Combined Model Predictive Control and Scheduling with Dominant Time Constant Compensation

Logan Beal  
Brigham Young University

Junho Park  
Brigham Young University

See next page for additional authors

Follow this and additional works at: https://scholarsarchive.byu.edu/facpub

Part of the Chemical Engineering Commons

Original Publication Citation

BYU ScholarsArchive Citation
Beal, Logan; Park, Junho; Petersen, Damon; Warnick, Sean C.; and Hedengren, John, "Combined Model Predictive Control and Scheduling with Dominant Time Constant Compensation" (2017). All Faculty Publications. 1905.
https://scholarsarchive.byu.edu/facpub/1905
Authors
Logan Beal, Junho Park, Damon Petersen, Sean C. Warnick, and John Hedengren

This peer-reviewed article is available at BYU ScholarsArchive: https://scholarsarchive.byu.edu/facpub/1905
Combined Model Predictive Control and Scheduling with Dominant Time Constant Compensation

Logan D. R. Beal\textsuperscript{a}, Junho Park\textsuperscript{a}, Damon Petersen\textsuperscript{a}, Sean Warnick\textsuperscript{b}, John D. Hedengren\textsuperscript{a,*}

\textsuperscript{a}Department of Chemical Engineering, Brigham Young University, Provo, Utah, USA
\textsuperscript{b}Department of Computer Science, Brigham Young University, Provo, Utah, USA

Abstract

Linear model predictive control is extended to both control and optimize a product grade schedule. The proposed methods are time scaling of the linear dynamics based on throughput rates and grade-based objectives for product scheduling based on a mathematical program with complementarity constraints. The linear model is adjusted with a residence time approximation to time-scale the dynamics based on throughput. Although nonlinear models directly account for changing dynamics, the model form is restricted to linear differential equations to enable fast online cycle times for large-scale and real-time systems. This method of extending a linear time-invariant model for scheduling is designed for many advanced control applications that currently use linear models. Simultaneous product switching and grade target management is demonstrated on a reactor benchmark application. The objective is a continuous form of discrete ranges for product targets and economic terms that maximize overall profitability.

Keywords: scheduling, model predictive control, dynamic pricing, time scaling, complementarity constraints

*corresponding author

Email address: john_hedengren@byu.edu (John D. Hedengren)
1. Introduction

Time-of-day energy pricing for electricity and natural gas pose a challenge and opportunity for industrial scale manufacturing processes. In many manufacturing processes in which Model Predictive Control (MPC) is well-established, such as downstream refining and petrochemicals, there is lost opportunity when advanced process control only operates at certain conditions but must be turned off when unit production is lowered [1]. A challenge with changing production rates is that the dynamics of a process often change dramatically with throughput. Empirical models identified at high throughput rates are often inaccurate and lead to poor control performance at low production rates. An opportunity with time-of-day pricing is to temporarily reduce the consumption of energy-intensive processes for periods when the energy costs are sufficiently high. During off-peak periods, the production rate is increased or more energy-intensive product grades are produced to take advantage of low energy costs. With typical daily cycles in energy costs, a cyclical operation forms that the original advanced control system may not be designed to follow. When production targets are not set to maximize to production constraints, MPC may switch to maximize energy efficiency or other secondary objectives. Set point targets traditionally come from a real-time optimization (RTO) application that optimizes a steady state operating point for the plant [2].

Segregated control and scheduling structure is historically due to computational factors that limit application complexity [3]. As a result, the control and scheduling fields have grown independently and without coordination, leading to a loss of opportunity from combining the applications [2, 4]. By combining process scheduling of set points with control, the inefficiencies of application layering are avoided. One such inefficiency that results from application layering is the infeasibility on the control layer of individual solutions pass from the supervisory layer [5]. Scheduling applications frequently do not consider grade transition times because of the large combinatorial look-up table that would be required to consider all possible transitions. Additionally, objectives
of individual solutions can oppose each other. For example, the controller does
not consider the most economical route to reach a target set point given by the
scheduler or steady-state optimizer [6].

The computational barriers for combined control and scheduling are dimin-
ished with improved computer hardware and adaptation of algorithms to the
hardware. Algorithmic barriers are being overcome with a number of key con-
tributions that are opening several fronts of development [7, 8]. Hardware or
network resources such as multi-core, cloud-based, and graphics processing units
(GPUs) provide access to previously inaccessible computing power. However,
advanced architectures such as GPUs for optimization impose some limitations
on the type of problems that can be solved because the algorithms have not yet
been adapted to take full advantage of the architecture [9].

Economic MPC (EMPC) [10, 11] uses an objective function that maximizes
a profit function rather than targeting a set point as in standard MPC. Including
the profit function directly in the MPC application ensures that decisions are
directly driven by economic considerations. The profit function also provides
guidance on product scheduling, although work on EMPC up to this point
has focused on single products. The drawback of this approach is that EMPC
generally requires a short time horizon such that the longer horizon required for
scheduling constraints and objectives cannot be met [10]. MPC for supply chain
management [12] is an alternative strategy that extends the control horizon to
schedule product movement through a distribution network.

Dynamic Real-Time Optimization (DRTO) also has an economic objective
function but augments a steady state Real-Time Optimization (RTO) with select
differential equations that capture the salient and dominant dynamics of
a process [13, 14, 15]. One drawback of RTO is that the process must be at
steady-state [16] to perform data reconciliation. RTO has traditionally been
applied to processes that do not have grade transitions but are dominated by
changing economics, disturbances, and slow dominant dynamics. With dynam-
ics included, DRTO can be solved more frequently than RTO applications and
can be solved during periods of transient disturbances, during startup, or during
shutdown periods. RTO calculations are typically performed every hour to every day while DRTO optimizes the transition between steady-state conditions. Like EMPC, DRTO does not manage multiple sequential product campaigns as a scheduler.

Complete integration of scheduling and control requires an extended prediction horizon to plan the production sequence as well as near-term control actions. Two integration approaches are referred to as top-down (add control and dynamics to a scheduling application) or bottom-up (add scheduling to a control algorithm) [3]. An early top-down implementation includes differential and algebraic equations in the scheduling application [17]. Another method is the scale-bridging model (SBM) in which a simplified model of process dynamics is embedded in the scheduling application [18, 19, 20]. A benefit of this method is disturbance rejection [21]. Algorithms include Benders’ decomposition [22] for problems that have a large-scale structured form and Dinkelbach’s algorithm [23] for non-convex problems that require global optimization methods. Applications of combined scheduling and control include batch processes [5, 24], polymer reactors [25], parallel Continuously Stirred Tank Reactors (CSTRs) [26], and an electrical grid that responds to current and future price signals [27].

Variable electrical pricing incentivizes reduced consumption during peak hours [28]. It is desirable to match generation to consumption, but the adoption of more renewable energy requires producers and consumers to respond to price signals [29, 30, 31]. Energy producers may expose consumers to time-of-day pricing to discourage consumption during peak hours [32]. Scheduling operation of chemical processing [33, 34], oil refining [35], and air separation [36] are some examples of industrial units that can shed electrical load during peak hours, typically in the middle of the day. Many cooling-limited processes also operate more efficiently at night [34]. Periodic constraints can be used to optimize a typical daily cycle.

Prior work in scheduling and control integration has been centered around slot-based, continuous-time scheduling formulations. The benefit of discrete-time formulation has been shown [34]. However, the non-linear discrete-time
formulation proved computationally difficult. The purpose of this work is to restrict the dynamic model to linear form while capturing benefits of the integration of scheduling and control. There is a large installed base of advanced controls that utilize linear models [1]. A unique aspect of this work is a time-scaling algorithm that adjusts the linear dynamic model based on residence time calculations with a theoretical foundation for first-order systems. The time-scaling approximation is applied to higher order, finite impulse response models that are common in industrial practice. These linear models are used in the combined scheduling and control application.

2. Time-Scaling with First-Order Systems

It is well known that linear MPC performance degrades with changes in the actual process time constant or gain [37]. This effect has been quantified for MPC where there is model mismatch in the time constant or gain of a first order system. A simple example is where the actual system is described by a single differential equation as \( \tau_p \frac{dy}{dt} = -y + K_p u \) with a process time constant of \( \tau_p = 1 \) and a process gain of \( K_p = 1 \). An MPC controller with objective \( \sum_{i=1}^{20} |y_i - 5| \) drives the response from a set point of 0 to 5. The controller model is similar to the process but with variable model time constant of \( \tau_m \) and gain \( K_m \) in the equation \( \tau_m \frac{dy}{dt} = -y + K_m u \). Common industrial practice is that acceptable MPC performance can be achieved with gain mismatch less than 30% \((0.7 \leq K_m \leq 1.3)\) and time constant mismatch less than 50% \((0.5 \leq \tau_m \leq 1.5)\) [37]. The dominant time constant for many industrial processes is characterized by the volume \( V \) divided by the volumetric flowrate \( q \) as \( \tau_p = V/q \). The explicit solution to the first order equation is given by Equation 1.

\[
y[k + 1] = \exp\left(-\frac{\Delta t}{\tau_p}\right) y[k] + \left(1 - \exp\left(-\frac{\Delta t}{\tau_p}\right)\right) K_p u[k]
\]

where \( y \) is the output, \( u \) is the input, \( k \) is the discrete time step, and \( t \) is the time. The integrated sum of absolute errors is computed for combinations of \( K_m \) and \( \tau_m \), each between 0.1 and 5.0. The 3D contour plot shows the control
performance over the range of time constant and gain mismatch (see Figure 1a). A mismatch in the gain (x-axis) and time constant (y-axis) are plotted versus the error. The vertical axis (z-axis) is the integrated absolute value of the objective function for a set point change from 0 to 5. A lower sum of absolute errors equates to better control performance with a minimum at $K_m = 1$ and $\tau_m = 1$ (no model mismatch) as shown in Figure 1b.

![MPC objective function](image-a)

![Control performance sensitivity](image-b)

(a) MPC objective function  
(b) Control performance sensitivity.

Figure 1: Performance degradation of MPC with model mismatch

Especially poor performance occurs when the model has a higher time-constant (slower) than the actual process. The model in MPC predicts that changes happen slower than are actually the case, leading to a controller that is more aggressive. This aggressiveness translates into overshoot of the set point or even instability. Likewise, a model with a lower gain than the actual process also exhibits poor control performance. The model predicts that larger changes in the Manipulated Variable (MV) are required to drive the process to the new set point. In reality, a smaller adjustment is required and the over-reaction of the controller leads to overshoot and possibly instability. A contour plot of the performance profile gives insight on the performance as shown in Figure 2a. Figure 2b shows the performance with the time-scaled model. The parallel contour lines show that there is no performance degradation of the controller when $\tau_m$ changes such as a production rate decrease or increase. The abil-
ity of the MPC to function over all production rates is required for processes that respond to utility price or product demand signals. This simple example shows the increased effectiveness over a wider range of operating conditions with time-scaling while still preserving the linear model.

![Contour plot of performance degradation of MPC](image)

Figure 2: Contour plot of performance degradation of MPC. An objective below 7 is acceptable, between 7-30 is marginal, and above 30 (red line) is poor performance.

This result nearly agrees with industry observations, shown with dashed box in Figure 2. Another feature of this result is that the combination of high mismatch in time-constant ($\tau_m$) and low mismatch in gain ($K_m$) combine to form a region of poor control performance. Acceptable or marginal performance is also possible if both $\tau_m$ and $K_m$ are both too high or too low. If there is low mismatch in time-constant ($\tau_m$) and high mismatch in gain ($K_m$), the controller degradation is manifest as sluggishness and only incremental moves in the MV but not instability.

The time scaling approach adjusts either the controller cycle time or the discrete model time step based on the change in unit throughput $q$ relative to the nominal throughput $\bar{q}$. $\bar{\tau}_p$ is the nominal time constant associated with $\bar{q}$. The modified process time constant is $\tau_p = q \times (\bar{\tau}_p/\bar{q})$ which now has a linear relationship to $q$. If the process model is not easily adjusted, the cycle time $\Delta t$ of the controller is adjusted to $\Delta t \times (q/\bar{q})$ to compensate for the changing process dynamics. For first-order systems, this gives an exact representation of
the nonlinear dynamics without modifying the original linear model.

3. Selective Time-Scaling

Multi-variate and higher-order systems may have certain MV to CV relationships that are known to scale with changing unit throughput while others are invariant to throughput changes. Prior work has focused on decomposition of fast and slow dynamics [38] or variable time-delay of measurements [39, 40]. For systems with multiple MVs and CVs, only the relationships that are sensitive to throughput are scaled. These can be identified with a dynamic process simulator or else by repeating plant identification tests at low and high production rates. A method to scale higher order systems is to transform the linear time-invariant (LTI) model into discrete form. In discrete form, the sampling time is scaled by \( q/\bar{q} \) and resampled to preserve the overall model sampling time. As an example of this time scaling approach, consider the 7th order system given by Equation 2 as a transfer function in terms of Laplace variable \( s \).

\[
G(s) = \frac{CV(s)}{MV(s)} = \frac{1.5}{(s^2 + 0.6s + 1)(0.5s + 1)^5}
\]  

(2)

Suppose that the dynamics of this system depend on the production rate and that the feed rate to the unit is reduced to half of the rate where the model is originally identified. When a time-scaling transformation of \( q/\bar{q} = 2 \) is applied, the new transfer function is also a 7th order system but with shifted dynamics. The steady state gain of the transfer function is preserved with this method of dynamic transformation. The resulting transfer function is Equation 3.

\[
G(s) = \frac{CV(s)}{MV(s)} = \frac{1.5}{(4s^7 + 21.2s^6 + 47s^5 + 57s^4 + 42s^3 + 20s^2 + 6.2s + 1)}
\]  

(3)

The continuous transfer function is first converted to discrete form with a sufficiently small sampling time. For this case, a sampling time of 0.5 sec is chosen for the continuous to discrete transformation. The discrete model sampling time is set to \( 2 \times 0.5 \text{ sec} = 1.0 \text{ sec} \) based on the time-scaling factor.
and then the model is resampled to 0.5 sec to be consistent with other unscaled MV/CV models. Continuous to discrete transformations and discrete model resampling for LTI models are standard methods [41] and are not repeated here for the sake of brevity. A graphical demonstration of this method is shown in Figure 3.

4. Scheduling and Control Formulation

An innovation of this work is to combine scheduling and control with linear models and quadratic objective functions for fast solution with Quadratic Programming (QP) [42] or Nonlinear Programming (NLP) solvers. There are multiple methods for formulating tiered price structures. The focus of this work is on creating mathematical expressions that have continuous first and second derivatives and fit within the QP or a QP with Quadratic Constraint (QPQC) framework. In contrast, most modern scheduling applications utilize integer variables, often as binary decision variables. Including integers requires MILP or MINLP solvers, which are significantly slower and more complex.
4.1. MPCC Steps for Product Pricing

The method uses a continuous formulation to logical decisions. It uses a Mathematical Programs with Complementarity Constraints (MPCC) formulation to avoid binary variables that would otherwise accommodate tiered pricing structures [43]. This formulation uses a step function MPCC, using constraints $v_0y = 0$ and $v_1(1 - y) = 0$. $v_0$ and $v_1$ are positive slack variables. The complementarity constraints force variables to their bounds – resulting in $y$ being a binary variable at its bounds $[0, 1]$.

This condition can be difficult for solvers to find a solution so the condition is typically either solved as an equivalent inequality ($v_0y \leq 0$), a relaxed inequality ($v_0y \leq \epsilon$), or included in the objective function ($\min v_0y$) [44]. The step function MPCC that turns $y$ from 0 to 1 at switching point $x_p$ is shown in Equation 4a-4d.

$$
\min_{v_0, v_1, y} v_0 y + v_1 (1 - y) \quad (4a)
$$

$$
x - x_p = v_1 - v_0 \quad (4b)
$$

$$
v_1, v_0 \geq 0 \quad (4c)
$$

$$
0 \leq y \leq 1 \quad (4d)
$$

To preserve the QP structure, the complementarity constraints are included in the objective function. If the complementarity constraints are included as inequalities a QPQC or NLP solver must be used. With sufficient weighting, the complementarity constraints in the objective function are zero at a final optimal solution but not necessarily along the search path to the solution. Being zero at the solution, the complementarity constraints do not influence other objective terms such as minimizing energy consumption or maximizing profit. The MPCC switching conditions are combined to create a tiered product pricing structure as shown in Figure 4.

Each product has a different value and potentially a different width of specification limits. The positive and negative step functions at different switching
points are combined with pricing information to create an objective function with multiple products. The individual steps are shown in the upper subplot of Figure 4 while the summation of all steps is shown in the lower subplot. Product A ranges from 0 to 1 and has a price of 1 per unit production. Product B ranges from 2 to 3 and has a price of 2 per unit production. Product C ranges from 2 to 3 and has the highest price of 3 per unit production. The facility has a capacity of two units of production per hour and the minimum amount of production for each product over a 7 hour time window is 2 units of A, 5 units of B, and 3 units of C. Because C is the most valuable product, any spare capacity should favor the production of C. In switching between products, a schedule should account for transition material between grades that does not have value. The speed of a transition is limited by the maximum move rate of the Manipulated Variable (MV) of 1.6 per hour and the dynamics of the process. The dynamics of the process are simply a linear first-order system as \( \tau \frac{d(CV)}{dt} = -CV + MV \) with a time constant \( \tau \) of 1.0.

The grade-specific objective function with a quadratic objective, linear equation, and simple inequality constraints is added to the process model to optimize the timing of grade transition switches. A complete statement of the QP
optimization problem is shown in Appendix A. The series of three products demonstrates a simple combined control and scheduling problem as shown in Figures 5 and 6. The cases have different initial conditions and constraints but the same underlying model. The first case has an initial condition for CV of 0.0 at the lower specification limit for product A while the second case starts at CV of 4.5.

![Figure 5: The control and scheduling optimization are combined to determine optimal MV movement along with the optimal order and quantity of production for each grade.](image)

The optimizer minimizes the amount of products A and B while meeting the minimum requirements before transitioning to the next grade. This allows the schedule to favor the production of C, the most profitable product. By minimizing the amount of products A and B, the scheduler creates a plan that produces an extra unit of product C. The combined controller and scheduler anticipates grade transitions by shifting the production specification to the upper limit to minimize the transition time to the next grade. This is apparent at time 1.0 hr and 4.2 hr where the product is already transitioning to the new grade just as the minimum required amount for products A and B are produced. The combined scheduler and controller adjusts both the control actions as well as the order in which the products are produced. In the next example, the initial condition is shifted to a different starting location to demonstrate this feature.
of the approach.

The computational time is an important factor for control problems where the algorithm must return a solution within the cycle time necessary to rejected disturbances and maintain process stability. The fully discretized combined scheduling and control problem with 3 products and discrete time points has 2730 variables, 1820 equations, and 910 degrees of freedom. The problem is solved on a Dell R815 server with an AMD Opteron 6276 Processor and 64 GB of RAM. The problem requires 8.6 s to converge with the APOPT solver (version 1.0) or 4.0 s with the IPOPT solver (version 3.10). MPC commonly uses a warm-start from a prior solution to improve efficiency. With a warm-start from the prior time-step, the optimizer requires 0.5 s with APOPT and 3.0 s with IPOPT. Active-set solvers (such as APOPT) and interior point solvers (such as IPOPT) are both effective for large-scale MPC applications although interior point solvers have better performance as degrees of freedom increase [7].

4.2. Dynamic Cyclic Schedule

A common method to optimize a schedule in current practice is to place products on a fixed grade wheel where each product is visited sequentially in a rotation. The cyclic schedule visits all grades in a forward and then a reverse order to end up at the original grade and begin the cycle again. If there is a sudden demand for a particular grade out of sequence, the product wheel is still typically traversed in the grade wheel order but there may be a minimal amount of the less desirable grades in favor of the desired grade production. In some processes, such as polyolefin production, there are multiple versions of the grade wheel. One version of the grade wheel may include only commodity products while other wheels or sub-wheels may be a more complex cycles that include less frequently produced products.

The contribution of this work is the discrete time, extended controller and scheduler that is also able to produce a cyclic schedule. However, this cyclic schedule is not fixed but adjusts the sequence of products automatically when updated economic or constraint information is available. The schedule and
control action may update every controller cycle (e.g. every minute). Some constraints for scheduling are periodic boundary constraints where the final condition must be equal to the initial condition, contracted quantities that must be produced by a certain date or time, particular equipment limitation, or time-of-day transition constraints. An example of a time-of-day constraint is that certain grade transitions need start-up or shut-down of auxiliary equipment. A constraint may be that the grade transition should happen only during a weekday day-time shift where there is adequate operator support.

The prior example problem is augmented with intermediate production targets and a periodic constraint. The scheduler and controller, shown in Figure 6 produces an optimized schedule and control actions. It is a dynamic rather than a static grade wheel because the order of product production and quantity is re-optimized every cycle of the controller. The product order or quantity may change based on changing customer demand, price signals for electricity or feed costs, or disturbances that drive the system to a different state.

![Optimized schedule with periodic condition, intermediate production targets, and final production targets.](image)

Figure 6: Optimized schedule with periodic condition, intermediate production targets, and final production targets. The schedule and control actions adjust to meet constraints and maximize profit (product C).

The optimized solution meets constraints as well as maximizes the production of product C, the highest value product. The intermediate constraints
originating from contracted delivery times and storage capacity of a particular product. In this case, the storage capacity of all products is less than 5 units of production. This requires two deliveries of product B that are scheduled at two equally spaced intervals of 5 hrs. The production of product C exceeds the required delivery amount because it is the highest value product. Although the initial condition is at product C, the controller immediately targets product B to meet the required delivery of 5 units of production at 5 hrs. Without the periodic constraint, production of product C would be maximized before transitioning because the scheduler evaluates the alternative that a transition back to product C results in lost profit potential. However, with the required transition back to product C, the scheduler puts excess production of product C at the end.

While this method is capable of producing cyclic schedules, the optimizer should begin from current conditions rather than steady-state product conditions to fully integrate control and scheduling. Cyclic schedules combined with online control may lead back to off-spec conditions because of disturbances or because the controller is in a transition. Instead, this method is better suited to a different set of constraints – production amounts and due dates. These constraints give more freedom to the optimizer so the economic objective will improve or be equal to the solution with periodic constraints.

As with adding any constraint, there is potential to make the problem infeasible. Multi-objective optimization statements, such as those used to refine a dynamic grade wheel sequence, can also be posed with a rank-ordered set of constraints with an explicit prioritization of objectives. The $\ell_1$-norm dead-band formulation is discussed in more detail in [7] and related to combined scheduling and control production targets in Section 4.3. Posing the constraints as multi-objective penalties versus hard constraints allows the problem to remain feasible yet still meet the most important objectives in order of priority.
4.3. Acceptable Range of Production Quantity

One drawback to the prior examples is that all spare production capacity is typically placed on the highest value product. Over-production of any product can have the effect of lowering the selling price because of supply and demand market forces. In scheduling, there is often a range of production quantity that is acceptable instead of just a single hard limit. To accommodate this, the scheduling and control algorithm can use an $\ell_1$-norm objective function to give a target region for the production quantity, rather than one specific hard limit. Equation 5 shows a generalized $\ell_1$-norm control formulation used in this work.

$$\min_{x, CV, MV} \Phi = w_{hi}^T e_{hi} + w_{lo}^T e_{lo} + x^T c_x + MV^T c_{MV} + \Delta MV^T c_{\Delta MV}$$

s.t. 
$$0 = f \left( \frac{dx}{dt}, x, \frac{dCV}{dt}, CV, MV \right)$$
$$e_{hi} \geq x - d_{hi}$$
$$e_{lo} \geq d_{lo} - x$$

In this formulation, $\Phi$ is the objective function, $x$ is the production quantity per grade. Parameters $w_{lo}$ and $w_{hi}$ are penalty matrices for solutions outside of the production target region. The slack variable $e_{lo}$ and $e_{hi}$ define the error of the dead-band low and high limits. Parameters $c_x$, $c_{MV}$, and $c_{\Delta MV}$ are cost vectors for the production quantity (positive values minimize production within region, negative values maximize production within region), the MVs (positive values minimize MV quantities such as energy use, negative values maximize MV quantity such as pump speed correlated to higher efficiency), and change of MVs, respectively. The function $f$ is an open-equation set of equations as functions of $x$, $CV$, $MV$, and time derivatives of $x$ and $CV$. The demand targets $d_{lo}$ and $d_{hi}$ define lower and upper target limits for production.

This range formulation is not used in this work but is presented to demonstrate one of many ways that this problem structure can be expanded to meet various scheduling needs.
5. Application: Continuously Stirred Tank Reactor

A continuously stirred tank reactor (CSTR) is a common benchmark used in nonlinear model predictive control and scheduling applications [34, 3, 16]. This nonlinear application is used to demonstrate the strength of the approach to time-scale based on throughput and combined scheduling & control.

The linear time-scaled model changes dynamics based on reactor flow rate, allowing linear MPC to be used instead of nonlinear MPC. The multi-product scheduling objective is used instead of a simple target tracking to combine the scheduling and control into one application. Rapid convergence is ensured with a linear model and quadratic objective because the problem is convex and because QP solvers are efficient for large-scale systems.

The CSTR application is highly nonlinear because of an exothermic reaction that has the potential to cause rapid reaction of stored reactant and thereby cause a temperature run-away. The CSTR shares characteristics of many industrial processes such as polymer reactors or many refining processes but with much simpler mathematics that are amenable to demonstrating a new approach for control and scheduling. The liquid full reactor is used to convert compounds \( A \Rightarrow B \) with constant liquid density \( (\rho) \) and heat capacity \( (C_p) \) as shown in Figure 7.

The reaction kinetics are first order and irreversible. Reaction of \( A \) to \( B \) is exothermic with the potential for temperature run-away because of the exponential dependence of reaction rate on temperature, typical of an Arrhenius form for reaction rates. The reactor is well-mixed with reactor concentration and temperature equally distributed and also equal to the outlet measured values. The reactor temperature is regulated with a cooling jacket liquid temperature, \( T_c \). The cooling jacket temperature is normally regulated by adjusting the rate of cooling or the coolant flow rate but in this model the jacket temperature is assumed to be controlled directly and the dynamics are approximated by a maximum rate of change.

The dynamics of the CSTR are dictated by a species and energy balance as
Figure 7: Diagram of the well-mixed and liquid-full CSTR. The $A \rightarrow B$ reaction is exothermic and controlled by a cooling jacket fluid.

shown in Equations 6-7 and in Figure 8.

$$V \frac{dC_A}{dt} = q(C_{A0} - C_A) - k_0 V e^{-\frac{E_A}{RT}} C_A$$  \hspace{1cm} (6)

$$\rho C_p V \frac{dT}{dt} = q \rho C_p (T_f - T) - k_0 V e^{-\frac{E_A}{RT}} C_A \Delta H_r - U A (T - T_c)$$  \hspace{1cm} (7)

where $V$ is the volume of the reactor, $C_A$ is the concentration of reactant $A$, $q$ is the volumetric flowrate, $C_{A0}$ is the inlet concentration of reactant $A$. The energy balance includes terms $U A$ as an overall heat transfer coefficient times the tank surface area, $C_p$ as the reactor fluid heat capacity, $\rho$ as the fluid density, $T_f$ as the temperature of the feed stream, $T$ as the temperature of reactor, and $T_c$ as the temperature of cooling jacket fluid. Terms related to the reaction include $\Delta H_r$ as the heat of reaction, $E_A$ as the activation energy, $R$ as the universal gas constant, and $k_0$ as the pre-exponential factor. Table 1 lists the CSTR parameters and the associated values.

A regression is shown with varying orders for an Output Error (OE) time-series model in Figure 8. Second and third order models have nearly the same fit to the nonlinear regression while a first order model is insufficient in capturing
The concentration response of the reactor at half production rate is shown in the bottom subplot of Figure 8 with the second order time-scaled model that originally fit with simulated data from the full production rate simulation. The time-scaled approach is effective at capturing the essential process dynamics without re-fitting a process model at low rates.

One reactor makes multiple products by varying the concentrations of $A$ and $B$ in the reactor. The cooling jacket temperature $T_c$ is the manipulated variable in this optimization. The production rate of the reactor changes throughout a typical 24-hour period because of time-of-day pricing that necessitates a cut-

---

### Table 1: Reactor Initial Conditions and Parameter Values

<table>
<thead>
<tr>
<th>States</th>
<th>Description</th>
<th>Initial Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_A$</td>
<td>Concentration of reactant $A$</td>
<td>$0.1 \frac{mol}{L}$</td>
</tr>
<tr>
<td>$C_B$</td>
<td>Concentration of product $B$</td>
<td>$0.9 \frac{mol}{L}$</td>
</tr>
<tr>
<td>$T$</td>
<td>Reactor temperature</td>
<td>$386.82 \text{ K}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Manipulated Variables</th>
<th>Description</th>
<th>Initial Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_c$</td>
<td>Cooling jacket temperature</td>
<td>$300 \text{ K}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q$</td>
<td>Volumetric flowrate</td>
<td>$12 \text{ m}^3/\text{hr}$</td>
</tr>
<tr>
<td>$V$</td>
<td>Tank volume</td>
<td>$40 \text{ m}^3$</td>
</tr>
<tr>
<td>$C_{A0}$</td>
<td>Feed concentration of reactant $A$</td>
<td>$1 \frac{mol}{L}$</td>
</tr>
<tr>
<td>$UA$</td>
<td>Overall heat transfer coefficient</td>
<td>$5000 \frac{W}{K}$</td>
</tr>
<tr>
<td>$C_p$</td>
<td>Heat capacity of reactor fluid</td>
<td>$0.239 \frac{J}{kg \text{ K}}$</td>
</tr>
<tr>
<td>$\rho$</td>
<td>Density of reactor fluid</td>
<td>$1000 \frac{kg}{m^3}$</td>
</tr>
<tr>
<td>$T_f$</td>
<td>Feed temperature</td>
<td>$350 \text{ K}$</td>
</tr>
<tr>
<td>$\Delta H_r$</td>
<td>Heat of reaction (exothermic)</td>
<td>$-11.95 \frac{MJ}{mol}$</td>
</tr>
<tr>
<td>$k_0$</td>
<td>Pre-exponential factor, rate constant</td>
<td>$1.8 \times 10^{-1} \frac{1}{\text{hr}}$</td>
</tr>
<tr>
<td>$E_A/R$</td>
<td>Activation energy divided by $R$</td>
<td>$8750 \text{ K}$</td>
</tr>
</tbody>
</table>
back of production during peak energy prices and when cooling capacity is limited. The production rate is modified by adjusting the total flow through the reactor \( (q) \) between half-rates at 6 \( \frac{m^3}{hr} \) and full-rates at 12 \( \frac{m^3}{hr} \).

There is demand for three products with quantities that must be met in the schedule over a 48-hour horizon. The product descriptions and quantities are shown in Table 2.

Table 2: Product Summary with Demand and Price

<table>
<thead>
<tr>
<th>Product</th>
<th>( C_A )</th>
<th>Demand</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(mol/L)</td>
<td>(m^3)</td>
<td>($) / m^3</td>
</tr>
<tr>
<td>( P_1 )</td>
<td>0.12 ± 0.01</td>
<td>120</td>
<td>9</td>
</tr>
<tr>
<td>( P_2 )</td>
<td>0.25 ± 0.01</td>
<td>130</td>
<td>11</td>
</tr>
<tr>
<td>( P_3 )</td>
<td>0.35 ± 0.01</td>
<td>150</td>
<td>6</td>
</tr>
</tbody>
</table>

The most valuable is product \( P_2 \) while the least valuable product is \( P_3 \). Although \( P_3 \) has the lowest price, it also has the highest required quantity. Spare capacity in the production facility favors product \( P_2 \). A potential drawback to always switching to \( P_2 \) at the end of a campaign is that there is lost material
during the transition back to $P_2$. An improved strategy is to make excess $P_2$ when the schedule requires it to meet a minimum target demand instead of transitioning back to $P_2$ near the end of the time period. The combined control and scheduling solution is shown in Figure 9 over a 48 hour period with 6 minute time intervals. The problem is discretized with orthogonal collocation on finite elements. Each 6 minute segment is integrated with Radau quadrature. The resulting QP or QPQC problem is solved with a nonlinear programming solver with a simultaneous solution of the objective and equations. If specific control action is needed at more frequent intervals, the first steps of the horizon could be adjusted to meet a required controller cycle time. This would develop a near-term move plan and simultaneously solve the scheduling optimization problem with one application. The feed flow rate is decreased to half each day between the hours of 08:00 and 18:00 as is done with some energy intensive processes that exploit time-of-day electricity pricing. The addition of production rate as a decision variable and the associated cooling constraints is outside the scope of this work because the model becomes nonlinear. The simultaneous control and scheduling of production rate and product grade sequence is the topic of a future publication (see [34] for preliminary results). The top sub-plot is the sequence of control moves to drive the system to produce on-spec products and transition between products. The middle sub-plot shows the grade specifications and the concentration in the reactor. The bottom sub-plot is the total production of each grade with the minimum required as indicated by the circle markers at hour 48. The production rate is non-zero during transition periods because the total rate includes production of off-spec as well as on-spec grade material.

The control influences the scheduling solution and the scheduling solution gives the controller target values. The controller adjusts the jacket temperature ($T_c$) to minimize the transition time between grade specifications and remain within the grade limits. The controller response improves with knowledge of the scheduling targets because pro-active pre-transition movement shows that the product specification are pushed to an upper limit right before the transition
begins. One factor that affects the schedule is the speed of the transition from one product to another. The control dynamics are due to limitations in the manipulated variable movement and the process dynamics affecting the speed of attaining concentration range. The speed of the control response is factored into the schedule as lost production time when transition (off-spec) material is produced. Blending of transition material is one strategy to dilute the off-test material in prime products. This strategy is not considered in this approach but could be included as a constraint on the amount of reblend fraction that is allowed in the final product.

The scheduling profit function is an application and adaptation of Equations 4a-4d. Figure 10 displays the slack and step function for each of the product limits along with the overall profit function.

The slack variables and complementarity conditions combine to create discrete steps with a function that has continuous first and second derivatives. For this problem, the complementarity constraints were included as constraints and in the objective function to assist the optimizer and ensure binary decision
variables. The transition time to each product is not specified beforehand but is a result of the optimization solver finding a sequence of grade transitions that maximize overall profit. The individual steps of Figure 10 occur when one of the product grade limits is crossed. One of each of the slack variables for each step is non-zero before or after the step. The slack variables guide the solver with a continuous form of a discrete grade switching function. The overall solution is intuitive because the optimizer produces excess product $P_2$ (highest volumetric price) but meets only the minimum production quantities for the lower value products ($P_1$ and $P_3$).

A weakness of this method is that the CSTR is nonlinear but the solution is computed with a linearized model. Linearization neglects features of the process that may be optimized if a suitable process model more accurately represented
the physical process. A key contribution of this work is that the solution to
the combined control and scheduling problem is relatively computationally in-
expensive in comparison to a full nonlinear solution. In this case, the linearized
model with time-scaling has a total of 22,048 variables and 16,640 equations
and is solved on a Dell R815 with an AMD Opteron 6276 Processor and 64
GB of RAM. After initialization, the problem requires 23.7 s to converge with
the IPOPT solver. This is representative of the cycle time of the application
as it repeatedly solves to reject disturbances and as new objective information
or demand constraint information is available. This speed, coupled with the
inclusion of a process model sufficient for control, allows this formulation to be
utilized for on-line control, on top of providing an advanced schedule. This is
the epitome of combining scheduling and control – a fully unified optimization
that can replace both layers.

A nonlinear version of this application is future work that will be reported
in a subsequent publication. A key difference with the full nonlinear solution
is that the solution in this work is several orders of magnitude faster because
of the quadratic objective and linear time-scaled model. Although combined
scheduling and control with linear models or feedback linearization does not
always accurately predict a highly nonlinear system, the linearized solution is a
valuable starting point to initialize a nonlinear solution. Also, many processes
are not highly nonlinear and this approach is likely suitable for systems that are
already controlled with linear MPC.

6. Conclusion

A combined scheduling and control application is enabled by an MPCC ob-
jective function, discrete-time, and linear time-scaling of process dynamics based
on production rate changes. The objective of this work is to extend traditional
linear MPC applications with a scheduling objective that allows for rapid con-
vergence for real-time applications. One drawback of this work is that nonlin-
earities are not included in the application. These nonlinearities are the subject
of future work. The method is tested on a CSTR application that includes three
grades over a 48 hour time horizon with 6 minute time intervals. The embed-
ded controller simulates realistic transition times between each of the products.
The scheduling objective determines the order and quantity of production at
each grade even with half-rate reduction during peak electricity demand. The
formulation is sufficiently fast enough, and includes enough process dynamics,
to be utilized in on-line control. This presents a fully unified optimization that
fulfills the roles of, and can replace, both control and scheduling for a compa-
orable system. Although the method is demonstrated on the CSTR application,
this formulation can be applied to other systems by replacing the model, pricing
structure, and constraints of the scheduler.

Acknowledgments

Financial support from the NSF Award 1547110, EAGER: Cyber-Manufacturing
with Multi-echelon Control and Scheduling, is gratefully acknowledged.

Appendix A. Combined Scheduling and Control Example

The following APMonitor model and Python script detail the variables,
equations, and commands necessary to reproduce the combined scheduling and
control presented in Section 4.1. The application uses two elements including
the model file (schedule.apm) and a data file (schedule.csv). The data file is
a list of times between 0 and 7 with time increment 0.1 and another column
labeled last that is 0 everywhere except the end point as 1. The parameter last
is to enforce the constraints that a certain amount of each product should be
produced. All source files are available from https://github.com/APMonitor.

Listing 1: Constants and Parameters

<table>
<thead>
<tr>
<th>Constants</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 3</td>
<td>% products</td>
</tr>
</tbody>
</table>

25
5 % transition points for steps
5 b[1] := 0
5 b[2] := 1
5 b[3] := 2
5 b[4] := 3
5 b[5] := 4
5 b[6] := 5
5 % step up (+) or down (-)
5 e[g][1] := 1
5 e[g][2] := -1
5 e[g][3] := 1
5 e[g][4] := -1
5 e[g][5] := 1
5 e[g][6] := -1
5 % magnitude of step function
5 a[1] := 1
5 a[2] := 1
5 a[3] := 2
5 a[4] := 2
5 a[5] := 3
5 a[6] := 3
5 % demand for each product
5 d[1] := 2
5 d[2] := 5
5 d[3] := 3
5 % flow rate
5 q := 2
5 % manipulated variable
5 u := 0.0 >= 0.0 <= 8.0
5 % zero everywhere except last point
5 last := 0

Listing 2: Variables and Equations

5 Variables
5 x := 0.0
5 % individual step functions
5 v[1:2*n] >= 0 , <= 1
5 % slack variables
5 % should be positive when x - b is negative
5 x[1:2*n] >= 0 , <= 1000
5 % should be positive when x - b is positive
5 x2[1:2*n] >= 0 , <= 1000
5 % profit function
5 p[0] := 0 >= 0 <= 3
5 % total profit at each time step
5 % which is product is being produced
5 prod[1:n] := 0
5 % integral of product
5 iprod[1:n] := 0

5 Intermediates
5 % sum steps
5 z[0] := 0
5 z[1:2*n] := z[0:2*n-1] + e[g][1:2*n] * a[1:2*n] + u[1:2*n]
E quations

\[ pfc_n = z[2+n] \]
\[ profit = pfc_n \times q \]
\[ \frac{d(x)}{dt} = -x + u \]
\[ \frac{d(iprod[1:n])}{dt} = \prod[1:n] \times q \]
\[ s - b[1:2+n] = x[1:2+n] - z[1:2+n] \]
\[ last + (iprod[1:n] - d[1:n]) >= 0 \]
\[ % include as alternative to objective version \]
\[ %s1[1:2+n] \times (1-w[1:2+n]) <= 0 \]
\[ %s2[1:2+n] \times (1-w[1:2+n]) <= 0 \]
\[ minimize \ 10000 \times s1[1:2+n] \times (1-w[1:2+n]) \]
\[ minimize \ 10000 \times s2[1:2+n] \times (1-w[1:2+n]) \]

Python Script for Combined Control and Scheduling

```python
from apm import *
s = 'http://byu.apmonitor.com'
a = 'products'
# set up problem
apm(s, a, 'clear all')
apm_load(s, a, 'schedule.apm')
csv_load(s, a, 'schedule.csv')
# create manipulated variable
apm_info(s, a, 'MV', 'u')
apm_options(s, a, 'u status', 1)
apm_options(s, a, 'u descent', 1.0)
apm_options(s, a, 'u dmax', 0.16)
apm_options(s, a, 'u upper', 8.0)
apm_options(s, a, 'u lower', 0.0)
# set options
apm_options(s, a, 'nlc imode', 6)
apm_options(s, a, 'nlc max_iter', 200)
apm_options(s, a, 'nlc solver', 3)
apm_options(s, a, 'nlc nodes', 2)
# solve combined scheduling and control
output = apm(s, a, 'solve')
print(output)
# retrieve solution
sol = apm_sol(s, a)
```

References


URL http://dx.doi.org/10.1016/j.ces.2013.07.039


URL http://dx.doi.org/10.1016/j.jprocont.2014.03.010


URL http://www.sciencedirect.com/science/article/pii/S0098135414000052


URL http://pubs.acs.org/doi/10.1021/acs.iecr.5b00909


URL http://dx.doi.org/10.1016/j.compchemeng.2013.12.001


URL http://dx.doi.org/10.1016/j.compchemeng.2012.06.035


URL http://doi.wiley.com/10.1002/aic.13738


[29] S. M. Safdarnejad, J. D. Hedengren, L. L. Baxter, Dynamic optimization of
a hybrid system of energy-storing cryogenic carbon capture and a baseline power generation unit, Applied Energy 172 (2016) 66–79.


