's NMPC platform.

Two additional applications will be presented in some more detail.

2.1   Manganese Metal Refining

This case is a metal refining process for removing carbon from manganese metal. The process is sketched in Figure 1. It consists of a ladle which is filled with liquid-phase high-carbon manganese metal. This implies that about 7% of the metal bath consists of carbon. In addition there is some iron and MnO in the metal bath. Carbon is removed by blowing O_2 into the ladle. The main overall reaction is

$$C + O_2 \rightarrow CO_2$$
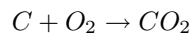
The refining process produces different products with a carbon content in the range $0.5\% - 1.5\%$. Downstream the refining process the metal is casted, crushed and screened before it is packed and shipped to customers.

In addition to the main reaction there are intermediate reactions as well as side reactions. One important side reaction is evaporation of manganese metal.

$$Mn_{(l)} \rightarrow Mn_{(g)}$$
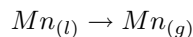
Fumes generated during the batch are collectected in an off-gas system and routed to a filter-system for removing dust.
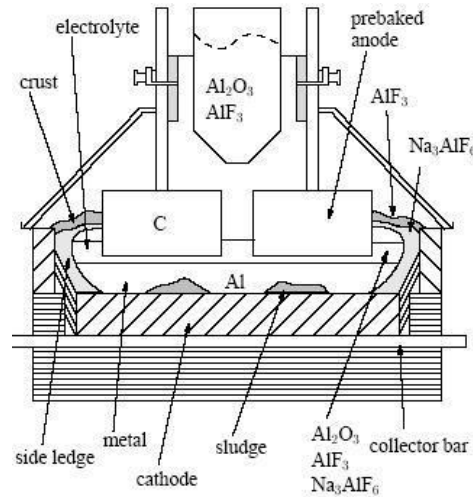
Fig. 2. Sketch of an aluminum electrolysis cell

The metal refining process is operated as a fed-batch process. The length of a batch sequence is in the order of $1-2$ hours. The same ladle is used from one batch to the next. It is replaced when the inner lining becomes too thin. The economic insentive for improved control is minimizing metal loss due to evaporation, while satisfying an upper limit on the concentration of carbon at the end of the batch. The control problem is challenging since this batch process is highly nonlinear and operational constraints are critical. An NMPC application was implemented in 2003 using Cybernetica's NMPC platform. More information on Mn decarburation can be found in [3].

2.2   Aluminum Electrolysis Cell

The Hall-Heroult process - a continuous process - is dominating worldwide in the production of aluminum [5]. The fundamentals of the process are to dissolve Al_2O_3 in molten cryolite, and electrically reduce complex aluminum containing ions to pure aluminum. The overall electro-chemical reaction in the electrolyte is

$$]2Al_2O_3 + 3C \longrightarrow 4Al + 3CO_2$$
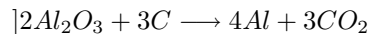
where carbon is fed to the reaction as consumable anodes. By the use of various additives, in particular AlF_3, the operating temperature of the electrolyte can be lowered from $1010C$ to approximately $960C$. Both decreased temperature and increased excess AlF_3 is believed to be beneficial for the current efficiency and the energy consumption. As molten cryolite is very corrosive, the only component of an acceptable cost presently capable of coexisting with it over time is frozen cryolite. It is therefore necessary to maintain a layer of frozen cryolite (side ledge) to prevent the carbon walls from eroding. In order to maintain the side ledge

there has to be a substantial heat loss through the side ledge and the carbon walls of the cell.

The cell voltage applied is typically $4.5V$, and the electric current through the cell is typically $150 - 200kA$. A sketch of a cell is shown in Figure 2. In real life 100-200 cells are placed and connected in series.

There are three control inputs to the process, anode beam adjustments (controlling energy input), addition of AlF_3 and addition of Al_2O_3, and three controlled variables, bath temperature, concentration of AlF_3 and concentration of Al_2O_3. A cell is regularly perturbed since liquid aluminium is tapped and one of the anode blocks is changed on a daily basis. This induces severe disturbances in the energy balance, and it implies that the operating conditions will vary significantly and hence provoke nonlinear cell effects. The process has strong internal couplings, for instance between the mass and energy balance through the side ledge. Control of an aluminum electrolysis cell is a challenging problem [2], particularly as cell capacity increases beyond $200kA$. An NMPC application is presently under development.

3   Modeling

A common denominator for the applications referenced above is the use of a non-linear first principles model. Empirical models have traditionally dominated the arena of MPC applications. A meaningful question would therefore be: "Why use mechanistic models?" The answer can be divided into two parts, and the first part can be found outside the model itself. In some of the applications a mechanistic model existed prior to the NMPC project initiation, and substantial resources had been used to develop and validate such a model. The fact that a model existed and considerable resources had been spent on development and validation are important reasons for extending the use of a model. Further, added use increases the odds for long-term survival of a model within a company. In the aluminum electrolysis case a model which had been developed over several years existed. This model had to be adjusted, only slightly however, to fit the NMPC application. No dynamic model existed prior to the NMPC project for the Mn metal refining reactor. The company, however, viewed the development of a mechanistic model in itself as important since such a model acts as a knowledge repository. One consequence of this was that the model development and validation phase was run as a joint activity between Cybernetica and the customer.

Second, in our experience a major advantage of first principles models is the reduced need for plant experimentation. As a matter of fact in the bulk of the above cases models have been developed purely on the basis of data from regular operation, ie. no dedicated measurement campaigns have been necessary. This implies that model structures have been selected, parameters have been estimated and models have been validated without resorting to often costly plant experiments. To further substantiate this, personnel from Borealis state "The model can be identified without doing plant experiments at all" [1] when

discussing their proprietary NMPC technology and its application to polyole-fine reactors. The need for reduced plant experimentation becomes particularly apparent when applying NMPC to several similar reactors as was the case for the suspension PVC polymerization reactors and the phenol-formaldehyde batch polymerization reactors referenced earlier.

Despite the above, the modeling process is a complex task and may be a bot-tleneck in the development of advanced computer-based applications, see eg. [4]. Without delving into this issue it should be noted that the process data used for offline tuning of parameters must, in some cases, include supplementary informa-tion to apply them in an appropriate manner. The reason for this is the fact that the process data does not necessarily contain sufficient information to uniquely define the process conditions applicable to the model. Examples of this may be changes in low level instrumentation and control loops due to maintenance, or merely the fact that different shifts use different operational strategies. One shift may for instance prefer to run a low level feeder control loop in manual as opposed the others running the same loop in automatic mode. The implication of the need for added information is that close ties between developers and key process personnel is important for efficient model development, in particular to swiftly provide the additional information if and when necessary.

In addition to prediction accuracy a model used for optimization-based con-trol should be smooth to facilitate the search algorithm for solving the online constrained nonlinear optimization problem. Hence, it is not adequate to develop a model with good prediction accuracy. The model should also be smooth with respect to the control inputs eligible for optimization. Our experience is that this has been a key issue to obtain robust and computationally efficient performance of the optimization algorithm both in the reference cases on metal refining and in the suspension PVC-application.

To elaborate on the metal refining case the basic kinetics models and thermo-dynamics are non-smooth. The non-smooth function were changed by applying sigmoid-functions. To illustrate assume the following kinetic model for the reac-tion rate r for the reaction $B \rightarrow A$.

$$r = \begin{cases} a(p_B - p_{equil}) \text{ if } p_B > p_{equil} \\ 0 \qquad\qquad\quad \text{ if } p_B \leq p_{equil} \end{cases}$$

$a > 0$ is some constant, p_B is the partial pressure of (gas) component B, and p_{equil} is the equilibrium partial pressure. A smooth approximate model for the reaction rate, which, however, does allow negative reaction rates, is

$$r = h(p_B, p_{equil}) \cdot [a(p_B - p_{equil})]$$
$$\text{where } h(p_B, p_{equil}) = \frac{1}{1 + e^{-\alpha(p_B - p_{equil})}}, \ \alpha > 0$$

4 State Estimation

The aim of the state estimator is to provide a robust and reliable estimate of the current states at all times. This is a challenging problem since the process data vector in our experience often is dynamic in the sense that data points routinely are biased, delayed or even missing. Examples of this are delayed or missing measurements at the startup of the metal refining batch, and delayed temparature and composition measurements in the aluminum electrolysis cell. The former typically happens on a sporadic basis while as the latter occurs regularly. The time delays in the process data from the aluminum electrolysis cell, however, may vary significantly from one sample to another.

The dynamic process data vector constitutes a challenge since an application requires a robust and reliable estimate of the current states at all times. We have applied two methods for state estimation, an extended and augmented Kalman filter (EAKF) with some modifications and recently a moving horizon estimator (MHE), see eg. [10]. The estimation software includes handling of asynchronous measurements with arbitrary sampling intervals and varying measurement delays.

It is our experience the Kalman filter has proved to work very well in several demanding applications, even if this simple estimation algorithm provides a crude approximative solution to the underlying nonlinear stochastic estimation problem. The performance of the EAKF for a specific application depends, however, crucially on the modelling of the stochastic process disturbances and on the choice of which model parameters to estimate recursively in addition to the model states. In Kalman filtering the process disturbances are modelled as filtered white noise, and this disturbance model should reflect how the true process disturbances and uncertainties are anticipated to influence the real process. Special attention should be directed towards fulfilling basic mass and energy balance requirements. It is, however, a shortcoming of the Kalman filter that these balances will generally not be exactly fulfilled even if the process disturbances are properly modelled. This is due to the linearization approximations involved in the calculation of model state updates from measurement prediction deviations.

The choice of which parameters to estimate in the EAKF should be guided by an identifiability analysis. Usually we cannot assume that the process excitations fulfil certain persistency requirements in order to ensure convergence of parameter estimates. Hence, we normally choose a set of parameters which is identifiable from stationary data, and which do not require any particular excitations in order to obtain convergence. By carefully selecting the set of model parameters to estimate, we can usually obtain zero steady-state deviations in measurement predictions.

The MHE has several advantages compared to the Kalman filter. Evident advantages are the ability to handle varying measurement delays as well as constraints in a consistent manner. As mentioned above varying delays occur in some of our applications. The ability to include constraints is also important since a nonlinear mechanistic model by definition includes physically related states and parameters, variables which often can be limited by a lower and upper bound.

Other advantages of the MHE are related to the increased accuracy in solving the underlying stochastic estimation problem. The stochastic estimation problem is solved exactly over the length of the horizon. Hence, as the estimation horizon increases towards infinity, the MHE estimate approaches the true solution of the underlying nonlinear estimation problem.

The main disadvantage with the MHE as compared to the Kalman filter is the increased computational requirements for the MHE. The estimation of stochastic process disturbances over a long estimation horizon, when compared to the length of the sampling interval, may lead to a nonlinear programming problem of untractable size. Hence, in practical applications it might be necessary to restrict the length of the horizon or to parameterize the process disturbances with a limited number of parameters over the estimation horizon. These modifications will generally reduce the accuracy of the MHE.

The bulk of our experience is based on the use of the EAKF. Because of the advantages of the MHE scheme, despite its drawback from a computational point of view, we foresee a shift towards this estimation scheme in future applications.

The state estimator in itself often provides interesting information about the process conditions. Hence, commissioning the state estimator, assuming that it provides reliable estimates, before the actual NMPC application is in our experience a favourable option. This provides at least three positive effects. First, the state estimator, a critical component of the NMPC application, is tested in its real environment. Such an environment will always provide some challenges not present in a testing environment. Second, the state estimates may provide important information to plant personnel. This is for instance the case for the aluminum electrolysis cell where estimates of internal cell states are highly interesting. Finally, the estimator builds trust and interest in the future NMPC application.

5   Control Formulation and Online Optimization

Formulating the control problem, ie. the online optimization problem, tends to be simpler than the modeling and estimation tasks described above. The online problem for an NMPC application does not in principle differ from the linear MPC case. The objective function will in some sense be related to economic conditions. For a batch reactor, in which batch capacity limits production, minimizing the batch time is in most cases equivalent to optimizing an economic criterion. This was the case both for the metal refining case, the PVC polymerization reactors and the phenol-formaldehyde batch polymerization process. The constraints will limit variables linked to safety and quality. A typical safety constraint is the net cooling capacity in the (exothermic) PVC polymerization reactors while the end point carbon content is an important quality constraint in the metal refining reactor. The choice of control inputs and controlled outputs is again a problem where the issues in linear MPC and NMPC are similar and will hence not be discussed further herein.

Online optimization in NMPC applications is a very different issue than the convex QP-problem normally encountered in linear MPC. A robust and reliable algorithm is critical. In our experience such an algorithm can only be developed by merging insight into nonlinear programming techniques with extensive trials. The NMPC algorithm is based on the Newton-type algorithm developed by Biegler and co-workers [8]. Their approach is to linearize a nonlinear state space model around a nominal trajectory determined by the input sequence computed at the previous sampling time. A new input sequence is computed by solving a quadratic program, once over the time horizon, followed by a line search where the quadratic optimization criterion is computed based on the nonlinear model. Through the action of the line search, global convergence of the method is enforced as long as the objective function exhibits descent directions. Sufficient conditions for global convergence and stability are developed by Li and Biegler [6]. Their development assumes that the states are available, hence state estimation is not considered in the referenced paper. The algorithm we use extends and modifies the Newton-type algorithm proposed by Biegler in several ways. The algorithm is based on a linearization of the nonlinear model around nominal input and output trajectories, which are computed at each time step. The linearization is usually performed once at each time step. The optimization criterion is quadratic and the constraints are linear in the process outputs and inputs. The outputs are, however, arbitrary nonlinear functions of the states and the inputs. Another extension of Biegler's algorithm includes more flexible parameterizations of inputs and outputs; each input and output variable is parameterized independently.

Input constraints are hard constraints in the optimization. Output constraints are handled as soft exact penalty type constraints as outlined by Oliveira and Biegler [7].

6   Implementation

Putting NMPC into industrial use requires competence and systems beyond NMPC theory and the algorithms themselves. This includes a project development plan which does not differ from a typical project plan for implementing other advanced controllers. A project will include a functional design specification task which describes the functionality and details the specifications for the delivery. Thereafter the application is developed, integrated into the existing control system, and finally formally accepted by the customer through a Site Acceptance Test.

A difference in the development of applications based on first principles models compared to data driven models is that the time spent at the plant, doing process experiments and application commissioning and testing, is shortened by the use of mechanistic models. The reason is that less experiments are required for the model and state estimation development and tuning. Usually, we have on-line secure internet connection to the application computer. Then the estimator can

be tested on-line and also the NMPC application can be tested in open loop before closed loop testing.

In the Cybernetica system an NMPC application will consist of one component, a model component, which is developed specifically for the type of process unit or process section to be controlled. The other parts of the system, such as the EAKF, MHE and NMPC algorithms as well as configuration interface and communication interface to the basic control system, consist of generic components. Usually, the operator interface is integrated into the same operator interface system as used by the basic control system. Hence, from the operators point of view, the introduction of an NMPC application is seen as an extension of the existing control system.

A close dialogue with key personnel at the plant has been critical in many of our reference projects. There are several reasons for this. First, as remarked earlier, personnel have access to information and they possess knowledge which is vital in specifying an application and in the model development stage. Second, the operating personnel may have the privilege to choose between an existing operating strategy as an alternative to a (new) NMPC application. In such a situation it is important that the operators understand the application so as to gain confidence in the new application. Further, insight and motivation definitely helps in a situation where an application needs to be modified due to some unforeseen problems.

In most of the reference cases the NMPC application replaces an existing application. Usually it is necessary to upgrade other parts of the system in conjunction with an NMPC project simply because the NMPC application normally requires more accurate information than what was the case prior to its installation. Improvements typically include upgrading of instrumentation and data collection routines, and retuning of low-level control loops.

7   Conclusions

This paper discusses issues that arise when implementing NMPC in the sense of a nonlinear mechanistic model, state estimation, and the solution of an on-line constrained nonlinear optimisation problem. Even though this technology presently is in the development stage several demanding applications have been developed with good industrial acceptance.

References

[1] K. S. Andersen, M. Hillestad and T. Koskelainen, Model Predictive Control of a BORSTAR Polyethylene Process, *In: 1st European Conference on the Reaction Engineering of Polyolefins (ECOREP)*, France, 2000.

[2] T. Drengstig, D. Ljungquist and B. A. Foss, "On the AlF_3 and Temperature Control of an Aluminum Electrolysis Cell," *IEEE Trans. of Control System Technology,* vol. 6, no. 2, pp. 157–171, 1998.

[3] W. Dresler, "Limitations to carburization of high carbon ferromanganese", *In: Proceedings 1989 Steelmaking conference*, 1989.

[4] B. A. Foss, B. Lohmann and W. Marquardt, "A field study of the industrial modeling process," *Journal of Process Control,* vol. 8, no. 5-6, pp. 325–338, 1997.

[5] K. Grjotheim and H. Kvande, eds., "Introduction to aluminium electrolysis". *Aluminum-Verlag*, 1993.

[6] W. C. Li and L. T. Biegler, "Multistep, Newton-type control strategies for constrained nonliner processes", *Chem. Eng. Res. Des.*, vol.67, pp. 562-577, 1989.

[7] N. M. C. de Oliveria and L. T. Biegler, "Constraint handling and stability properties of model-predictive control," *AIChE Journal*, vol. 40, pp. 1138-1155, 1994.

[8] N. M. C. de Oliveria and L. T. Biegler, "A finite-difference method for linearization in nonlinear estimation algorithms", *Automatica*, vol.31, pp. 281-286, 1995.

[9] S. J. Qin and T. A. Badgwell, "An Overview of Nonlinear Model Predictive Control", *In: Nonlinear Model Predictive Control*, F. Allgower and A. Zheng Birkhauser, eds., 2000.

[10] C. V. Rao and J. B. Rawlings, "Nonlinear Moving Horizon Estimation", *In: Nonlinear Model Predictive Control*, F. Allgower and A. Zheng Birkhauser, eds., 2000.