NONLINEAR MPC COMPUTATIONAL REDUCTION FOR REAL-TIME CONTROL APPLICATIONS

John D. Hedengren and Thomas F. Edgar

Department of Chemical Engineering

University of Texas at Austin

Prepared for Presentation at the 2003 Annual Meeting, San Francisco, CA

Date 4/2003

Unpublished

Abstract

This paper proposes a method to implement nonlinear model predictive control (NMPC) in real-time control applications. Nonlinear model identification is generally seen as a major obstacle to implementing NMPC. However, once an accurate nonlinear model is identified the computational effort is often too great to implement the model in a real-time application. The approach in this paper is a two step process, model reduction followed by computational reduction. Model reduction is accomplished by computing balanced empirical gramians. Computational reduction is accomplished by using the method of *in-situ* adaptive tabulation (ISAT). ISAT was previously developed for computational reduction of turbulent flame direct numerical simulations (DNS) and is extended to the sequential NMPC framework. A case study is performed with a binary distillation column model with 32 states. By computing balanced empirical gramians the number of states is reduced to 5. NMPC with ISAT approaches the accuracy of the full state NMPC with the speed of linear MPC. In addition, the ISAT method can be used with hard or soft constraints and ODE or DAE models.

1. Introduction

In model predictive control (MPC) there is a tradeoff between computation speed and accuracy of the model. While many processes are better characterized by nonlinear models, the effort to implement nonlinear MPC (NMPC) in real-time process control is often prohibitive (Marquardt, 2001). The objective of this project is to develop a combined model and computational reduction approach to make NMPC computationally feasible for real-time applications. Comparisons are made with the speed and accuracy of linear MPC (LMPC).

2. Model Reduction

Recent efforts have focused on nonlinear model reduction. In particular, nonlinear model reduction using balancing of empirical gramians has proven effective (Hahn and Edgar, 2002). This method reduces the nonlinear model to a variable subspace that captures the most important dynamics of input/output behavior. One system studied by Hahn and Edgar is a binary distillation column with the reflux ratio (u) as the manipulated variable and distillate composition (x_1) as the controlled variable. The simulated column contains 30 trays, a reboiler, and condenser. The 32 states are the compositions of the liquid at each stage. The ODE model is placed in the general nonlinear form of Equation 1.

$$\begin{aligned} \dot{x} &= f(x,u) \\ y &= h(x) \end{aligned} \tag{1}$$

A similarity transform (T) is computed from the balancing of empirical gramians. The transformed variables are in order from most important to the least important for input/output behavior. The transformed system is shown in Equation 2.

$$\begin{aligned} \dot{\bar{x}} &= T^{-1} f(Tx,u) \\ y &= T^{-1} h(Tx) \end{aligned} \quad \text{or} \quad \begin{aligned} \dot{\bar{x}} &= \bar{f}(\bar{x},u) \\ y &= \bar{h}(\bar{x}) \end{aligned} \tag{2}$$

Hahn and Edgar showed that a reduced system with 3 transformed variables shows excellent agreement with the full 32 state model on step tests. The first 3 transformed variables are shown in Equation 3.

$$\begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{bmatrix} = \begin{bmatrix} 9.1 & \cdots & 0.015 \\ 49.5 & \cdots & -0.060 \\ -4.9 & \cdots & -0.202 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_{32} \end{bmatrix} \tag{3}$$

The relative importance of the first state (reflux drum composition) on input/output behavior is much greater than the last state (reboiler composition). This relative importance is shown by the magnitude of the numbers in the first column compared with the last column. The similarity transform from the balancing of empirical gramians confirms that most of the important dynamics for control are found in the states at the top of the column. With the transformation, the nonlinear system can be reduced by truncation or residualization as shown in Table 1.

Reduction Method	Reduced Model
Truncation	$\begin{bmatrix} \dot{\bar{x}}_1 \\ \dot{\bar{x}}_2 \\ \dot{\bar{x}}_3 \\ \dot{\bar{x}}_4 \\ \vdots \\ \dot{\bar{x}}_{32} \end{bmatrix} = \begin{bmatrix} \bar{f}_1(\bar{x},u) \\ \bar{f}_2(\bar{x},u) \\ \bar{f}_3(\bar{x},u) \\ 0 \\ \vdots \\ 0 \end{bmatrix}$
Residualization	$\begin{bmatrix} \dot{\bar{x}}_1 \\ \dot{\bar{x}}_2 \\ \dot{\bar{x}}_3 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \bar{f}_1(\bar{x},u) \\ \bar{f}_2(\bar{x},u) \\ \bar{f}_3(\bar{x},u) \\ \bar{f}_4(\bar{x},u) \\ \vdots \\ \bar{f}_{32}(\bar{x},u) \end{bmatrix}$

Table 1: Truncation vs. Residualization

By reducing the number of dynamic variables through truncation, the computational requirements are reduced by ~ 40%.

3. Computational Reduction

Computational reduction has a different focus than model reduction. In computational reduction, the goal is to retain all of the dynamics of the original model while substantially lowering the computational costs. An example of computational reduction is the method of *In-Situ* Adaptive Tabulation (ISAT) (Pope, 1997). In performing turbulent flame simulations, Pope demonstrated a speed-up factor of 1000 times using ISAT.

Analogies to the ISAT method exist in many different industries and products. For example, computer systems are built with multi-layers of caching. One of the reasons that Pentium® processors are considered superior to Celeron® processors is the larger amount of cache. This cache stores and accesses frequently computed instructions and thereby improves the processor performance. As another example, the computer can speed-up the effective download speed for internet connections by storing web pages on the hard disk. When a web site is visited again, the page can be loaded from the much faster hard disk. The common characteristics of these speed-up technologies are:

1. The first time through there is no speed-up. In fact, there may be some slow-down associated with building and storing the database.
2. Search time is generally fast compared to repeating the operation.
3. Storage costs for the database are low compared to the cost of repeating the operations.
4. There is a sufficient probability that the operation will be repeated, otherwise the database would serve mainly as an archive.
5. The system performance increases as the database matures and more operations are repeated.

In this work, the ISAT method from combustion simulations is applied to NMPC. An important assumption for ISAT is that nearby integrations will likely be repeated. For a single step test or impulse response, very few integrations will likely be repeated and ISAT would likely show poor performance. In control applications, similar disturbances to the system can occur frequently or step changes to switch between product grades can happen regularly. The ISAT method is a framework for the controller to access data from previous MPC calculations.

3.1 Database Searching

A vector ϕ_o defines the initial state of the distillation column given by Equation 4 where u is the reflux ratio.

$$\phi_o = \begin{bmatrix} u \\ \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{bmatrix} \tag{4}$$

Before integrating to obtain ϕ_f, a search is performed to obtain a record close to ϕ_o. If a record is found that is sufficiently close to ϕ_o then ϕ_f can be approximated with the nearby integration that has been previously computed and stored (see Figure 1).

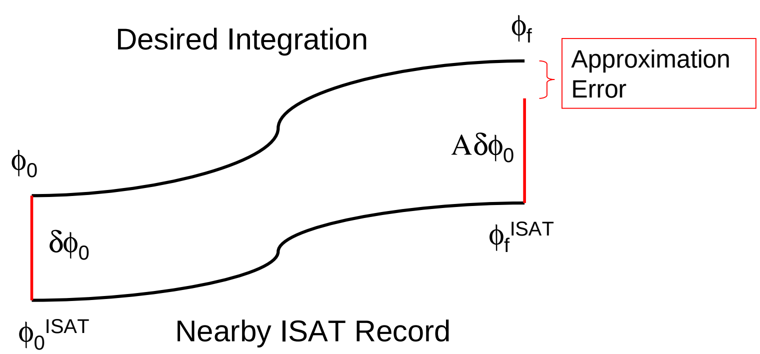

Figure 1: ISAT Search

When $\delta\phi_o$ is within an initial state region of accuracy, the nearby ISAT record ($\phi_f^{approx} = \phi_f^{ISAT} + A\delta\phi_o$) is accepted as an approximation to ϕ_f, where A is a matrix of state sensitivities.

When $\delta\phi_o$ is outside of the region of accuracy the region of accuracy is either grown or a new record is added to the database.

As the database of integrations is built, the time required to find a close record grows. Sequential searching requires O(N) operations where N is the number of records stored in the database. An alternative to sequential searching is through a binary tree structure that requires $O(\log_2(N))$ operations. The binary tree has a structure shown in Figure 2.

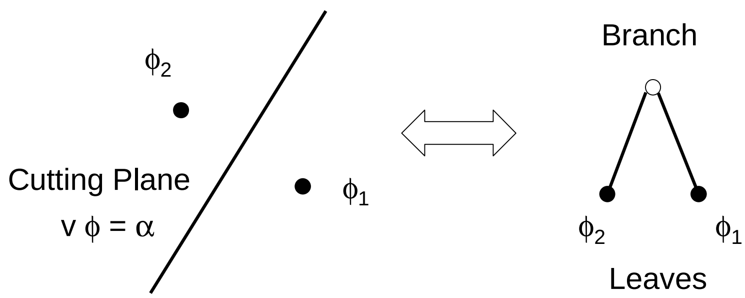

Figure 2: Binary Tree Structure

The cutting plane is defined by the vector v and scalar α (Equation 5).

$$\begin{aligned} v &= \phi_2 - \phi_1 \\ \alpha &= v^T\left(\frac{\phi_2 + \phi_1}{2}\right) \end{aligned} \tag{5}$$

When $v^T\phi_o$ is less than α, ϕ_o is closer to ϕ_1 and when $v^T\phi_o$ is greater than α, ϕ_2 should be chosen as the closer record. If ϕ_o is closer to ϕ_1, but still not within the specified tolerance then ϕ_o is integrated to ϕ_f and the tree is grown (see Figure 3).

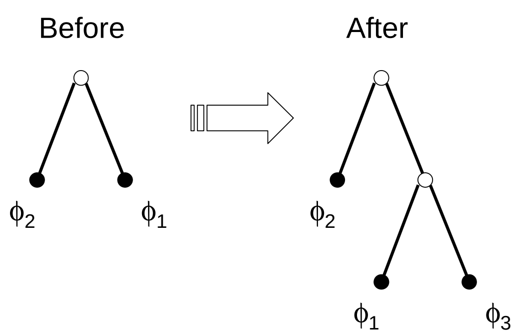

Figure 3: Binary Tree Growth

One of the limitations of the ISAT method is that the storage requirements are proportionally to n^2 where n is the total number of states. Therefore, models with a reduced number of variables are better suited to computational reduction through ISAT.

4. Combined Model and Computational Reduction for NMPC

Combining model reduction through empirical gramians and computational reduction through ISAT exploits the strengths of both methods. Generally, the model reduction step decreases the number of dynamic variables but does not have sufficient computational speed-up. Computational reduction is more effective with a low number of dynamic variables and can reduce the computational time significantly. Figure 4 provides an overview of the combined approach.

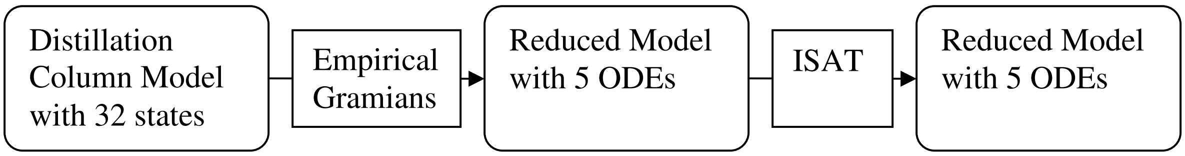

Figure 4: Combining Model Reduction and Computational Reduction

A case study has been performed with the distillation column model comparing NMPC/LMPC for the following models:

1. Nonlinear reduced model with 5 dynamic states and ISAT
2. Nonlinear reduced model with 5 dynamic states
3. Nonlinear model with 32 states
4. Linear model with 32 states

For 1 and 2 the full state model with 32 dynamic variables is reduced through truncation down to 5 dynamic variables. Using the sequential approach to MPC, the distillation column models are integrated multiple times in order to find optimal control moves that minimize a quadratic objective cost function.

Certain operational, safety, or economic constraints must be considered when developing control solutions to real systems. These limitations can be implemented as either hard or soft constraints in the MPC framework. Soft constraints are costs added to the objective function. In the author's opinion, soft constraints are the more intuitive method because the solver can choose to violate a constraint if the economic performance of the entire plant will be improved. In addition, the relative importance of each soft constraint is automatically considered. Hard constraints may be more desirable for some situations such as when safety is a concern. In this sequential NMPC approach, hard constraints can be implemented on the dependent variables.

The nominal operating point for the reflux ratio is 3. Soft constraints limit the operating region to between 2 and 4. The reflux ratio (manipulated variable) is adjusted every 5 minutes. The control horizon is 10 minutes and the prediction horizon is 15 minutes. Figure 5 shows the closed loop responses.

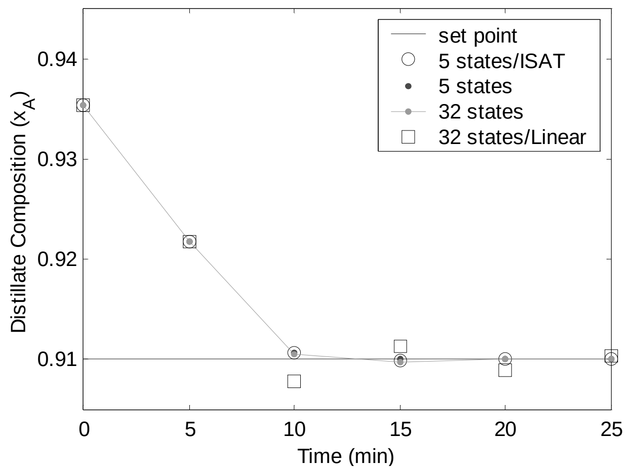

Figure 5: Closed Loop Response

During the first control move, all MPC results are at the reflux ratio lower bound of 2. Figure 6 shows the speed-up factor (compared to 32 state NMPC) for the 5 optimization steps of Figure 5.

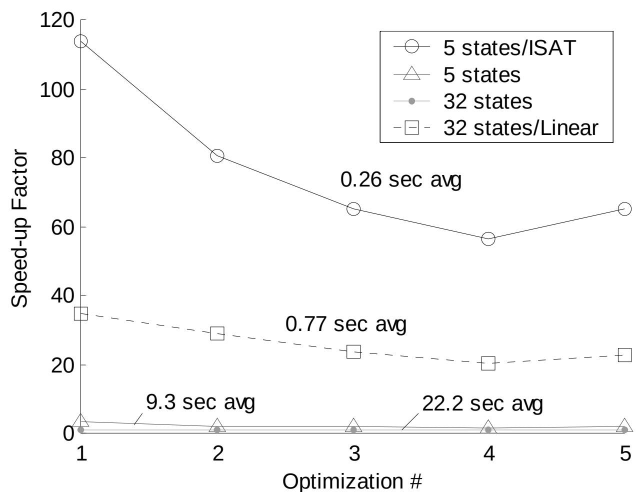

Figure 6: Speed-up Factor

The cpu times shown on the graph are from computations on a 2 GHz Celeron processor. This case study shows that ISAT can exhibit significant computational reduction while preserving the accuracy of the nonlinear model.

4. Discussion

The objective of this project is to make NMPC feasible for real-time applications. In the first step, the dynamic degrees of freedom are reduced by the balancing of empirical gramians. Next, the ISAT method is applied to reduce the computational time. The ISAT method shows computational times comparable to LMPC with the accuracy of NMPC.

Although applied with a model reduced through balanced empirical gramians, ISAT for NMPC can be used with any model reduction technique that reduces the number of dynamic degrees of freedom. In the case where the model already has a low number of variables, ISAT can be applied directly without a model reduction step. In addition, the ISAT approach can be applied to ODE or DAE models. The ISAT technique also works with hard or soft constraints.

References

Hahn, J. and T. F. Edgar. An Improved Method for Nonlinear Model Reduction Using Balancing of Empirical Gramians. *Computers and Chemical Engineering* **26**, No. 10, pp. 1379-1397 (2002).

Marquardt, W. Nonlinear Model Reduction for Optimization Based Control of Transient Chemical Processes. *Proceedings of the CPC VI*, Tucson, Arizona, pp. 30-60 (2001).

Pope, S. B. Computationally Efficient Implementation of Combustion Chemistry Using *In Situ* Adaptive Tabulation. *Combustion Theory Modeling* **1**, pp. 41-63 (1997).