Approximate Nonlinear Model Predictive Control with In Situ Adaptive Tabulation

John D. Hedengren [a] and Thomas F. Edgar [b],*

[a] *PAS, Inc., 16055 Space Center Blvd., Suite 600, Houston, TX 77062, USA*

[b] *Department of Chemical Engineering, University of Texas at Austin, Austin, TX 78712, USA*

Abstract

In situ adaptive tabulation (ISAT) is applied to store and retrieve solutions of nonlinear model predictive control (NMPC) problems. ISAT controls approximation error by adaptively building the database of NMPC solutions with piecewise linear local approximations. Unlike initial state or constraint parameterized constrained quadratic programming (QP) solutions, ISAT approximates NMPC solutions within a specified tolerance, thereby easing the dimensionality difficulties of these other techniques. In the limit as the specified tolerance is reduced to zero and for linear models with quadratic objective functions, ISAT becomes an adaptive version of state parameterized storage and retrieval of mp-QP problems.

Key words: In situ adaptive tabulation, Nonlinear model predictive control, Computational reduction

* Corresponding author. Tel.: +1 512 471 3080
  Email address: edgar@che.utexas.edu (Thomas F. Edgar).

1 Introduction

MPC is now suggested as a candidate to replace PID control due to recent developments in computational reduction of the MPC algorithm (Pannocchia et al., 2005) (Pistikopoulos et al., 2002). By computing all possible LP solutions off-line, the on-line portion is reduced to some conditional checking and simple matrix multiplications. This modification extends the potential of MPC to fast processes and simple computers (e.g. integrated circuit chips).

The contribution of this work is to adaptively perform the formerly off-line computations on-line and reduce the dimensionality of the store. By adaptively building the store, only areas accessed in practice are included in the database. With in situ adaptive tabulation (ISAT), the storage requirements are reduced by using local linear solutions to approximate a hyper-ellipsoid region where the local solution is within a specified error tolerance of the actual solution.

2 Explicit MPC (Linear Models)

The linear quadratic regulator (LQR) with a linear model and quadratic objective function is a special case of MPC without constraints. Without constraints the linear solution of the Riccati equation is optimal for all possible initial states. An on-line implementation of LQR would consist of simply multiplying the state vector by the gain matrix to obtain the optimal control vector. With constraints, the optimal solution is a piecewise affine (PWA) linear function of the initial states. The linear regions are often referred to as characteristic or critical regions (CRs). Each region is bounded by a set of constraints. When the constraint boundary is crossed, the linear solution may

2

no longer be exact. On-line retrieval of explicit MPC with constraints includes one extra step: location of the region with the correct active set of constraints. Once this region is located (via the checking of several conditions), the rest of the computation is identical to the LQR implementation.

2.1 Parameterization of Initial States

The development of multi-parametric linear programs (mp-LPs) started with the formulation of Gal and Nedoma (1975). Acevedo and Pistikopoulos (1997) extended sensitivity analysis to mixed-integer linear programming (MILP) by solving mp-LP problems. Dua and Pistikopoulos (1999) generalized the model form by developing multi-parametric analysis of mixed-integer nonlinear programming (MINLP). Bemporad et al. (2000) applied the mp-LP work to MPC applications with linear objective functions and mixed-integer models. Pistikopoulos et al. (2000) extended the theory of mp-LPs to include multiparametric quadratic programs (mp-QPs). This extension made possible the explicit LQR solution subject to constraints or in other words, explicit MPC (Bemporad et al., 2002).

Even though an exact explicit solution is possible in theory for convex problems, there were some serious implementational issues that limited applications of explicit MPC to small systems, few constraints, and short control horizons (Pistikopoulos et al., 2002). A significant effort has been exerted to reduce these limitations. Bemporad and Filippi (2001) introduced suboptimal explicit MPC. Adjacent critical regions are merged when an error tolerance can be met. Rossiter and Grieder (2005) used an interpolation scheme to reduce the storage requirements by 2-3 times and reduce the on-line computational

costs by 10 times. Johansen and Grancharova (2003) proposed a technique to logarithmically limit the on-line search times with a structured binary tree. Off-line, the regions are divided into successively smaller hypercubes until the error tolerances are met at each of the vertices. Grieder and Morari (2003) performed a complexity analysis of the on-line implementation to reduce the controller complexity by orders of magnitude at a performance cost of $<\%1$. Tondel et al. (2003) increased the efficiency of the off-line calculation by deriving a new exploration strategy for sub-dividing the parameter space. Even with all of these improvements, the largest MPC problem reported in the literature is control of a laboratory model helicopter. The problem has 6 states, 2 manipulated variables, 8 constraints, is discretized in 0.01 second segments, has a control horizon of 0.5 seconds, and 4 input parameters (Tondel et al., 2003).

2.2   Parameterization of Active Sets

In deriving state parameterized explicit MPC, the problem is transformed into a quadratic program form. In this form, Seron et al. (2000) suggested that the optimal control can be parameterized by the active set instead of current states. While Seron, et al. proposed an analytical solution, Pannocchia et al. opened the approach to non-trivial problems by creating a numerical algorithm to solve the active set parameterized problem (Pannocchia et al., 2004). Each of the constraints can either be inactive, at the lower bound, or at the upper bound. Off-line a table of all possible solutions is generated. The on-line portion consists of finding the table value that predicts non-negative Lagrange multipliers and manipulated variables (MVs) inside the constraint bounds.

4

Storage and retrieval of a constrained linear quadratic controller solution for SISO systems has been proposed to replace PID control (Pannocchia et al., 2005). Two limitations of this algorithm are (1) constraints are restricted to lower and upper bounds on the MVs and (2) problem scaling is 3^N, where N is the horizon length. The theory for MIMO systems follows by simple extension, but full enumeration of all active sets is prohibitive due to 3^{mN} scaling, where m is the number of constrained inputs.

Muske and Badgwell (2002) developed offset free control in MPC by creating input or output integrating disturbances. Pannocchia (2003) showed that an integrating disturbance must exist for every measurement to guarantee offset free control. The offset free control is included in the constrained LQ control of Pannocchia et al. (2005). Sakizlis et al. (2004) followed by incorporating offset free control into state parameterized explicit MPC.

3 Explicit NMPC (Nonlinear Models)

Because there is rarely an exact explicit solution to NMPC, all computational reduction techniques for NMPC are approximate. The effectiveness of a particular technique depends on the control of the approximation error, storage requirements, speed of the off-line algorithm, speed of the on-line algorithm, and guarantees of stability. An explicit solution of NMPC in this section refers to an explicit numerical solution through storage and retrieval of previous computed solutions. An analytic explicit solution is not attempted.

3.1 Dynamic Programming

Dynamic programming was originally proposed by Bellman to solve optimal control problems (Bellman, 1962). The goal of dynamic programming is to find an optimal cost-to-go function, which can be used to solve for an optimal trajectory of inputs as a function of initial states. Recent approaches such as sequential reinforcement learning avoid dynamic programming dimensionality problems by operating on states as they occur sequentially (Barto, 1997). Also, neuro-dynamic programming (Kaisare et al., 2003) (Bertsekas, 2001) overcomes the curse of dimensionality by approximating the cost-to-go function with a neural net. Yet another technique balances accuracy with computational speed by approximate dynamic programming (Lee and Lee, 2004). In summary, dynamic programming's curse of dimensionality has been partially remedied by algorithms that seek to reduce the storage and search times. However, applications to large scale problems are still infeasible.

3.2 Multi-Parametric NMPC

For multi-parametric analysis, suboptimal explicit MPC techniques have been developed to allow nonlinear models, nonlinear constraints, and non-quadratic objective functions. Bemporad (2003) introduced multi-parametric approximation of MINLP problems. Johansen (2002) formerly utilized an mp-QP approximation to solve the mp-NLP sub-problem but later decided to use the increased accuracy and computational expense of NLP sub-problems (Johansen, 2004). Hale and Qin (2004) take a similar approach as Johansen but use simplices instead of hypercubes to map the nonlinear surface. A predictor-

corrector method is used to obtain new points. The predictor is a linear ex-trapolation from an existing point to a new point of interest. If the active set of constraints changes, a condition is applied to find the active set boundary (Hale, 2005). The corrector uses a Newton's method type algorithm to solve the NLP that converges rapidly because of the linear predictor initialization. One drawback is poor computational scaling with increasing number of parameters (in this case, number of states), but polynomial scaling in other dimensions.

4 Approximate Nonlinear MPC

Consider the continuous-time nonlinear differential algebraic equation (DAE) system

$$0 = f(\dot{x}(t), x(t), u(t), \theta) \tag{1}$$

where $\dot{x}(t) \in \mathbb{R}^p$ is the state derivative, $x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}^m$ is the input, and $\theta \in \mathbb{R}^q$ is a set of parameters. The dimension of $\dot{x}$ is equal to that of x for ODE models. The discrete-time nonlinear DAE system can be obtained by numerically integrating Equation 1 as an initial value problem (IVP), resulting in the explicit form that is solved sequentially on a sub-node level in optimal control problems

$$x_{k+1} = f(x_k, u_k, \theta) \tag{2}$$

or by orthogonal collocation, creating an implicit form that is solved on a sub-node level simultaneously in optimal control

$$0 = f(x_{k+1}, x_k, u_k, \theta) \tag{3}$$

7

where $x_k \in \mathbb{R}^n$ and $u(t) \in \mathbb{R}^m$. The indices (k) refer to the discretized step with the current time being 0. In optimal control, when a sampling instant occurs the current time is shifted to zero.

4.1   NMPC Formulation

For the current state (x_0) and parameters (θ), a typical NMPC algorithm solves the optimization problem

$$\Phi^*(x_0, u_{-1}, \theta) = \min_{x,u} \left(\sum_{i=1}^{N} f_i(x_i, u_{i-1}, \theta) \right) \tag{4a}$$

subject to

$$x_0 \ given \tag{4b}$$

$$u_{-1} \ given \tag{4c}$$

$$0 = f(x_{k+1}, x_k, u_k, \theta) \quad k = 1, \ldots, N-1 \tag{4d}$$

$$Dx_k \leq d \quad k = 1, \ldots, N \tag{4e}$$

$$Eu_k \leq e \quad k = 0, \ldots, N-1 \tag{4f}$$

$$G\left(u_k - u_{k-1}\right) \leq g \quad k = 0, \ldots, N-1 \tag{4g}$$

where D, E, and G are matrices and d, e, and g are vectors of appropriate dimension. The quantities x and u refer to the sequence of vectors $(x_1, x_2, \ldots, x_N)$ and $(u_0, u_1, \ldots, u_{N-1})$, respectively. The optimal solution to the NMPC problem is a unique function of the current states x_0, previous input u_{-1}, and the adjustable parameters, θ. The adjustable parameters can be feed-forward or feedback model variables. An example of feedback variables are input or output integrating disturbances for offset free control. Feed-forward parameters accommodate anticipated shifts in process dynamics or multiple model switching.

After the optimal control problem is solved the first input (u_0^*) is injected into the process. At the next sampling instant, a new estimate of the current states and parameters is obtained. NMPC is often referred to as receding horizon control (RHC) because the horizon of the optimal control problem shifts as time advances. The same optimal control problem is solved at every sampling instant, deterministically dependent on the updated variables assembled in ϕ.

$$
\phi = \begin{bmatrix} x_0 \\ u_{-1} \\ \theta \end{bmatrix} \tag{5}
$$

Even though the entire trajectory of optimal inputs are solved $(u^* = \{u_0^*, u_1^*, \ldots, u_{N-1}^*\})$, the only one required for optimal control is the first input, u_0^*. The storage and retrieval of optimal control can therefore be simplified to u_0^* as a unique function of ϕ.

4.2 ISAT Approximate Control

In situ adaptive tabulation (ISAT) dynamically stores and retrieves nonlinear functions with piecewise linear approximations. ISAT is applied to store and retrieve solutions of NMPC problems, producing approximate optimal solutions bounded by a specified error tolerance. The focus of this work is two-fold: (1) dynamically generate the store of approximate solutions and (2) reduce the dimensionality of the store while limiting the error to an acceptable range. By retrieving approximate solutions, NMPC can be applied with substantially less computational resources.

The error control strategy proposed in Pope (1997) and with further details given by Hedengren and Edgar (2005) may be ineffective for problems with constraints. The constraints can lead to a non-continuously differentiable or non-continuous control moves as a function of current states, parameters, set points, etc. These discontinuities occur at the junction between active sets. This non-smooth solution surface reduces the accuracy of local linear approximations and limits the assumption about an initial estimate of the region of accuracy (ROA). An initial estimate of the ROA is necessarily set to zero volume after a database addition. This potentially causes the database to perform more growths but is a protection to improve the ability of ISAT to keep the approximation error with the specified tolerance. The initial estimate of the ROA is eliminated for nonlinear problems and restricted to the active set constraint region of constrained LQ problems.

Ellipsoid of accuracy (EOA) expansions are made only after an expanded validity check is performed. An ISAT record is expanded symmetrically with each growth. The expanded validity check examines both sides of the expansion.

In summary, ISAT is customized for nonlinear functions that are discontinuous or non-continuously differentiable. These customizations were made specifically for NMPC problems. With linear models and a quadratic objective function, the initial estimate of the ROA is exact within the active set. For these problems, the initial estimate of the ROA is taken as the active set region. The basic unit of the ISAT database is the record. An ISAT record consists of the independent variables (ϕ), the dependent variables (u_0^*), a sensitivity matrix ($\frac{\partial u_0^*}{\partial \phi}$), an ellipsoid of accuracy (EOA), and a critical region (CR) (see Table 1). The memory required to store an individual ISAT record scales with $O((n + m + q)^2)$.

Table 1

Elements of the ISAT record for NMPC storage and retrieval

ISAT Record Element	Symbol and Dimension
Independent variables	$\phi \in \mathbb{R}^{n+m+q}$
Dependent variables	$u_0^* \in \mathbb{R}^m$
Sensitivity	$A \in \mathbb{R}^{m \times (n+m+q)}$
Ellipsoid of accuracy	$M \in \mathbb{R}^{(n+m+q) \times (n+m+q)}$
Critical region (cLQ only)	$CR \in \mathbb{R}^{(n+q) \times (n+m+q)}$

The level of effort required to set up the ISAT database is variable. At a minimum, ISAT needs only the values of the independent and dependent variables. The sensitivity can be estimated by finite differencing or estimated from a store of nearby data by linear regression. An exact sensitivity is preferable and is available from the Hessian at the converged solution. Appendix A details the derivation of the exact sensitivity and defines the critical region for constrained LQ problems.

4.2.1 First Scenario: Retrieval

When ISAT receives a database request, it performs one of three scenarios. In the first scenario, the query (ϕ_{q1}) is inside a region of accuracy termed the ellipsoid of accuracy (EOA), centered about a close stored record, ϕ_s (see Figure 1). Retrievals are extremely fast because computations are limited to a binary tree search, conditional checking, and matrix-vector multiplications. When the ISAT database is mature most of the operations are retrievals.
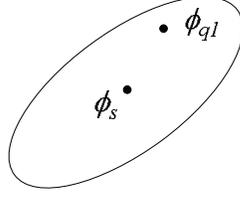
Fig. 1. A retrieval occurs when the query point (ϕ_{q1}) is within the ellipsoid of accuracy (EOA)

4.2.2   Second Scenario: EOA Growth

In the second scenario, the query is outside the EOA but inside the error tolerance for u_q^* and u_{s-q}^*. In this case, the EOA is expanded to include the tested query point (see Figure 2). For the second and third scenarios, ISAT
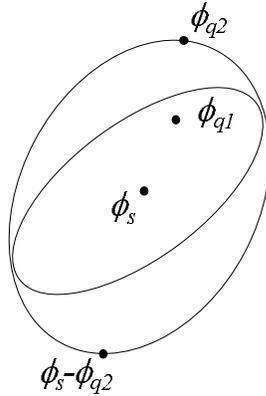

Fig. 2. The EOA is grown when the query point is outside the EOA but within the error tolerance for u_0^*

has no computational advantage over directly solving the original NLP problem on-line. If real-time requirements prohibit an on-line NLP solution, an approximation to the optimal control can be obtained by using $\hat{u}_q^*$ anyway, but no guarantees of accuracy or stability are provided.

4.2.3   Third Scenario: Addition

In the final scenario, the query is outside the EOA and outside the error tolerance for u_q^* or u_{s-q}^*. A new ISAT record is added with an initial estimate of the ROA (see Figure 3). For constrained LQ problems, the optimal control
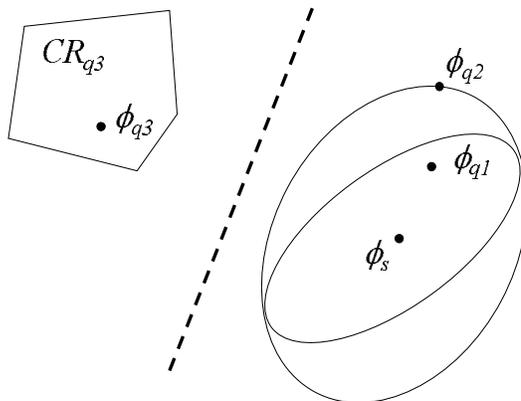

Fig. 3. A record is added to the ISAT database when query point is outside the EOA and error tolerance for u_0^*

solution is linear with respect to ϕ. Therefore, an initial estimate of the ROA is the active set state space. For nonlinear problems with constraints, there is no accuracy guarantee. In this case, the initial estimate of the ROA is a zero-volume ellipsoid centered at u_q^*.

4.3   Summary of the ISAT Algorithm

ISAT can be summarized in 13 steps. Steps 1-5 are the retrieval steps, 6-11 attempt growth of the EOA, and 12 is a database addition. The last step is to inject either u_0^* or $\hat{u}_0^*$ into the process. Retrievals produce approximate optimal control within the desired error tolerance ϵ_{tol} whereas growths and additions produce exact answers.

13

(1) locate nearby records with multiple binary tree searches

(2) compute $\hat{u}_q^* = u_s + A\left(\phi_q - \phi_s\right)$

(3) (for QP problems) if $\lambda_q \leq 0$ and $g_I(x_q, u_q, \phi) \leq 0$ go to 5

(4) if $\phi_q^T M \phi_q \leq \epsilon_{tol}$ go to 5

(5) set $\hat{u}_0^* = \hat{u}_q^*$, go to 13

(6) solve the NLP (or QP) for ϕ_q

(7) if $\left|\hat{u}_q^* - u_q^*\right| > \epsilon_{tol}$, go to 12

(8) solve the NLP (or QP) for $(2\phi_s - \phi_q)$ to get u_{2s-q}^*

(9) compute $\hat{u}_{2s-q}^* = u_s + A\left(\phi_s - \phi_q\right)$

(10) if $\left|u_{2s-q}^* - \hat{u}_{2s-q}^*\right| > \epsilon_{tol}$, go to 12

(11) grow EOA, set $u_0^* = u_q^*$, go to 13

(12) add a new record to the database with a zero volume EOA (if QP, initial ROA is given by $\lambda_q \geq 0$) and $g_I(x_q, u_q, \phi) \leq 0$

(13) inject u_0^* for optimal control or $\hat{u}_0^*$ for approximate optimal control

5   Temperature Control of an Exothermic CSTR

A simple test problem is considered to show the applicability of ISAT to storage and retrieval of optimal control. A perfectly mixed, adiabatic CSTR has an exothermic reaction of compound A transformed into compound B. Temperature control of the reactor is a challenge due to the highly exothermic reaction $(\Delta H_{rxn} = 50{,}000\ \frac{J}{mol})$. The temperature of the fluid in the jacket surrounding the CSTR is manipulated to control the temperature of the reactor fluid. The dynamics of the reactor are described by a set of ODEs generated from a mole balance on A and an energy balance on the reactor.

At a constant cooling temperature of 305 K, the reactor temperature spikes
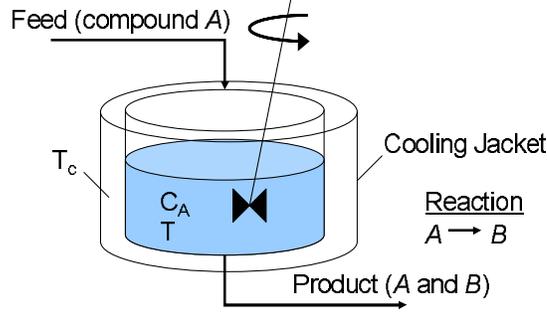
Fig. 4. Diagram of the exothermic CSTR. The two state variables reactor concentration c_A and temperature T are controlled by the jacket cooling temperature T_c continuously as the reactor goes through cycles of concentration buildup followed by moments of intense reaction (see Figure 5). The unsteady response of
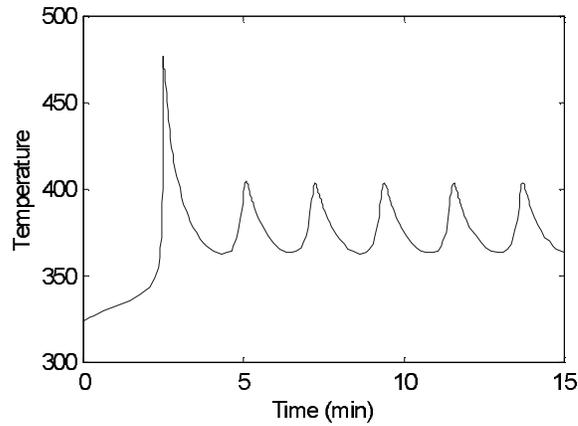

Fig. 5. Unsteady response of the reactor temperature due to moments of intense reaction followed by periods of gradual cooling.

the reactor with a constant cooling jacket temperature suggests that unsteady control may be necessary when pushing the reactor to the stability limit. A sequential direct single shooting approach to dynamic optimization is used as the control algorithm. The N-step finite horizon NMPC is given by the following NLP problem.

$$\Phi^*(x_0, T_{sp}) = \min_{x,u} \left(\sum_{i=1}^{N} (x_i - \theta)^T Q (x_i - \theta) \right) \qquad (6a)$$

15

subject to

$$x_0 \ given \tag{6b}$$

$$x_{k+1} = f(x_k, u_k) \quad k = 1 \ldots N \tag{6c}$$

$$E u_k \leq e \quad k = 0 \ldots N - 1 \tag{6d}$$

where

$$N = 40 \quad Q = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad E = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad e = \begin{bmatrix} 320 \\ -280 \end{bmatrix} \quad \theta = \begin{bmatrix} 0 \\ T_{sp} \end{bmatrix} \tag{6e}$$

Here x and u refer to the sequence of vectors $(x_1, x_2, \ldots, x_N)$ and $(u_0, u_1, \ldots, u_{N-1})$, respectively. In this problem formulation, Φ (the cost function) is quadratic in x (states) and therefore strictly convex. The source of non-linearity comes from the model function $f(x_k, u_k)$ that is solved by integrating the ODE model. With a constant reactor temperature set point, the first optimal control step u_0^* is a unique function of the current concentration and temperature of the reactor. The optimal cooling jacket temperature $(u_0^* = T_c^*)$ to drive the reactor temperature to 320 K was calculated for reactor concentrations between 0 and $1 \frac{mol}{m^3}$ and reactor temperatures between 310 and 330 K (see Figure 6). Even though the model is highly nonlinear (reaction rate depends exponentially on temperature), the optimal control surface is surprisingly linear with respect to ϕ. With clipping of the ISAT predicted value to meet the control constraints, only one ISAT record is required to store all of the optimal control solutions with an error tolerance of 1.0 K (see Table 2). A realistic control problem was set up to test ISAT for a few set point changes. The control horizon is discretized into 1 minute segments. The estimator horizon is 40 minutes and the regulator horizon is 60 minutes. The temperature is sampled every 5 seconds and includes Gaussian distribution noise with a standard deviation of 2 K.
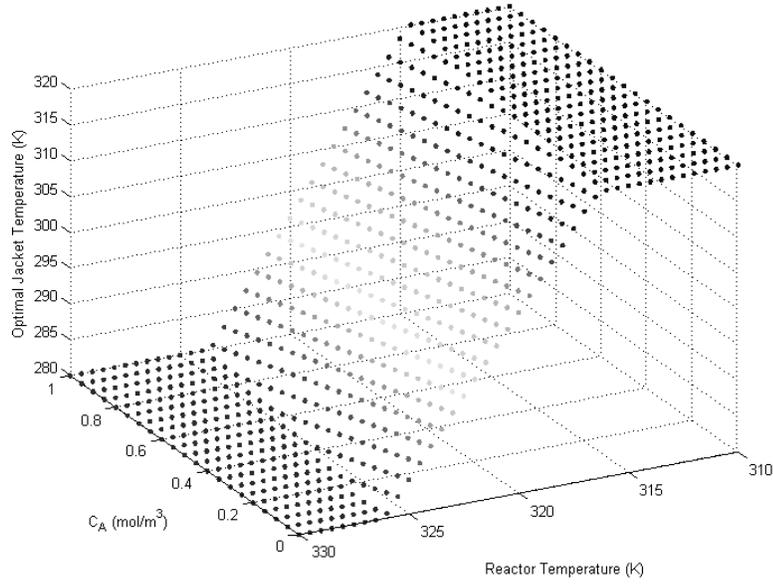
16

Fig. 6. The optimal jacket temperature (T_c^*) is a unique function of reactor concentration (c_A), reactor temperature (T), and reactor temperature set point (T_{sp}). In this figure, the set point is fixed $(T_{sp} = 320K)$ and c_A and T are varied.

Table 2

Elements of the ISAT record for the CSTR example

Element	Value
ϕ^T	$[c_A \ T \ T_{sp}] = [0.9 \ 315.0 \ 318.0]$
u_0^*	$T_c = 306.8$
A	$[\frac{\partial c_A}{\partial T_c} \ \frac{\partial T}{\partial T_c} \ \frac{\partial T_{sp}}{\partial T_c}] = [-6.227 \ -4.081 \ 4.889]$
M	$0_{3 \times 3}$
CR	N/A

Concentration is sampled every 10 seconds with a standard deviation of 0.1 $\frac{mol}{m^3}$. Plant-model mismatch is introduced by using an activation energy of the first order $(A \rightarrow B)$ reaction of 8750 $\frac{J}{mol}$ for the model and 8740 $\frac{J}{mol}$ for the

17

plant. At the first sampling time the plant state is 0.951 $\frac{mol}{m^3}$ and 312.8 K. The estimated model states are 0.9 $\frac{mol}{m^3}$ and 300 K. The initial set point is 315 K. At 50 minutes the set point changes to 300 K, an unreachable set point. At 100 minutes the set point changes to 328 K, close to the NMPC closed loop stability limit. At 150 minutes the set point changes to 308 K (see Figure 7).
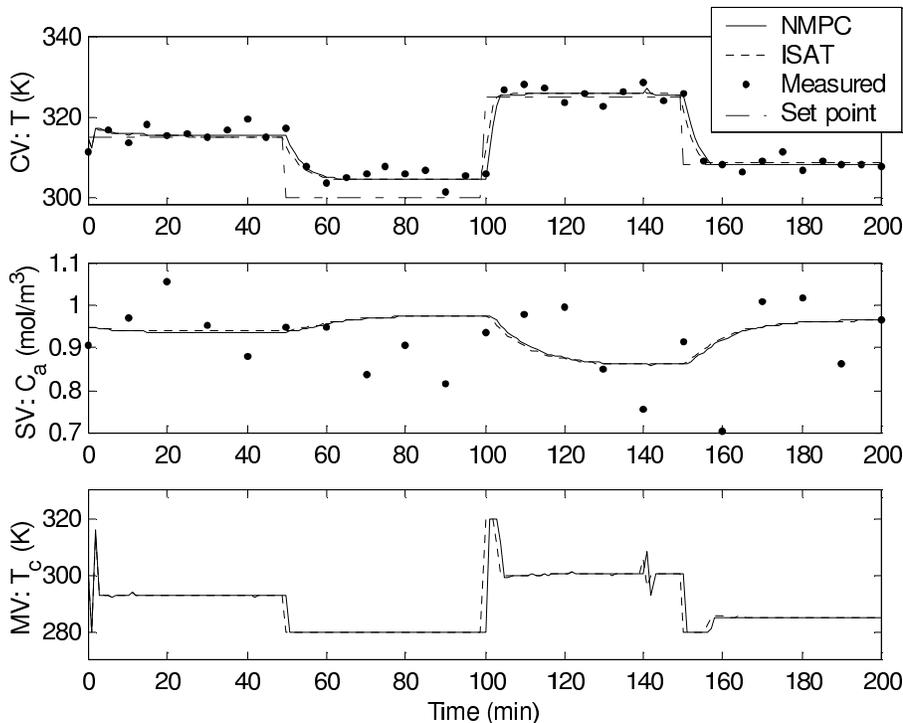

Fig. 7. Control performance of ISAT compared to NMPC. The controlled variable (CV) is the reactor temperature, the state variable (SV) is the reactor concentration, and the manipulated variable (MV) is the temperature of the cooling jacket.

While the two control performances are virtually indistinguishable, ISAT performance is actually slightly better because there is no time delay associated with computing the optimal control solution. Because ISAT operates one step ahead, it responds faster to set point changes and disturbances. This, however, is not the main advantage of using ISAT. The main advantage is that NMPC can be applied to processes with fast sampling times (<1 μsec) or simple com-

puters (IC chips). In addition, enumerating the entire control solution off-line
can reveal infeasible regions, stability limits, and other closed loop properties.
For this example, the CPU times are shown in Figure 8. NMPC consists of
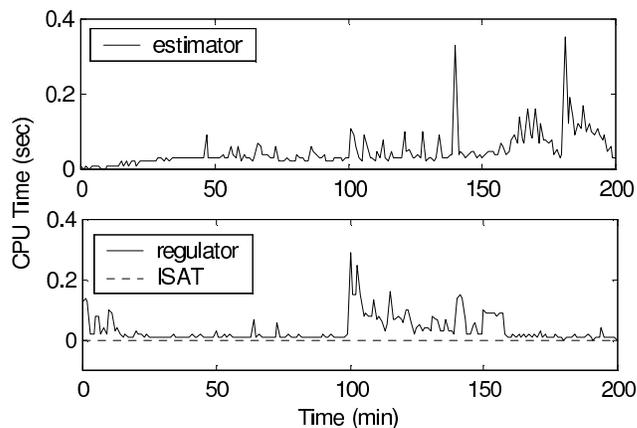

Fig. 8. Computational times of the estimator and regulator at each sampling instant.
ISAT is fast because computation is limited to a matrix multiplication.

at least two principal calculations: estimation and regulation. The estimator
and regulator calculations averaged under 0.1 seconds with maximum calcu-
lation time of about 0.3 seconds. Both the estimator and regulator NLPs were
solved with the VF13 SQP solver in FORTRAN using a direct single shoot-
ing solution approach. All calculations were performed on a 2.0 GHz Celeron
processor. The estimation problem is not reduced with ISAT in this example
although an approach has been developed for explicit moving horizon estima-
tion (MHE) (Hedengren and Edgar, 2006). Parameterizing the current states
with all previous measurements is one possible solution. Another approach
improves the solution speed but does not eliminate the need to solve an NLP
problem at every sampling instant (Hedengren and Edgar, 2004).

6 Summary and Conclusions

MPC is now suggested as a candidate to replace PID control due to recent developments in off-line calculations for efficient on-line implementation. Up to this point, the proposed algorithms suffer from dimensionality problems. For state parameterization, control applications are limited to small models and short control horizons. For constraint parameterization, control applications are limited to short control horizons and low number of inputs. The ISAT algorithm proposed in this work overcomes the dimensionality problems by adaptively storing only those regions accessed in practice. ISAT efficiently handles both NLP problems and constrained LQ problems. ISAT reduces to an adaptive version of state parameterized storage and retrieval of mp-QP problems when the error tolerance is reduced to zero.

References

Acevedo, J., Pistikopoulos, E., 1997. A multiparametric programming approach for linear process enginering problems under uncertainty. Ind. Eng. Chem. Res. 36, 717–728.

Barto, A., 1997. Neural Systems for Control. Academic Press, San Diego, CA.

Bellman, R., 1962. Applied Dynamic Programming. Princeton University Press, Princeton, NJ.

Bemporad, A., December 2003. Multiparametric nonlinear integer programming and explicit quantized optimal control. In: Proceedings of the 42nd IEEE Conference on Decision and Control. Maui, Hawaii, pp. 3167–3172.

Bemporad, A., Borrelli, F., Morari, M., December 2000. Optimal controllers for hybrid systems: stability and piecewise linear explicit form. In: Pro-

ceedings of the 39th IEEE Conference on Decision and Control. Sydney, Australia, pp. 1810–1815.

Bemporad, A., Filippi, C., December 2001. Suboptimal explicit MPC via approximate multiparametric quadratic programming. In: Proceedings of the 40th IEEE Conference on Decision and Control. Orlando, FL, pp. 4851–4856.

Bemporad, A., Morari, M., Dua, V., Pistokopoulos, E., 2002. The explicit linear quadratic regulator for constrained systems. Automatica 38, 3–20.

Bertsekas, D., 2001. Neuro-dyanamic programming: An overview. In: Proc. CPC VI. pp. 92–96.

Dua, V., Pistikopoulos, E., 1999. Algorithms for the solution of multiparametric mixed-integer nonlinear optimization problems. Ind. Eng. Chem. Res. 38, 3976–3987.

Gal, T., Nedoma, J., 1975. Multiparametric linear programming. Management Science 18, 406–442.

Ganesh, N., Biegler, L., 1987. A robust technique for process flowsheet optimization using simplified model approximations. Computers and Chemical Engineering 11, 553–565.

Grieder, P., Morari, M., December 2003. Complexity reduction of receding horizon control. In: Proceedings of the 42nd IEEE Conference on Decision and Control. Maui, Hawaii, pp. 3179–3190.

Hale, E., 2005. Multi-parametric nonlinear programming. Ph.D. thesis, University of Texas at Austin.

Hale, E., Qin, S., July 2004. Multi-parametric nonlinear programming and the evaluation of implicit optimization model adequacy. In: Proceedings of the 7th International Symposium on the Dynamics and Control of Process Systems. Cambridge, MA.

Hedengren, J., Edgar, T., 2004. In situ adaptive tabulation for real time control. In: Proceedings of the American Control Conference. Boston, MA.

Hedengren, J., Edgar, T., 2005. In situ adaptive tabulation for real time control. I&EC Research 44 (8), 2716–2724.

Hedengren, J., Edgar, T., 2006. Moving horizon estimation - the explicit solution. In: Proceedings of Chemical Process Control (CPC) VII Conference. Lake Louise, Alberta, Canada.

Johansen, T., December 2002. On multi-parametric nonlinear programming and explicit nonlinear model predictive control. In: Proceedings of the 41st IEEE Conference on Decision and Control. Las Vegas, NV, pp. 2768–2773.

Johansen, T., 2004. Approximate explicit receding horizon control of constrained nonlinear systems. Automatica 40, 293–300.

Johansen, T., Grancharova, A., 2003. Approximate explicit constrained linear model predictive control via orthogonal search tree. IEEE Transactions on Automatic Control 48 (5), 810–815.

Kaisare, N., Lee, J., Lee, J., 2003. Simulation based strategy for nonlinear optimal control: Application to a microbial cell reactor. International Journal of Robust and Nonlinear Control 13, 347–363.

Lee, J., Lee, J., 2004. Approximate dynamic programming strategies and their applicability for process control: A review and future directions. International Journal of Control 2 (3), 263–278.

Muske, K. R., Badgwell, T. A., 2002. Disturbance modeling for offset-free linear model predictive control. Journal of Process Control 12, 617–632.

Pannocchia, G., 2003. Robust disturbance modeling for model predictive control with application to multivariable ill-conditioned processes. Journal of Process Control 13 (8), 693–701.

Pannocchia, G., Laachi, N., Rawlings, J., 2004. A fast, easily turned, siso,

model predictive controller. In: DYCOPS. Boston, MA, p. 163.

Pannocchia, G., Laachi, N., Rawlings, J., 2005. A candidate to replace PID control: SISO constrained LQ control. AIChE Journal.

Pistikopoulos, E., Dua, V., Bozinis, N., Bemporad, A., Morari, M., 2000. On-line optimization via off-line parametric optimization tools. Computers and Chemical Engineering 24, 183–188.

Pistikopoulos, E., Dua, V., Bozinis, N., Bemporad, A., Morari, M., 2002. On-line optimization via off-line parametric optimization tools. Computers and Chemical Engineering 26, 175–185.

Pope, S., 1997. Computationally efficient implementation of combustion chemistry using in situ adaptive tabulation. Combustion Theory Modelling 1, 41–63.

Rossiter, J., Grieder, P., 2005. Using interpolation to improve efficiency of multiparametric predictive control. Automatica.

Sakizlis, V., Dua, V., Perkins, J., Pistikopoulos, E., 2004. Robust model-based tracking control using parametric programming. Computers and Chemical Engineering 28, 195–207.

Seron, M., Doná, J. D., Goodwin, G. C., December 2000. Global analytical model predictive control with input constraints. In: Proceedings of the 39th IEEE Conference on Decision and Control. Sydney, Australia, pp. 154–159.

Tondel, P., Johansen, T., Bemporad, A., 2003. An algorithm for multi-parametric quadratic programming and explicit mpc solutions. Automatica 39, 489–497.

Wolbert, D., Joulia, X., Koehret, B., Biegler, L., 1994. Flowsheet optimization and optimal sensitivity analysis using analytical derivatives. Computers and Chemical Engineering 18 (11/12), 1083–1095.

A   Elements of an ISAT Record

Two optional elements of the ISAT record are the exact sensitivity of optimal control parameters and the critical region for constrained LQ problems. When an exact sensitivity is not readily available, it must be estimated. This can be accomplished by perturbing the input parameters and re-converging the NLP or through linear regression using nearby stored data. There are significant drawbacks to each of the approximate techniques. Preferably, the sensitivity information is obtained from the Hessian matrix at the solution.

A.1   Sensitivity of Optimal Control to Parameters

The parameters of interest are any independent variable such as current state, model coefficient, or tuning constant. Ganesh and Biegler (1987) developed a reduced Hessian strategy for sensitivity analysis of optimal flowsheets. A part of their sensitivity derivation is given here. Sensitivities locally approximate the optimal solution with a 1st order solution. NMPC can be expressed more compactly with adjustable parameters ϕ, inequality constraints g, and equality constraints h.

$$\Phi^*(\phi) = \min_{x,u} \left(\Phi(x, u) \right) \tag{A.1a}$$

subject to

$$\phi \; given \tag{A.1b}$$

$$g(x, u, \phi) \leq 0 \tag{A.1c}$$

$$h(x, u, \phi) = 0 \tag{A.1d}$$

where x and u refer to the sequence of vectors $(x_1, x_2, \ldots, x_N)$ and $(u_0, u_1, \ldots, u_{N-1})$, respectively. The NLP is solved by minimizing the Lagrangian L

$$L(x, u, \phi) = \Phi(x, u, \phi) + \lambda g(x, u, \phi) + \nu h(x, u, \phi) \tag{A.2}$$

where Φ is the objective function, λ is the inequality constraint multiplier, and ν is the equality constraint multiplier. The Karush-Kuhn-Tucker (KKT) conditions are satisfied at the optimal solution.

$$\nabla\Phi(x, u, \phi) + \nabla g(x, u, \phi)\lambda + \nabla h(x, u, \phi)\nu = 0 \tag{A.3a}$$

$$\lambda g(x, u, \phi) = 0 \tag{A.3b}$$

$$\lambda \geq 0 \tag{A.3c}$$

$$g(x, u, \phi) \leq 0 \tag{A.3d}$$

$$h(x, u, \phi) = 0 \tag{A.3e}$$

The solution sensitivity reveals how the optimal solution changes with deviations in the the parameters ϕ. In order for a local sensitivity to exist, a few conditions must be met. First, the Lagrangian must be twice continuously differential in x and u and once in ϕ. Second, the constraint gradients must be linearly independent at the optimal solution. Finally, the second-order sufficiency conditions must be met. In generating the local sensitivities it is assumed that the active set does not change. The active constraints are

$$\nabla_x L(x, u, \phi) = 0 \tag{A.4a}$$

$$\nabla_u L(x, u, \phi) = 0 \tag{A.4b}$$

$$g_A(x, u, \phi) = 0 \tag{A.4c}$$

$$h(x, u, \phi) = 0 \tag{A.4d}$$

where g_A is the subset of g that are at the equality bound. The sensitivities are derived by taking the total derivative of the active constraints listed in Equation A.4.

$$\mathrm{d}[\nabla_x L(x, u, \phi)] = \nabla_{xx} L \mathrm{d}x + \nabla_{ux} L \mathrm{d}u + \nabla_x g_A \mathrm{d}\lambda + \nabla_x h \mathrm{d}\nu + \nabla_{\phi x} L^T \mathrm{d}\phi = 0$$

$$\text{(A.5a)}$$

$$\mathrm{d}[\nabla_u L(x, u, \phi)] = \nabla_{xu} L \mathrm{d}x + \nabla_{uu} L \mathrm{d}u + \nabla_u g_A \mathrm{d}\lambda + \nabla_u h \mathrm{d}\nu + \nabla_{\phi u} L^T \mathrm{d}\phi = 0$$

$$\text{(A.5b)}$$

$$\mathrm{d}g_A(x, u, \phi) = \nabla_x g_A^T \mathrm{d}x + \nabla_u g_A^T \mathrm{d}u + \nabla_\phi g_A^T \mathrm{d}\phi = 0 \qquad \text{(A.5c)}$$

$$\mathrm{d}h(x, u, \phi) = \nabla_x h^T \mathrm{d}x + \nabla_u h^T \mathrm{d}u + \nabla_\phi h^T \mathrm{d}\phi = 0 \qquad \text{(A.5d)}$$

Each of the equations in A.5 is divided by $\mathrm{d}\phi$. In the limit as $\mathrm{d}\phi$ shrinks to zero the local sensitivities become (with some rearrangement)

$$\begin{bmatrix} \nabla_\phi x^T \\ \nabla_\phi u^T \\ \nabla_\phi \lambda^T \\ \nabla_\phi \nu^T \end{bmatrix} = - \begin{bmatrix} \nabla_{xx} L & \nabla_{xu} L & \nabla_x g_A & \nabla_x h \\ \nabla_{ux} L & \nabla_{uu} L & \nabla_u g_A & \nabla_u h \\ \nabla_x g_A^T & \nabla_u g_A^T & 0 & 0 \\ \nabla_x h^T & \nabla_u h^T & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \nabla_{\phi x} L^T \\ \nabla_{\phi u} L^T \\ \nabla_\phi g_A^T \\ \nabla_\phi h^T \end{bmatrix} \qquad \text{(A.6)}$$

where $\nabla_\phi x$ is the state sensitivity, $\nabla_\phi u$ is the input sensitivity, $\nabla_\phi \lambda$ is the active inequality constraint multiplier sensitivity, and $\nabla_\phi \nu$ is the equality constraint multiplier sensitivity. Equation A.6 shows that the only requirement for a sensitivity calculation is the Lagrangian second partial derivatives with respect to the parameters. With analytical derivatives through automatic differentiation the sensitivity calculation speed can be greatly improved (Wolbert et al., 1994).

A.2 Defining the Critical Region

For unconstrained LQ problems the local sensitivity gives an exact optimal solution over all state space. In this case, the sensitivity is equivalent to the unconstrained LQR gain matrix. For constrained LQ problems the optimal solution is linearly dependent on the adjustable parameters ϕ within the same active constraint region. An individual query point ϕ_q can be tested to determine if it lies within this critical region (CR). A 1st order approximation of optimal variables at ϕ_q is determined.

$$x_q = x + \nabla_\phi x \left(\phi_q - \phi\right) \tag{A.7a}$$

$$u_q = u + \nabla_\phi u \left(\phi_q - \phi\right) \tag{A.7b}$$

$$\lambda_q = \lambda + \nabla_\phi \lambda \left(\phi_q - \phi\right) \tag{A.7c}$$

Equation A.8 gives the qualifications for a point within the CR.

$$g_I(x_q, u_q, \phi_q) \leq 0 \tag{A.8a}$$

$$\lambda_q \geq 0 \tag{A.8b}$$

where g_I is the set of inactive inequality constraints. If any of these qualifications are not met it indicates that the active set changed and the point lies outside the CR.