In Situ Adaptive Tabulation for Real-Time Control

John D. Hedengren and Thomas F. Edgar
Department of Chemical Engineering
University of Texas at Austin
Austin, TX 78712
E-mail: edgar@che.utexas.edu

Abstract **- This paper outlines 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 and is extended to the sequential NMPC framework in this work. 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 five. With ISAT, the computational speed is 85 times faster than the original NMPC while maintaining the accuracy of the nonlinear model. Since ISAT is a storage and retrieval method, it is compared to artificial neural networks in another case study. This case study is performed with a dual CSTR model with 6 states. Open loop and closed loop step tests are performed to demonstrate the superior quality of ISAT in extrapolating outside of the training domain.**

I. 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 [1]. The objective of this research 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).

II. MODEL REDUCTION

Recent efforts have focused on nonlinear model reduction. In particular, nonlinear model reduction using balancing of empirical gramians has proven effective [2]. 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 (1).

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

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 (2).

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

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 (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 I.

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

Table I
Comparison of truncation and residualization with 3 dynamic variables remaining.

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

III. 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) [3]. 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.

A. Database Searching

A vector ϕ_o defines the initial state of the distillation column given by (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 Fig. 1).
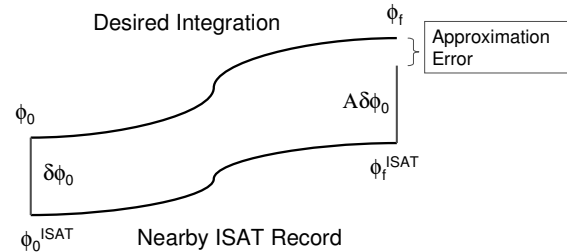

Fig. 1. Approximation of the desired integration final state with a nearby ISAT record.

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.

B. Database Growth

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 Fig. 2.
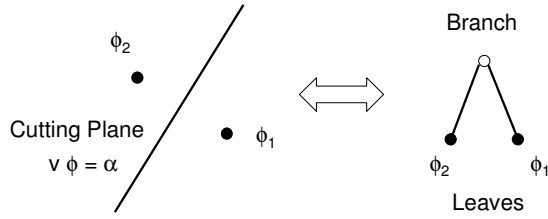

Fig. 2. An illustration of the binary tree structure in the cutting plane format (on the left) and the tree format (on the right).

The cutting plane is defined by the vector v and scalar α.

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

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 Fig. 3).
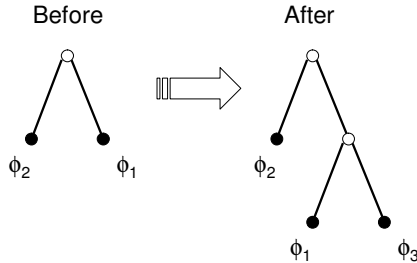

Fig. 3. Binary tree growth. A tree with one branch and two leaves is grown to include another leaf.

One of the limitations of the ISAT method is that the storage requirements are proportional to n^2 where n is the total number of states [4]. Therefore, models with a reduced number of variables are better suited to computational reduction through ISAT.

IV. 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. Fig. 4 provides an overview of the combined approach.
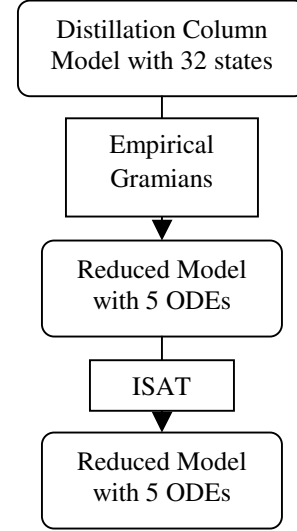

Fig. 4. Model and computational reduction flowchart

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

Models 1 and 2 are from full state model with 32 dynamic variables, 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 manipulated 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. Fig. 5 shows the closed loop responses.
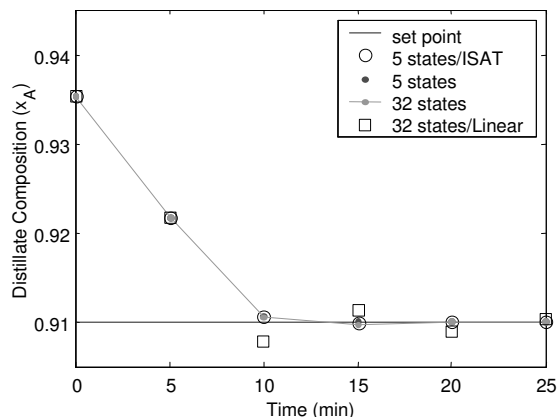

Fig. 5. Closed loop response comparison for nonlinear MPC with ISAT with 5 states, nonlinear MPC with 5 states, nonlinear MPC with 32 states, and linear MPC.

During the first control move, all MPC results are at the reflux ratio lower bound of 2.

Fig. 6 shows the speed-up factor (compared to 32 state NMPC) for the 5 optimization steps of Fig. 5.
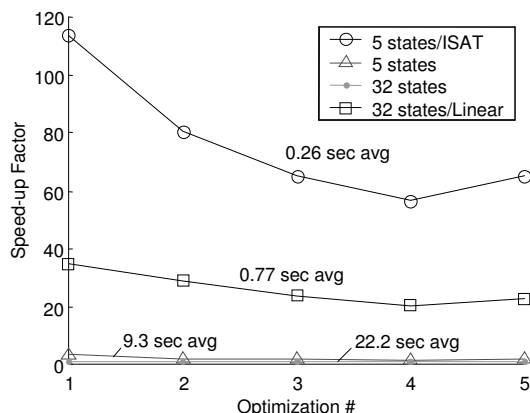

Fig. 6. Speed-up factor for each of the optimizations shown in Fig. 5. The number above each curve indicates the average optimization cpu time on a 2 GHz processor.

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.

V. ISAT vs. NEURAL NETWORKS

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.

As an illustrative example, ISAT and a neural net are compared in an open loop step test and a closed loop control application. The example model was a dual CSTR model (see Fig. 7) with 1 manipulated variable (heat addition to the first tank), 6 states, and 1 controlled variable (temperature of the second reactor). The model was developed by Hahn and Edgar [2] as a benchmark model for nonlinear model reduction.
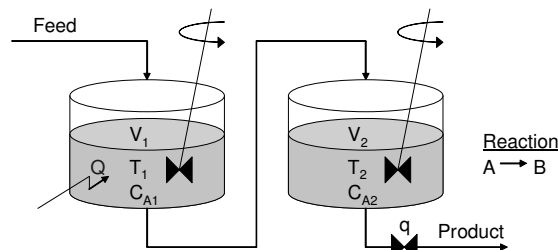

Fig. 7. Diagram of two CSTRs in series with a first order reaction. The manipulated variable is the cooling rate to the first CSTR.

The data were gathered from ISAT training. For the sake of comparison, the neural net used the ~1600 ISAT records for training. The neural net was constructed with one hidden layer of a hyperbolic tangent sigmoid transfer function and an output layer of a linear transfer function (see Fig. 8).
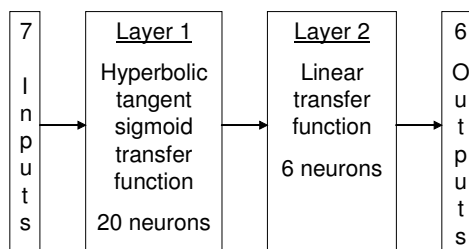

Fig. 8. Neural net with one hidden layer and one output layer. The hidden layer is a hyperbolic tangent function and the output layer is a linear function. This neural net relates 7 inputs to 6 outputs.

Before the training, the data were appropriately scaled for efficient implementation in the neural net. Fig. 9 shows a large open loop step test, one that is outside those found in the training data. In this step test, the cooling is increased to the point that the irreversible reaction is extinguished and a large temperature step results.

Up to about 5 minutes of simulated time, the neural net and ISAT perform similarly. To this point both accessed data that were within the training domain. Beyond 5 minutes ISAT is superior in agreement with the non-reduced model due to a built in error checking strategy. Before 5 minutes, the ISAT method shows that it performs mostly retrievals

indicated by the circles. Once ISAT detects large errors from retrievals, it starts adding records to the database, indicated by the squares. If the step response were to extend further in time to a new steady state, the ISAT points would again become circles, indicating database retrievals.
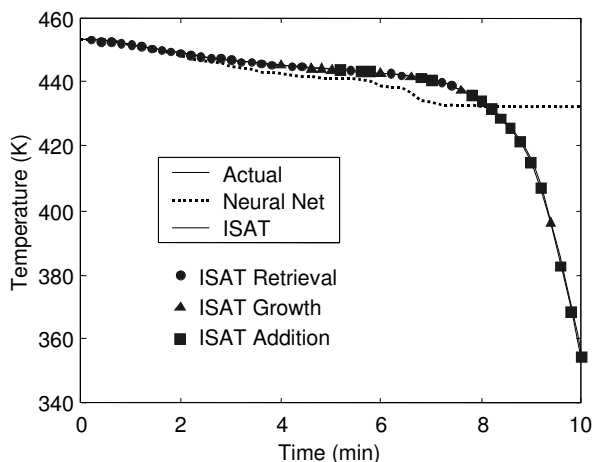

Fig. 9. Open loop step test for the dual CSTR model. The error control of ISAT indicates that additional records must be added to the database

ISAT and the neural net were compared in a closed loop simulation with a 'small' set point change inside the training domain (see Fig. 10).
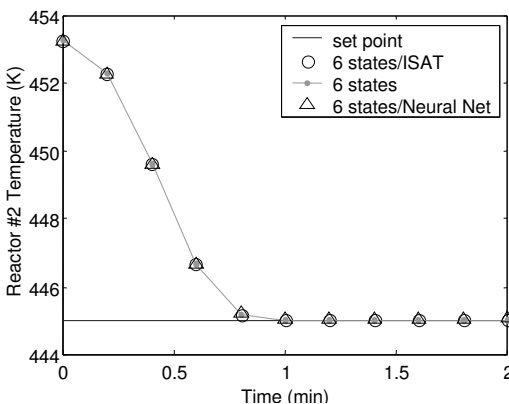

Fig. 10. Small closed loop set point change within the training domain.

All three show excellent agreement as they reach the new set point along the same trajectory. Next, a 'large' set point change was performed to access a region of state space outside of the training domain (see Fig. 11).

For this step change, the neural net controller eventually becomes unstable. This is because the neural net does not have the capability to extrapolate outside of the data that was used to train it. In this respect, the ISAT method is superior because it detects when it has gone outside of the training domain and integrates the model to generate and add new data to the training set.
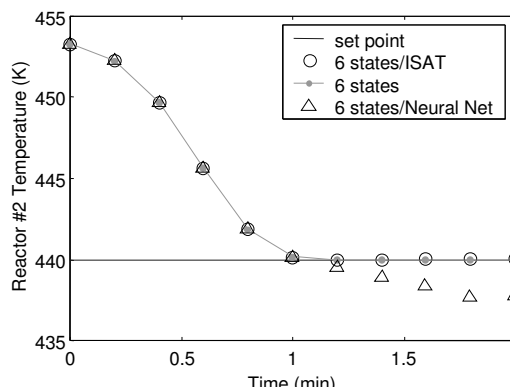

Fig. 11. Large closed loop set point change outside of the training domain.

VI. CONCLUSIONS

This paper outlines a new technique for computational reduction for NMPC. In this approach model reduction through balanced empirical gramians is followed by computational reduction through ISAT. Although previously developed for turbulent flame simulations, ISAT can be directly applied because many open loop simulations are performed to find optimal inputs to the control problem. A case study with a binary distillation column model showed a speedup of 85 over the original NMPC. Like neural nets, ISAT reduces the computational cost through storage and retrieval. Another case study with a dual CSTR showed the advantage of using ISAT over neural nets when the simulation accessed data outside the training domain.

REFERENCES

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

[2] Hahn, J. and T. F. Edgar. An Improved Method for Nonlinear Model Reduction Using Balancing of Empirical Gramians. *Computers and Chemical Engineering,* vol. 26, no. 10, pp. 1379-1397, 2002.

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

[4] Chen, J.-Y., Blasco, J. A., Fueyo, N., and C. Dopazo, An economical strategy for storage of chemical kinetics: fitting *in situ* adaptive tabulation with artificial neural networks, *Proceedings of the Combustion Institute*, vol. 28, no. 1, pp. 115-121, 2000.