A Bilevel NLP Sensitivity-based Decomposition for Dynamic Optimization with Moving Finite Elements

Weifeng Chen
College of Information Engineering, Zhejiang University of Technology, Hangzhou, P.R. China

Zhijiang Shao
Dept. of Control Science and Engineering, Zhejiang University, Hangzhou, P.R. China

Lorenz T. Biegler
Dept. of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213

DOI 10.1002/aic.14339
Published online January 16, 2014 in Wiley Online Library (wileyonlinelibrary.com)

Optimal control has guided numerous applications in chemical engineering, and exact determination of optimal profiles is essential for operation of separation and reactive processes, and operating strategies and recipe generation for batch processes. Here, a simultaneous collocation formulation based on moving finite elements is developed for the solution of a class of optimal control problems. Novel features of the algorithm include the direct location of breakpoints for control profiles and a termination criterion based on a constant Hamiltonian profile. The algorithm is stabilized and performance is significantly improved by decomposing the overall nonlinear programming (NLP) formulation into an inner problem, which solves a fixed element simultaneous collocation problem, and an outer problem, which adjusts the finite elements based on several error criteria. This bilevel formulation is aided by a NLP solver (the interior point optimizer) for both problems as well as an NLP sensitivity component, which provides derivative information from the inner problem to the outer problem. This approach is demonstrated on 11 dynamic optimization problems drawn from the optimal control and chemical engineering literature. © 2014 American Institute of Chemical Engineers *AIChE J*, 60: 966–979, 2014
Keywords: optimization, process control, numerical solutions

Introduction

Dynamic optimization and optimal control have guided numerous applications of chemical engineering. In *AIChE Journal* alone, over 200 studies have appeared since the publication of Pontryagin's classical contribution.[1] Optimal control applications abound in the development of optimal trajectories for separation processes, including distillation[2–4] and crystallization,[5–7] determination of optimal material and energy inputs, temperature and catalyst loading profiles in continuous reactors[8,9] and operating strategies and recipe generation for batch processes.[10–12] Once these optimal profiles are obtained, they serve as essential targets and reference trajectories to assess process performance, treatment of uncertainty and variability, and timely adaptation to real-time data. Optimal control strategies are challenged by the need to capture discontinuities (breakpoints) in control profiles exactly, and handle constraints on state and control profiles. Also, for singular problems, control profiles cannot be determined directly from the optimality conditions, and additional conditions need to be developed.

Most optimal control/dynamic optimization problems do not have an analytical solution. Consequently, at the heart of optimal control strategies is the need to determine accurate optimal trajectories for complex system models with efficient and reliable nonlinear programming (NLP) methods. In this study, we develop an efficient strategy for the solution of a class of optimal control problems that leads to numerically exact input trajectories.

For dynamic optimization and optimal control, simultaneous collocation methods have seen considerable development over the last two decades. Here, both the discretized differential-algebraic equation (DAE) model and the optimal control problem are formulated as a single NLP problem. The DAEs are discretized to a set of algebraic equations over finite elements in time, using, for instance, implicit Runge–Kutta (IRK) methods for the discretization. With the development of large-scale NLP solvers, direct transcription has been applied to challenging dynamic optimization problems. A key feature of simultaneous collocation is that the embedded solution of the DAE model is no longer required. Instead, exact first and second derivative information for the NLP are readily obtained from the discretized algebraic formulation. Consequently, time-consuming sensitivity calculations tied to the DAE solver are avoided and failures of the embedded DAE solver are bypassed. This issue is especially important for DAE systems that have unstable forward modes. Finally, the large-scale NLP formulation is often sparse and structured, and flexible decomposition strategies can be applied for efficient solution of the NLP.

Correspondence concerning this article should be addressed to L. T. Biegler at biegler@cmu.edu.

An important issue in the development of simultaneous collocation methods is the choice of the finite element mesh. This choice is guided by two objectives. First, the finite element mesh is selected to ensure accuracy of the DAE model. Given a suitably accurate IRK discretization, accurate solutions to a (semiexplicit index-1) DAE can be obtained with a sufficiently fine mesh. Second, a mesh must also be selected that determines the optimal location of the control variable breakpoints. In Ref. 13, mesh selection and successive refinement is performed through the solution of a sequence of NLPs, each with an equally spaced fixed mesh. This approach leads to a well-defined set of optimization problems, but it can be slow in the determination of breakpoints, as mesh refinement may require many more finite elements than needed to determine accurate solution profiles. This can be especially expensive if structural discontinuity is not built into discretization of the control profiles.

This study explores an alternate simultaneous collocation strategy where moving finite elements are embedded within the solution of the optimization problem. The approach also incorporates a mesh refinement stage, but it contains two important features: direct determination of the breakpoints as part of the finite element mesh and a stopping criterion based on the profile of the Hamiltonian function. As noted in previous work,[14–19] handling moving finite elements provides some challenges. Because a variable mesh introduces additional nonlinearity into the equality constraints, and additional nonconvexity into the NLP formulation, we explore a bilevel decomposition strategy with an inner optimization problem that deals with a fixed mesh, and an outer optimization problem that adjusts the finite elements. This bilevel approach leverages NLP sensitivity capabilities for the inner problem and allows for flexible problem formulations in the outer problem for mesh placement. As a result, it decomposes a poorly posed problem into two parts: a large-scale problem that is easier to solve, and a smaller, more difficult problem that can be tailored to specific problem classes and challenges.

The next section provides a detailed introduction of our direct transcription approach as well as additional conditions for moving finite elements. The section Decomposition for Moving Finite Element Optimization Strategy then develops the bilevel decomposition strategy along with mesh initialization. The sections Optimal Control of Batch Reactors and Optimal Control of Modified Singular Problems present numerical results for batch reactor optimization and modified singular control problems, respectively, that demonstrate the effectiveness of the bilevel approach. The section Conclusions and Future Directions concludes the article and offers directions for future work.

NLP Formulations and Solution

We describe the class of dynamic optimization problems for this study as the following multiperiod representation over a number of variable time periods, $l=1,\ldots,N, t\in[t_{l-1}, t_l]$ with $t_N=t_f$

$$\min\ \Phi(z^N(t_f)) \tag{1a}$$

$$\text{s.t.}\quad \frac{dz^1}{dt}=f(z^1(t),y^1(t),u^1(t)),\ z^1(t_0)=z_0 \tag{1b}$$

$$\frac{dz^l}{dt}=f(z^l(t),y^l(t),u^l(t)),\ z^l(t_{l-1})=z^{l-1}(t_{l-1}),\ l=2,\ldots,N \tag{1c}$$

$$g(z^l(t),y^l(t),u^l(t))=0,\ l=1,\ldots,N \tag{1d}$$

$$u_L\le u^l(t)\le u_U, t\in(t_{l-1},t_l],\ l=1,\ldots,N \tag{1e}$$

$$\psi(z^N(t_f))\le 0 \tag{1f}$$

The DAE model 1b–1d is given in semiexplicit form and is assumed to be index-1. Also, we assume that the differential and algebraic state profiles $z^l(t)$ and $y^l(t)$, respectively, and control profiles $u^l(t)$ are bounded, sufficiently differentiable functions of time within each time period, l with $t\in(t_{l-1},t_l)$, and $z^l(t)$ is required to be continuous at t_l with $l=1,\ldots,N-1$. Constraints on state profiles are not considered in this study. Also, we assume, without loss of generality that t_l can be chosen so that the solution of Problem (1) has control profiles with $u_j^l(t)=u_{L,j}, u_j^l(t)=u_{U,j}$ or $u_{L,j}<u_j^l(t)<u_{U,j}$, $t\in(t_{l-1},t_l), j=1,\ldots n_u$. With this assumption, all touchpoints (at the bounds of $u(t)$) are located only at t_l.

For Problem (1) we define the Hamiltonian function in each period as

$$H^l(t)\equiv\lambda^l(t)^T f(z^l(t),y^l(t),u^l(t))+\eta(t)^T g(z^l(t),y^l(t),u^l(t)) +\alpha_l^L(t)^T(u(t)-u_L)+\alpha_l^U(t)^T(u_U-u(t)) \tag{2}$$

where $\lambda^l,\eta^l,\alpha_l^U$, and α_l^L are the adjoint profiles. As a result, the necessary conditions of optimality (NCO) for (1) can be written as

$$\frac{d\lambda^l}{dt}=-\frac{\partial H^l}{\partial z^l},\ l=1,\ldots,N \tag{3a}$$

$$\lambda^N(t_f)=\frac{\partial\Phi(t_f)}{\partial z}+\frac{\partial\psi(t_f)}{\partial z}\eta_f \tag{3b}$$

$$\lambda^l(t^l)=\lambda^{l+1}(t^l), l=1,\ldots,N-1 \tag{3c}$$

$$\frac{\partial H^l}{\partial y^l}=0,\frac{\partial H^l}{\partial u^l}=0,\ l=1,\ldots,N \tag{3d}$$

$$0\le\alpha_l^L(t)\perp(u(t)-u_L)\ge 0,\quad 0\le\alpha_l^U(t)\perp(u_U-u(t))\ge 0 \tag{3e}$$

$$0\le\eta_f\perp\psi(t_f)\le 0 \tag{3f}$$

along with the additional constraints (1b)–(1f). As Problem (1) is not an explicit function of t, it is well-known that the Hamiltonian is continuous at t_l, even if $u(t)$ is not, and constant over time,[20] that is

$$H^l(t_l)=H^{l+1}(t_l) \tag{4a}$$

$$H^l(t)=H^N(t)=\bar{H},\ l=1,\ldots,N-1 \tag{4b}$$

Simultaneous collocation

To define our simultaneous collocation formulation, we introduce an orthogonal collocation formulation based on Lagrange polynomials using either Gauss or Radau points on finite elements. Here, we choose $K+1$ interpolation points in element i and approximate the state in a given element i, as the following polynomial

$$z^K(t)=\sum_{j=0}^{K}\ell_j(\tau)z_{ij},\ \ell_j(\tau)=\prod_{k=0,\neq j}^{K}\frac{(\tau-\tau_k)}{(\tau_j-\tau_k)},\ t\in[t_{i-1},t_i],\ \tau\in[0,1] \tag{5}$$

where $t=t_{i-1}+h_i\tau, \tau_0=0, 0<\tau_j\le 1, j=1,\ldots,K$ are shifted Gauss or Radau points and h_i is the length of element i. This

polynomial representation has the desirable property that $z^K(t_{ij})=z_{ij}$, where $t_{ij}=t_{i-1}+\tau_j h_i$. For $N>1$, we also enforce continuity of the differential state profiles across element boundaries, written as

$$z_{i+1,0}=\sum_{j=0}^{K}\ell_j(1)z_{ij}, i=1,\ldots,N-1 \tag{6a}$$

$$z_f=\sum_{j=0}^{K}\ell_j(1)z_{Nj},\ z_{1,0}=z_0 \tag{6b}$$

In addition, the control variables and algebraic states can also be represented by Lagrange interpolation profiles at K points

$$u^K(t)=\sum_{j=1}^{K}\bar{\ell}_j(\tau)u_{ij}, y^K(t)=\sum_{j=1}^{K}\bar{\ell}_j(\tau)y_{ij},\quad \text{where } \bar{\ell}_j(\tau)=\prod_{k=1,\neq j}^{K}\frac{(\tau-\tau_k)}{(\tau_j-\tau_k)} \tag{7}$$

Substituting (5) and (7) into the DAEs in Problem (1) leads to the collocation equations written as

$$\sum_{j=0}^{K}\dot{\ell}_j(\tau_k)z_{ij}-h_i f(z_{ik},y_{ik},u_{ik})=0 \tag{8a}$$

$$z_{i+1,0}-\sum_{j=0}^{K}\ell_j(1)z_{ij}=0 \tag{8b}$$

$$g(z_{ik},y_{ik},u_{ik})=0,\ i=1,\ldots,N,\ k=1,\ldots,K \tag{8c}$$

where $\dot{\ell}_j(\tau)=\frac{d\ell_j(\tau)}{d\tau}$. With fixed elements, h_i, the NLP formulation corresponding to the optimal control problem can be written as

$$\min\ \Phi(z_f) \tag{9a}$$

$$\text{s.t.}\ \sum_{j=0}^{K}\dot{\ell}_j(\tau_k)z_{ij}-h_i f(z_{ik},y_{ik},u_{ik})=0 \tag{9b}$$

$$g(z_{ik},y_{ik},u_{ik})=0 \tag{9c}$$

$$u_L\le u_{ik}\le u_U,\ k=1,\ldots,K,\ i=1,\ldots,N \tag{9d}$$

$$z_{i+1,0}=\sum_{j=0}^{K}\ell_j(1)z_{ij}, i=1,\ldots,N-1 \tag{9e}$$

$$z_f=\sum_{j=0}^{K}\ell_j(1)z_{Nj},\ z_{1,0}=z(t_0) \tag{9f}$$

$$\psi(z_f)\le 0 \tag{9g}$$

Using collocation equations in each element, along with continuity of the differential variables and breakpoint locations for the controls, once can relate the periods in (1) indexed by l to the finite elements indexed by i in (9). Therefore, these two problems are essentially the same as long as the collocation approximation is a sufficiently accurate approximation of the DAE system (1b)–(1d). As this is a fully IRK method, order results for the approximation error are well known and can be enforced with the addition of error constraints discussed in the next subsection. Moreover, the NLP (9) provides an accurate approximation to (1). Therefore, if we maintain h_i sufficiently small, the multipliers and the controls with Radau collocation converge to the true solution at the rate of $O(h^K)$, while the Lagrange multipliers scaled by ω_j provide estimates of the adjoint profiles for Problem (1).[21]

Treatment of finite elements

An important concern in the solution of the simultaneous collocation Problem (9) is the appropriate selection of element lengths $h_i=t_i-t_{i-1}$. Clearly, if h_i is fixed in the NLP, then the resulting problem is generally easier to solve. In fact, for a linear DAE system, fixing h_i, leads to a linearly constrained NLP.

On the other hand, treating h_i as variables in (9), where they can "move" during the solution of the NLP, may provide a more accurate approximation to (1) with fewer finite elements. Provided that values of h_i remain sufficiently small, these variables can locate breakpoints in the optimal control profile along with profile segments (i.e., periods) with active bounds. Moreover, for Runge–Kutta methods state profiles need to remain smooth only within a finite element; this allows Problem (9) to deal with breakpoints at finite element boundaries in a well-defined way.

Placement of finite elements is often done through comparison with numerically integrated state profiles. This concept guides the initialization step of our overall moving element strategy as well. In addition, moving finite element formulations have been described extensively in Refs. 22 and 23 for the solution of boundary value problems. Assuming that the state profiles are smooth over the domain of interest, the global error from the polynomial approximation, $e(t)=z(t)-z^K(t)$, can be estimated from

$$\max_{t\in[0,t_f]}||e(t)||\le C_1\max_{i\in\{1,\ldots,N\}}\left(\max_{t\in[t_{i-1},t_i]}||T_i(t)||\right)+O(h_i^{K+p}) \tag{10}$$

where $p=2$ is for Gauss–Legendre points and $p=1$ for Radau points, C_1 is a computable constant and $T_i(t)$ can be computed from the polynomial solution. Choices for $T_i(t)$ are reviewed in Russell and Christensen.[23] In particular, we may compute the residual DAEs at noncollocation points, $t_{i,\mathrm{nc}}$ as follows

$$T_i(t_{i,\mathrm{nc}})=\begin{bmatrix}\frac{dz^K}{d\tau}(t_{i,\mathrm{nc}})-h_i f(z^K(t_{i,\mathrm{nc}}),y^K(z(t_{i,\mathrm{nc}})),u^K(t_{i,\mathrm{nc}}))\\ g(z^K(t_{i,\mathrm{nc}}),y^K(z(t_{i,\mathrm{nc}})),u^K(t_{i,\mathrm{nc}}))\end{bmatrix} \tag{11}$$

With $t_{i,\mathrm{nc}}=t_{i-1}+h_i\tau_{i,\mathrm{nc}}$, $\tau_{i,\mathrm{nc}}\in[0,1]$ this leads to $||e_i(t)||\le\bar{C}\,||T_i(t_{i,\mathrm{nc}})||$ with the constant $\bar{C}$ given by

$$\bar{C}=\frac{1}{A}\int_0^{\tau_{i,\mathrm{nc}}}\prod_{j=1}^{K}(s-\tau_j)ds, A=\prod_{j=1}^{K}(\tau_{i,\mathrm{nc}}-\tau_j)$$

as shown in Ref. 23. With this estimate, N sufficiently large and a user-specified error tolerance ε, appropriate values of h_i can be determined by adding the error constraints (12) to Problem (9)

$$\sum_{i=1}^{N}h_i=t_f, h_i\ge 0 \tag{12a}$$

$$\bar{C}\,||T_i(t_{i,\mathrm{nc}})||\le\varepsilon \tag{12b}$$

Because these constraints may be severely nonlinear, successful implementation within a simultaneous collocation approach requires careful solution strategies, as considered in Refs. 14–19.

In addition, we consider two further restrictions on the control profile approximations given by (7). First, a general

polynomial representation of the control profile may violate its bounds between collocation points. Second, the set of active bounds may change within a given element, and this may lead to nonsmoothness in the state profiles within that element, and poor approximations using (5). These problems are avoided if we represent the control profile in each element i as a monotonic polynomial, $\sigma(\tau, v_i)$ with $u_{ik}=\sigma(\tau_k, v_i)$ and v_i as a vector of decision variables that define the polynomial, so that the extremes of $\sigma(\tau, v)$ are at the element boundaries. This choice includes control representations that are rational or exponential functions. For this study, we choose only constant or linear representations within each element. The combined NLP formulation can, therefore, be completed by adding the constraints (12) and $u_{ik}=\sigma(\tau_k, v_i)$ to (9).

Finally, we note from Refs. 15 and 21, that the Karush-Kuhn-Tucker (KKT) multipliers from the solution of (9) can be related to discretized approximations of adjoint profiles, $\lambda(t_{ik})$ and $\eta(t_{ik})$. In particular, the quantities $\bar{\lambda}_{ik}=\omega_k\lambda_{ik}$ and $\bar{\eta}_{ik}=\omega_k h_i\eta_{ik}$ are the multipliers for (9b) and (9c), respectively, where $\omega_k=\int_0^1 \bar{\ell}_k(\tau)d\tau$ is the quadrature weight.

We could also develop criteria similar to (12b) to monitor whether $\lambda(t), \eta(t)$ and NCO are approximated accurately. Here, we develop an error criterion that corresponds to (3) by eliminating the bound multipliers. We define polynomial approximations to the adjoint profiles as follows

$$\lambda^K(t)=\sum_{j=0}^{K}\ell_j(\tau)\lambda_{ij}, \eta^K(t)=\sum_{j=1}^{K}\bar{\ell}_j(\tau)\eta_{ij}$$

$$\tilde{H}^K(t)=\lambda^K(t)^T f(z^K(t), y^K(t), u^K(t))+\eta^K(t)^T g(z^K(t), y^K(t), u^K(t))$$

Evaluating these criteria at noncollocation points $t_{i,\mathrm{nc}}$ leads to

$$S_i(t_{i,\mathrm{nc}})=\begin{bmatrix} \dfrac{d\lambda^K}{d\tau}(t_{i,\mathrm{nc}})+h_i\dfrac{\partial\tilde{H}^K}{\partial z}(t_{i,\mathrm{nc}}) \\ \dfrac{\partial\tilde{H}^K}{\partial y}(t_{i,\mathrm{nc}}) \\ \min\left[\left|\left|\dfrac{\partial\tilde{H}^K}{\partial u}(t_{i,\mathrm{nc}})\right|\right|, u_U^l-u(t_{i,\mathrm{nc}}), u(t_{i,\mathrm{nc}})-u_L^l\right] \end{bmatrix} \tag{13}$$

From this condition, we can impose the error constraint on the optimality conditions

$$\bar{C}\,||S_i(t_{i,\mathrm{nc}})|| \le \varepsilon \tag{14}$$

The mesh refinement difficulties posed by breakpoint selection and satisfaction of (12), (14) lead us to consider a bilevel optimization strategy. Here, we consider an inner optimization problem with a fixed mesh, whereas the outer problem considers the placement of the finite elements. This is similar to some of our previous studies. In Refs. 17 and 18, a bilevel strategy was proposed that uses (12), but NLP solvers and sensitivity were limited. This led to a nondifferentiable outer problem along with a specialized bundle solver. Conversely, in Ref. 14, a bilevel strategy was proposed where only the collocation equations were considered in the inner problem, whereas the outer problem was solved in the space of the controls and finite elements. Finally, in Refs. 15 and 16, the overall problem was solved in full space with a barrier method, and refinement of the finite elements was made along with adjustment of the barrier parameter. In this way, a bilevel strategy was formulated naturally as part of the nested solution of the barrier problem.

In this study, we extend our previous studies to a bilevel approach enabled by a barrier NLP solver [interior point optimizer (IPOPT)[24]] that is coupled seamlessly to an NLP sensitivity strategy (sIPOPT[25]). Additionally, we incorporate Hamiltonian-based constraints in the outer problem as well as efficient sensitivity calculation of the inner problem, for both the state and adjoint variables.

Decomposition for Moving Finite Element Optimization Strategy

To develop the bilevel strategy, we first generate a feasible mesh that is consistent with the discretized optimization problem and feasible for a specified level of approximation error. For this, we begin by considering two initialization approaches, one based on single shooting and one based on multiple shooting.

Mesh initialization

In the single shooting initialization, we generate the initial mesh by fixing $u(t)$ to initial (guessed) values $\bar{u}(t)$ and solving a sequence of NLPs indexed by $i=1,2,3,\ldots$ given by

$$\bar{h}_i=\arg\{\max\ h_i \tag{15a}$$

$$\text{s.t.}\ \sum_{j=0}^{K}\dot{\ell}_j(\tau_k)z_{ij}-h_i f(z_{ik}, y_{ik}, \bar{u}_{ik})=0 \tag{15b}$$

$$z_{i+1,0}=\sum_{j=0}^{K}\ell_j(1)z_{ij} \tag{15c}$$

$$g(z_{ik}, y_{ik}, \bar{u}_{ik})=0,\ k=1,\ldots,K \tag{15d}$$

$$-\varepsilon \le \bar{C}\,T_i(t_{i,\mathrm{nc}}) \le \varepsilon \tag{15e}$$

$$0 \le h_i \le \min\left(h_{\max}, t_f-\sum_{i'=1}^{i-1}\bar{h}_{i'}\right) \tag{15f}$$

where $z_{1,0}=z_0$. This approach mimics an initial value DAE solver with time steps $\bar{h}_i$, and leads to a feasible initialization for (9). It is particularly suitable for DAEs with stable forward modes, and is our default strategy as long as the state variables remain bounded.

The multiple shooting initialization is required if the state variables do not remain bounded in the single shooting strategy. Here, we select a suitable bounded region (often problem dependent) for the states $z(t)\in\mathcal{Z}$, apply the single shooting strategy and terminate the sequence of NLPs (15) when $z_{\hat{i}+1,0}\notin\mathcal{Z}$ for $\hat{i}>1$. We then define $N_p=\lceil t_f/(\sum_{i=1}^{\hat{i}-1}h_i)\rceil$ time periods and replicate the state profiles from the first $\hat{i}-1$ elements over the elements $i=j(\hat{i}-1)+k$ with $j=1,\ldots,N_p, k=1,\ldots,\hat{i}-1$. These profiles, with $N\le N_p(\hat{i}-1)$ elements then remain feasible and bounded in each period.

The inner problem

For the inner level of our optimization strategy, we rewrite (9) as the following NLP with h_i fixed to $\bar{h}_i$ and the approximation error constraints (12b) replaced by placeholder variables T_i, that is

$$\min\ \ \Phi(z_f) \tag{16a}$$

$$\text{s.t.} \quad \sum_{j=0}^{K} \dot{\ell}_j(\tau_k) z_{ij} - \bar{h}_i f(z_{ik}, y_{ik}, u_{ik}) = 0 \tag{16b}$$

$$g(z_{ik}, y_{ik}, u_{ik}) = 0 \tag{16c}$$

$$u_{ik} = \sigma(\tau_k, v_i) \tag{16d}$$

$$u_L \le u_{ik} \le u_U,\ k \in \{1, \ldots, K\},\ i \in \{1, \ldots, N\} \tag{16e}$$

$$z_{i+1,0} = \sum_{j=0}^{K} \ell_j(1) z_{ij},\ i = 1, \ldots, N-1 \tag{16f}$$

$$z_f = \sum_{j=0}^{K} \ell_j(1) z_{Nj},\ z_{1,0} = z(t_0) \tag{16g}$$

$$\psi(z_f) \le 0 \tag{16h}$$

$$T_i = \begin{bmatrix} \dfrac{dz^K}{d\tau}(t_{i,\mathrm{nc}}) - h_i f(z^K(t_{i,\mathrm{nc}}), y^K(z(t_{i,\mathrm{nc}})), u^K(t_{i,\mathrm{nc}})) \\ \\ g(z^K(t_{i,\mathrm{nc}}), y^K(z(t_{i,\mathrm{nc}})), u^K(t_{i,\mathrm{nc}})) \end{bmatrix} \tag{16i}$$
$$i = 1, \ldots, N$$

This problem has a feasible starting point due to the above initialization strategy. Also, the constraints (16i) are easy to satisfy because the placeholder variables T_i are not constrained, but merely used to represent the error constraints used in the outer problem. The optimality conditions of this problem are given by the constraints (16a)–(16i) as well as the following stationarity conditions

$$z_f: \quad \frac{\partial \Phi}{\partial z_f} + \frac{\partial \psi}{\partial z_f} \gamma_f + v_f = 0 \tag{17a}$$

$$0 \ge \psi(z_f) \perp \gamma_f \ge 0 \tag{17b}$$

$$z_{i0}: \quad -\sum_{k=1}^{K} \dot{\ell}_0(\tau_k) \bar{\lambda}_{ik} + (v_i - \ell_0(1) v_{i+1}) = 0,\ i = 1, \ldots, N-1 \tag{17c}$$

$$z_{N0}: \quad -\sum_{k=1}^{K} \dot{\ell}_0(\tau_k) \bar{\lambda}_{Nk} + (v_N - \ell_0(1) v_f) = 0 \tag{17d}$$

$$\begin{aligned} z_{ij}: \quad & -\sum_{k=1}^{K} \dot{\ell}_j(\tau_k) \bar{\lambda}_{ik} + \bar{h}_i \frac{\partial f}{\partial z_{ij}} \bar{\lambda}_{ij} + \frac{\partial g}{\partial z_{ij}} \bar{\eta}_{ij} \\ & - \ell_j(1) v_{i+1} = 0,\ i = 1, \ldots, N-1 \end{aligned} \tag{17e}$$

$$z_{Nj}: \quad -\sum_{k=1}^{K} \dot{\ell}_j(\tau_k) \bar{\lambda}_{Nk} + \bar{h}_i \frac{\partial f}{\partial z_{Nj}} \bar{\lambda}_{Nj} + \frac{\partial g}{\partial z_{Nj}} \bar{\eta}_{Nj} - \ell_j(1) v_f = 0 \tag{17f}$$

$$y_{ij}: \quad \bar{h}_i \frac{\partial f}{\partial y_{ij}} \bar{\lambda}_{ij} + \frac{\partial g}{\partial y_{ij}} \bar{\eta}_{ij} = 0 \tag{17g}$$

$$u_{ij}: \quad \bar{h}_i \frac{\partial f}{\partial u_{ij}} \bar{\lambda}_{ij} + \frac{\partial g}{\partial u_{ij}} \bar{\eta}_{ij} - \alpha_{ij}^L + \alpha_{ij}^U + \zeta_{ij} = 0 \tag{17h}$$

$$0 \le u_{ij} - u_L \perp \alpha_{ij}^L \ge 0, 0 \le u_U - u_{ij} \perp \alpha_{ij}^U \ge 0 \tag{17i}$$

$$v_i: \quad \sum_{j=1}^{K} \frac{\partial \sigma(\tau_j, v_i)}{\partial v_i} \zeta_{ij} = 0. \tag{17j}$$

Alternately, ζ_{ij} can be eliminated, and (17h) and (17j) can be combined to form

$$\sum_{j=1}^{K} \frac{\partial \sigma(\tau_j, v_i)}{\partial v_i} \left(\bar{h}_i \frac{\partial f}{\partial u_{ij}} \bar{\lambda}_{ij} + \frac{\partial g}{\partial u_{ij}} \bar{\eta}_{ij} - \alpha_{ij}^L + \alpha_{ij}^U \right) = 0 \tag{18}$$

Also, note that the error definition (16i) does not appear in the optimality condition as the corresponding multipliers on these equations (and variables T_i) equal zero.

Solution and Sensitivity of the Inner Problem

To solve the inner problem, we apply an efficient barrier solver, such as IPOPT. A key advantage of this solver is that NLP sensitivity information with respect to the element lengths h is generated automatically. Here, we represent the inner problem (16) with fixed element lengths as

$$\min_x f(x) \tag{19a}$$

$$\text{s.t.}\ c(x; \bar{h}) = 0,\ x \ge 0 \tag{19b}$$

and we obtain the solution to this problem by solving the following sequence of problems with $\mu_\ell \to 0$

$$\min_x f(x) - \mu_\ell \sum_{i=1}^{n_x} \ln(x^{(i)}) \tag{20a}$$

$$\text{s.t.}\ c(x; \bar{h}) = 0 \tag{20b}$$

Moreover, we need not solve the inner problem completely by letting $\mu_\ell \to 0$. Instead, a sequence of solutions to Problem (20) with μ sufficiently small still provides the following properties:

Theorem 1. *(Properties of the central path/barrier trajectory)*[26] *Consider Problem (19) with $f(x;h)$ and $c(x;h)$ at least twice differentiable in x and once in h. Let x^* be a local constrained minimizer of (19) with the following sufficient optimality conditions at x^*:*

- *x^* is a KKT point.*
- *the Mangasarian–Fromowitz constraint qualification and strict complementarity hold at x^*, for some bound multiplier v^* that satisfies KKT conditions.*
- *the strong second-order sufficient conditions (SSOSC) hold at the KKT point.*

If we now solve a sequence of barrier problems (20) with $\mu_\ell \to 0$, then:

1. *There is at least one subsequence of unconstrained minimizers $(x(\mu_\ell))$ of the barrier function converging to x^*.*
2. *For every convergent subsequence, the corresponding sequence of barrier multiplier approximations is bounded and converges to multipliers satisfying the KKT conditions for x^*.*
3. *A unique, continuously differentiable vector function $x(\mu)$ of the minimizers of (20) exists for $\mu > 0$ in a neighborhood of $\mu = 0$.*
4. $\lim_{\mu \to 0^+} x(\mu) = x^*$
5. $||x(\mu) - x^*|| = O(\mu)$

Having the primal–dual solution vector $s(\mu, \bar{h}) = (x(\mu, \bar{h}), \vartheta(\mu, \bar{h}))$ of (20), we now require its sensitivity to changes in the mesh h. For this, we apply the following property.

Theorem 2. *(Sensitivity Properties)*[27] *For Problem (20) assume that $f(x;h)$ and $c(x;h)$ are k times differentiable in h and $k+1$ times differentiable in x. Also, assume $x(\mu; \bar{h})$ is a KKT point of (20), where linearly independent constraint qualification (LICQ) and SSOSC hold.*

- *$x(\mu, \bar{h})$ is an isolated minimizer and the associated multipliers $\vartheta(\mu, \bar{h})$ are unique.*

- *For some h in a neighborhood of $\bar{h}$, there exists a k times differentiable function $s(\mu,h)^T=[x(\mu,h)^T;\vartheta(\mu,h)^T]$ that corresponds to a locally unique minimum for (20).*

This property allows us to compute the sensitivity information of x^* with respect to h, that is, $\frac{dx(\mu,\bar{h})}{dh}$. Calculation of the sensitivity of the inner problem variables with respect to h now proceeds from the implicit function theorem (IFT) applied to the optimality conditions of (20) at $\bar{h}$. As shown in Ref. 25, we define the quantities

$$M(s(\mu,\bar{h}))=\begin{bmatrix} W(s;\mu,\bar{h}) & A(x;\mu,\bar{h}) & -I \\ A(x,\mu,\bar{h})^T & 0 & 0 \\ V(\mu,\bar{h}) & 0 & X(\mu,\bar{h}) \end{bmatrix} \tag{21}$$

and

$$N_h(s(\mu,\bar{h}))=\begin{bmatrix} \nabla_{xh}L(s(\mu,\bar{h})) \\ \nabla_h c(x(\mu,\bar{h})) \\ 0 \end{bmatrix}, N_\mu=\begin{bmatrix} 0 \\ 0 \\ -\mu e \end{bmatrix} \tag{22}$$

where $e^T=[1,1,\ldots,1]$, $s(\mu,\bar{h})=[x(\mu,\bar{h}),\vartheta(\mu,\bar{h}),v(\mu,\bar{h})]^T$, $W(s(\mu,\bar{h}))$ denotes the Hessian $\nabla_{xx}L(s;h)$ of the Lagrangian function evaluated at $s(\mu,\bar{h}), A(x(\mu,\bar{h}))=\nabla_x c(x;\bar{h})$ evaluated at $x(\mu,\bar{h}), X=\text{diag}\{x\}$ and $V=\text{diag}\{v\}$. Application of IFT leads to

$$M(s(\mu,\bar{h}))\frac{ds(\mu,\bar{h})^T}{dh}+N_h(s(\mu,\bar{h}))=0 \tag{23}$$

When LICQ and SSOSC hold, $M(s(\mu,\bar{h}))$ is nonsingular and the sensitivities are calculated by sIPOPT from

$$\frac{ds(\mu,\bar{h})^T}{dh}=-M(s(\mu,\bar{h}))^{-1}N_h(s(\mu,\bar{h})) \tag{24}$$

Moreover, in IPOPT $M(s(\mu,\bar{h}))$ is directly available in factored form from the solution of (20), so the sensitivity can be calculated through a simple backsolve. Once these sensitivities are calculated, we obtain gradient information used directly in the solution of the outer problem, considered next.

Outer problem

The initialization approach in the previous section leads to feasible finite element mesh for Problem (9). However, the addition of error constraints (14) leads to a highly constrained problem that is generally difficult to converge to the optimum. Moreover, the solution of (9) may lead to suboptimal profiles if there are insufficient elements to represent the optimal solution. Consequently, a provision is needed for additional elements as the optimization progresses.

To address this issue, we choose an alternate optimality criterion based on (4b) instead of (3a)–(3f). As shown in Ref. 21, KKT multipliers for the constraints (16b), (16c) provide an $O(h^K)$ approximation of the adjoint variables. Using these multipliers as well as constraint values from (16), we approximate the Hamiltonian at the collocation points as follows

$$\begin{aligned} H_{(i-1)K+k}=H^K(t_{ik})=\lambda_{ik}^T f(z_{ik},y_{ik},u_{ik}) \\ +\eta_{ik}^T g(z_{ik},y_{ik},u_{ik})=\frac{\bar{\lambda}_{ik}^T}{\omega_k h_i}\sum_{j=0}^{K}\dot{\ell}_j(\tau_k)z_{ij} \end{aligned} \tag{25}$$

where $\lambda_{ik}=\bar{\lambda}_{ik}/\omega_k$, $\bar{\lambda}_{ik}$ is the corresponding KKT multiplier from the inner problem (16) and ω_k is the quadrature weight [Note that at t_{ik} the complementarity terms in (2) are zero and are omitted in (25)]. To ensure that the optimal control solution is sufficiently accurate, we enforce the condition that the Hamiltonian profile, represented by (25), is constant over time.

We now consider the following outer problem, where only the variables h_i remain to locate the breakpoints and satisfy the error constraints. This problem is stated as follows

$$\min_{\bar{H},h_i,w_j}\ \Phi(z_f(h))+\rho\left(\sum_{j=1}^{NK}w_j+\sum_{i=1}^{N}v_i\right) \tag{26a}$$

$$-(v_i+\varepsilon)\le \bar{C}\,T_i(h)\le \varepsilon+v_i,\ v_i\ge 0,\ i=1,\ldots,N \tag{26b}$$

$$H_{(i-1)K+k}(h)=\frac{\bar{\lambda}_{ik}(h)^T}{\omega_k h_i}\sum_{k'=0}^{K}\dot{\ell}_j(\tau_k)z_{ik'}(h) \tag{26c}$$

$$-(\varepsilon_h+w_j)\le H_j-\bar{H}\le(\varepsilon_h+w_j),\ w_j\ge 0,\ j=1,\ldots,NK \tag{26d}$$

$$0\le h_i\le h_{\max} \tag{26e}$$

$$\sum_{i=1}^{N}h_i=t_f \tag{26f}$$

where the variables $z_{ik}(h),\bar{\lambda}_{ik}(h),T_i(h)$ arise from fully converged solutions of (16) with a fixed mesh $\bar{h}$. Hence, solution of Problem (16) must be nested within the solution of (26). In solving (26) gradients with respect to $z_f(h),z_{ik}(h)$, $\bar{\lambda}_{ik}(h)$ and $T_i(h)$ are determined from the sensitivity terms $\frac{ds(\mu,\bar{h})}{dh}$, which are well-defined for $\mu>0$. Second-order sensitivities are not calculated here as Problem (26) is relatively small, and an NLP solver with quasi-Newton Hessian update is sufficient to find a solution to (26).

A key objective in (26) is to enforce a constant Hamiltonian profile (26d) using the variable $\bar{H}$. However, this constraint may require additional finite elements. As a result artificial variables w_j and v_i, and an ℓ_1 penalty, with parameter $\rho>0$ sufficiently large, are introduced to enforce a solution to the outer problem with error tolerances ε for the approximation error and ε_h for the Hamiltonian profile.

Last, we analyze the quality of the solution of (26). In addition to the constraints in (26) the KKT conditions of the outer problem are given by

$$\begin{aligned} h_i:\frac{d\Phi}{dh_i}+\sum_{j=1}^{NK}\frac{dH_j(h)}{dh_i}(\gamma_j^+-\gamma_j^-) \\ +\bar{C}\sum_{i'=1}^{N}\frac{dT_{i'}(h)}{dh_i}(\delta_{i'}^+-\delta_{i'}^-)+(\beta_i^{\max}-\beta_i^0) \\ +\bar{\gamma}=0 \end{aligned} \tag{27a}$$

$$w_j:\rho=\gamma_j^++\gamma_j^-+\beta_j^w \tag{27b}$$

$$v_i:\rho=\delta_i^++\delta_i^-+\beta_i^v \tag{27c}$$

$$\bar{H}:\sum_{j=1}^{N}(\gamma_j^--\gamma_j^+)=0 \tag{27d}$$

$$0\le w_j+\varepsilon_h+H_j-\bar{H}\perp\gamma_j^-\ge 0 \tag{27e}$$

$$0\le w_j+\varepsilon_h-H_j+\bar{H}\perp\gamma_j^+\ge 0 \tag{27f}$$

$$0\le h_i\perp\beta_i^0\ge 0 \tag{27g}$$

$$0\le h_i^{\max}-h_i\perp\beta_i^{\max}\ge 0 \tag{27h}$$

$$0\le v_i+\varepsilon+\bar{C}T_i(h)\perp\delta_i^-\ge 0 \tag{27i}$$

$$0 \le v_i + \varepsilon - \bar{C}T_i(h) \perp \delta_i^+ \ge 0 \tag{27j}$$

$$0 \le w_j \perp \beta_j^w \ge 0 \tag{27k}$$

$$0 \le v_i \perp \beta_i^v \ge 0 \tag{27l}$$

Based on this solution, the following observations lead to useful heuristics for the management of additional finite elements.

- If $\beta_j^0 > 0$, it is clear that corresponding $h_j = 0$ can be removed without changing the solution or the multipliers for the other elements.
- If $w_j > 0$, then $\beta_j^w = 0$ and $\rho = \gamma_j^+ + \gamma_j^-$, where ρ is allowed to be arbitrarily large. Also, as $0 \le \gamma_j^+ \perp \gamma_j^- \ge 0$, we see that either γ_j^+ or γ_j^- becomes very large. A similar argument can be made for δ_i^+, δ_i^- if $v_i > 0$.
- From the structure of (27a), we do not expect all multipliers to be affected by ρ, as at least some elements would be inactive. Conversely, for elements where $w_j > 0$ or $v_j > 0$, it is clear that $\beta^{\max} > 0$ and the upper bounds (26e) must be active. Consequently, indices i in (27a) with large multipliers correspond to elements that need to be bisected or supplemented with additional elements.
- Finally, if $v_i = 0$ and $w_j = 0$, and the constraints (26b) and (26d) are inactive, then we have a solution given by

$$-\varepsilon_h \le H_j - \bar{H} \le \varepsilon_h, \; -\varepsilon \le T_i \le \varepsilon, \tag{28}$$

$$\begin{aligned} \frac{d\Phi}{dh_i} + \sum_{j=1}^{NK} \frac{dH_j(h)}{dh_i}(\gamma_j^+ - \gamma_j^-) &= \frac{d\Phi}{dh_{i'}} \\ + \sum_{j=1}^{NK} \frac{dH_j(h)}{dh_{i'}}(\gamma_j^+ - \gamma_j^-), \; i, i' &= 1, \ldots, N \end{aligned} \tag{29}$$

and the elements are optimally placed.

Bilevel strategy

The resulting bilevel mesh refinement strategy is given as follows:

1. (Optional) Divide the time domain equally with a small number of fixed finite elements and solve (9) to provide a value for $h^{\max}$ and an initial guess for $u(t)$.
2. Having an initial guess $\bar{u}(t)$, solve a sequence of NLPs (15) for $i = 1, 2, 3, \ldots$, while $t_f - \sum_{i'=1}^{i-1} \bar{h}_{i'} > 0$, to determine an initial finite element mesh $\bar{h}_i$ with the single shooting approach. If state profiles become unbounded, apply the multiple shooting approach.
3. For a given error tolerance ε and (minimum) barrier parameter μ, solve the outer problem (26) with embedded solutions of the inner problem (16) and the sensitivity system (23).
4. If $\beta_i^0 > 0$, delete element i and renumber the elements.
5. If $\sum_{j=1}^{NK} w_j + \sum_{i=1}^{N} v_i = 0$, stop. Otherwise, examine the multipliers in (27a) and bisect element i that corresponds to active constraints (26e). Reinitialize the variables for Problem (16) with the newly inserted element and return to Step 3.

This bilevel optimization strategy was implemented in AMPL and IPOPT[24] was used as the NLP solver for both problems (16) and (26). For the inner problem, exact first and second derivatives were supplied by AMPL. For (26), gradients were supplied from the NLP sensitivity code sIPOPT from the solution to (16). The limited memory BFGS option in IPOPT was used to solve the outer problem. In Problem (16), three collocation points ($K = 3$) were used for the state variables and piecewise linear control profiles were applied. Also, the error constraints (16i) were applied at two noncollocation points within each element with $\varepsilon = 10^{-4}$. The inner problem was converged to a KKT tolerance of 10^{-12} to ensure accurate sensitivities. For the outer problem (26) $\varepsilon_h = 10^{-4}|\bar{H}|, \rho = 10^4$ and the KKT tolerance was set to 10^{-5}. All numerical experiments are carried out on Intel(R) Core(TM) 2 Duo CPU P8800 processors (2.66 GHz and 4.0 G RAM) running Windows 7.

Optimal Control of Batch Reactors

In this section, we apply the bilevel optimization strategy to five case study examples drawn from optimal temperature control in batch reactors. These nonsingular problems have well-defined solutions and similar examples are considered in Ref. 16. The first three problems have well-conditioned profiles that can be solved with the bilevel approach. The fourth and fifth problems exhibit steep control profiles and demonstrate how the bilevel strategy must add finite elements to obtain accurate optimal control profiles. In all cases, the single shooting initialization based on Problem (15) was able to determine the initial mesh. Then, the bilevel strategy is converged to optimal solutions within the stated tolerance.

Temperature profiles for batch reactor with Parallel kinetics

We consider a nonisothermal batch reactor with first-order parallel reactions $A \rightarrow B, A \rightarrow C$ where the goal is to find a (transformed) temperature profile that maximizes the final amount of product B after 1 h ($t_f = 1$). The optimal control problem can be stated as

$$\min \quad -b(1) \tag{30a}$$

$$\text{s.t.} \quad \frac{da}{dt} = -a(t)(u(t) + u(t)^2/2) \tag{30b}$$

$$\frac{db}{dt} = a(t)u(t) \tag{30c}$$

$$a(0) = 1, \; b(0) = 0, \; u(t) \in [0, 5] \tag{30d}$$

The initial number of finite elements for Step 1 is set to 20 and the bilevel approach increases this number to 28. This requires four solutions of the outer problem and a total of 115 IPOPT iterations for (16). The total time for solving the outer problems is 43.2 CPU seconds. The value of $\bar{H}$ from (26) is 0.1585412 and the optimal objective has a value of -0.5735449921555. The final element distribution along with the state, control, and Hamiltonian profiles are shown in Figure 1.

Temperature profiles for batch reactor with van de Vusse kinetics

The nonisothermal batch reactor with van de Vusse kinetics is an extension of the series reactions with $A \rightarrow B \rightarrow C$ and $2A \rightarrow D$, where the goal is again to find a (transformed) temperature profile that maximizes the final amount of product B after 1 h ($t_f = 1$). The optimal control problem can be stated as

$$\min \quad -b(1) \tag{31a}$$

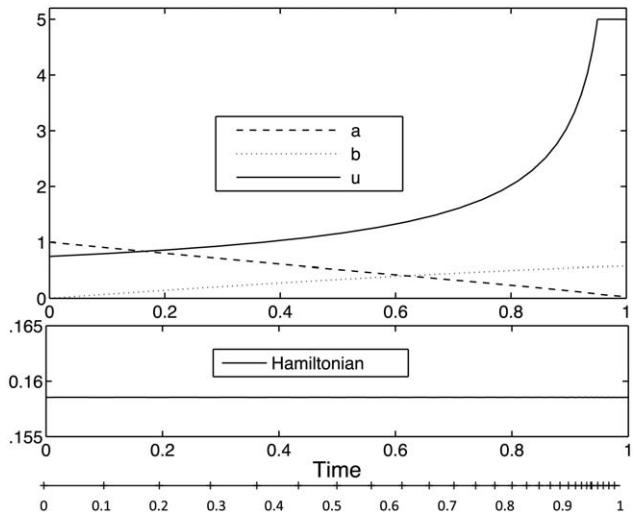

Figure 1. Control and Hamiltonian profiles for batch reactor with parallel reactions.

$$\text{s.t.} \quad \frac{da}{dt} = -(k_1 + k_3)a(t) \tag{31b}$$

$$\frac{db}{dt} = k_1 a(t) - k_2 b(t) \tag{31c}$$

$$a(0) = 1, b(0) = 0, u(t) \in [0, 0.5] \tag{31d}$$

and $k_i = \alpha_i u^{\beta_i}, i = 1, \ldots, 3$ and $\alpha = [1, 10, 1], \beta = [1, 2, 1.5]$. The initial number of finite elements for Step 1 is set to 20 and the final number increases to 21 with the bilevel approach. This requires two solutions of the outer problem and a total of 42 IPOPT iterations for (16). The total time for solving the outer problems is 11.7 CPU seconds. The value of $\bar{H}$ from (26) is 0.06390922 and the optimal objective has a value of -0.1585018546365. The final element distribution along with the state, control, and Hamiltonian profiles are shown in Figure 2.

Temperature profiles for batch reactor with Denbigh kinetics

The nonisothermal batch reactor problem for van de Vusse kinetics is now extended to deal with Denbigh kinetics with $A \to B \to C, A \to D, B \to E$ where the goal is to find a (transformed) temperature profile (transformed) temperature profile that maximizes the final amount of product B after 1 h ($t_f = 1$). The optimal control problem can be stated as

$$\min \quad -b(1) \tag{32a}$$

$$\text{s.t.} \quad \frac{da}{dt} = -(k_1 a(t) + k_3)a(t) \tag{32b}$$

$$\frac{db}{dt} = \frac{1}{2} k_1 a(t)^2 - (k_4 b(t) + k_2)b(t) \tag{32c}$$

$$a(0) = 1, b(0) = 0, \tag{32d}$$

and $k_i = \alpha_i u^{\beta_i}, i = 1, \ldots, 4$ and $\alpha = [1, 10, 1, 2], \beta = [1, 2, 1.5, 3]$. The initial number of finite elements from Step 1 was set to 20 and the final number remains at 20. The bilevel approach requires only one solution of the outer problem and a total of 11 IPOPT iterations for (16). The total time for solving the outer problems is 2.5 CPU seconds. The value of $\bar{H}$ from (26) is 0.02382238 and the optimal objective has a value of -0.06573002542291. The final element distribution along with the state, control, and Hamiltonian profiles are shown in Figure 3.

Temperature profiles for batch reactor with series kinetics

We consider a nonisothermal batch reactor with first-order series reactions $A \to B \to C$. For optimal reactor operation, we seek a temperature profile that maximizes the final amount of product B after 1 h of operation $t_f = 1$.

The optimal control problem can be stated as

$$\min \quad -b(1) \tag{33a}$$

$$\text{s.t.} \quad \frac{da}{dt} = -a(t)u(t) \tag{33b}$$

$$\frac{db}{dt} = a(t)\,u(t) - 2\,b(t)\,u(t)^2 \tag{33c}$$

$$a(0) = 1, b(0) = 0 \tag{33d}$$

$$0 \le u(t) \le 5 \tag{33e}$$

The initial number of finite elements for Step 1 is set to 20. The two-step algorithm is executed and terminates with the ℓ_1 penalty term at zero. The number of finite elements

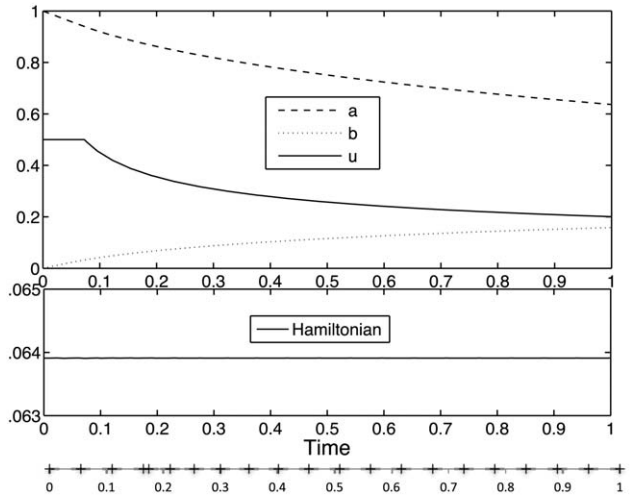

Figure 2. Control and Hamiltonian profiles for batch reactor with van de Vusse reactions.

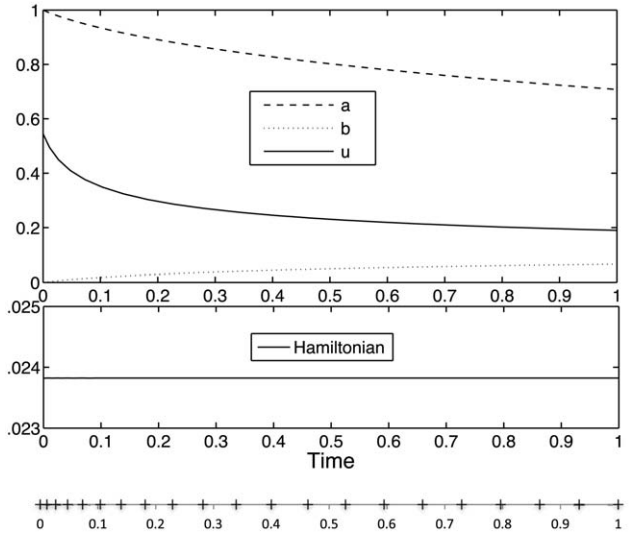

Figure 3. Control and Hamiltonian profiles for batch reactor with Denbigh reactions.

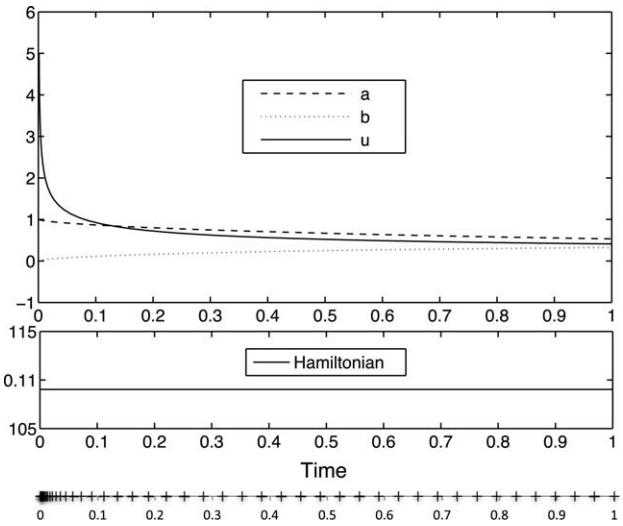

Figure 4. Control and Hamiltonian profiles for batch reactor with series reactions.

increases to 47. This requires nine solutions of the outer problem and a total of 479 iterations for the inner problems. The value of $\bar{H}$ from (26) is 0.1090224625156 and the optimal objective has a value of -0.324143015236. The total time for solving the outer problems is 439.7 CPU seconds. The final element distribution along with the state, control, and Hamiltonian profiles are shown in Figure 4. It is interesting to note the initial mesh was determined with a flat control profile; therefore, the resulting steep optimal control profile leads to considerable computation to determine the final mesh. These difficulties can be better understood by considering two additional cases

- In Step 1, we initialize the problem with 30 elements, and therefore, start with a smaller approximation error. The solution procedure is terminated with penalty term at zero and the number of finite elements increases to 31. The total time for solving the outer problems is now only 22.5 CPU seconds and the solution profiles remain the same as in Figure 4 with $\bar{H}$ from Problem (26) equal to 0.1090232175342 and the objective equal to -0.3241429813715.
- We impose an upper bound of $u(t) \leq 3.5$ and set the initial number of finite elements from Step 1 to 20. The bilevel approach increased the number of finite elements to 21 and the total effort for solving the outer problem is only 11.9 CPU seconds. Except for a slightly less steep initial control profile, the solution profiles remain virtually the same as in Figure 4, with $\bar{H}$ from Problem (26) equal to 0.1091037694239 and the objective equal to -0.32410050151.

Temperature profiles for batch reactor with Williams–Otto kinetics

We now consider a nonisothermal batch reactor with more complex reactions $A+B \rightarrow C, B+C \rightarrow P+E, C+P \rightarrow G$ as originally posed in Ref. 28. Here, the goal is again to find a (transformed) temperature profile that maximizes the final amount of product P after 1 h ($t_f = 1$). The optimal control problem can be stated as

$$\min \quad -p(1) \tag{34a}$$

$$\text{s.t.} \quad \frac{da}{dt} = -r_1 \tag{34b}$$

$$\frac{db}{dt} = -(r_1 + r_2) \tag{34c}$$

$$\frac{dc}{dt} = 2r_1 - 2r_2 - r_3 \tag{34d}$$

$$\frac{dp}{dt} = r_2 - r_3 \tag{34e}$$

$$a(0)=1, b(0)=1, c(0)=0, p(0)=0, u(t) \in [0, 100] \tag{34f}$$

where $r_1 = k_1 a(t) b(t), r_2 = k_2 b(t) c(t), r_3 = k_3 c(t) p(t), k_i = \alpha_i u^{\beta_i}, i = 1, \ldots, 3$ and $\alpha = [1, 1.21394, 0.733], \beta = [1, 1.25, 5/3]$. The initial number of finite elements for Step 1 is set to 20. Here, the algorithm is executed but the outer problem fails to meet the Hamiltonian criterion and the number of elements increases to only 23. Nevertheless, it is instructive to view one of converged inner problem solutions with the state and control profiles, and element distribution shown in Figure 5. From this figure, we see that the control profile has a very steep slope, the length of first finite element is 0.000395414, and the control value at $t = 0$ is near a large upper bound. We believe that this very steep control profile contributes to failure of the method.

To investigate this behavior, we reduce the upper bound on the control profile to $u(t) \leq 0.6$ and, again, initialize with 20 elements. With the bilevel approach, the final number of finite elements remains at 20 with only one solution of the outer problem and only six IPOPT iterations for (16). The total time for solving the outer problems is 4.6 CPU seconds. The value of $\bar{H}$ from (26) is 0.1272836433559 and the optimal objective has a value of -0.1254534249595. The final element distribution along with the state, control, and Hamiltonian profiles are shown in Figure 6.

Sensitivity to convergence parameters

As part of this numerical study, we solved an additional set of test cases to assess the sensitivity of the bilevel approach to its convergence parameters. In particular, we considered several values for $\rho > 10^4$ as well as values of ε and ε_h at 10^{-5} and 10^{-6}. The results can be summarized as follows:

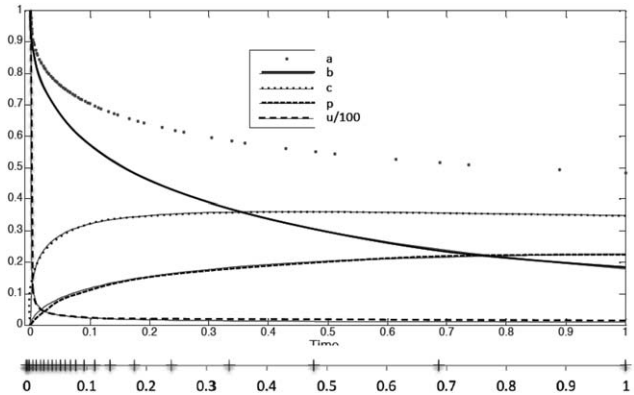

Figure 5. Control and Hamiltonian profiles for batch reactor with Williams–Otto reactions and $u(t) \leq 100$.

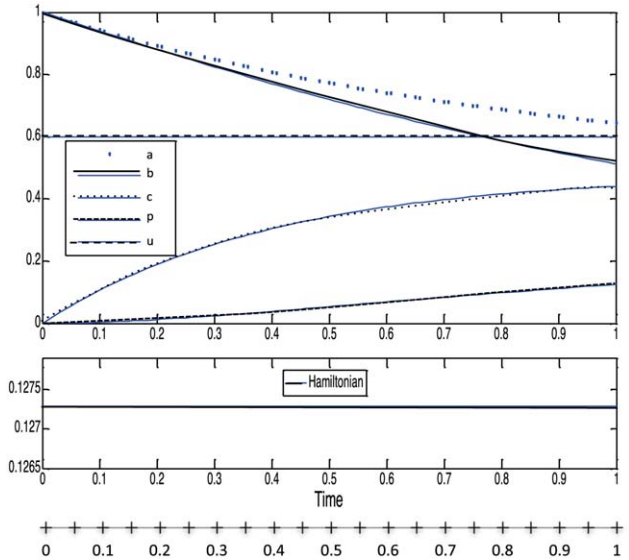

Figure 6. Control and Hamiltonian profiles for batch reactor with Williams–Otto reactions and $u(t) \leq 0.6$.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

- To demonstrate the effectiveness of the Hamiltonian criterion (26d), the NCO error criterion (14) was checked for all of the above solutions. We found that the NCO errors were all within the tolerance $\varepsilon = 10^{-4}$.
- All of the above results remain unchanged when a larger value of ρ is used, so we consider $\rho = 10^4$ to be sufficiently large. This is consistent with properties of ℓ_1 penalty functions.
- For the smaller values of ε and ε_h, the first three problems (parallel, van de Vusse, and Denbigh) solve without difficulty with solution profiles essentially the same as those presented above and (very) slight improvements in the objective functions.
- Due to the steep control profile, the series problem was more difficult to solve with smaller values of ε and ε_h. Convergence was successful only for $\varepsilon_h = 10^{-6}$ for $u(t) \leq 5$ and for $\varepsilon_h = 10^{-5}$ for $u(t) \leq 3.5$. Convergence was unsuccessful for $u(t) \leq 5$ with $\varepsilon_h = 10^{-5}$ and for $u(t) \leq 3.5$ with $\varepsilon_h = 10^{-6}$.
- Due to the steep control profile, the Williams–Otto problem with $u(t) \leq 100$ could not be solved with smaller values of ε and ε_h. On the other hand, for $u(t) \leq 0.6$ these cases solve without difficulty with solution profiles essentially the same as those presented above and very slight improvements in the objective function.

Optimal Control of Modified Singular Problems

We now consider six test examples modified from singular control problems; similar examples are also considered in Ref. 16. Convergence properties for mesh refinement with the collocation approach require coercivity conditions on the control problem,[21,29] analogous to SSOSC conditions for the discretized Problem (16). As singular control problems violate coercivity conditions, convergence of our bilevel approach cannot be guaranteed for these problems. Instead, we consider singular examples with an additional regularization term, $\hat{\varepsilon} \int_0^{t_f} u^2 dt$ in the objective, with $\hat{\varepsilon} > 0$, in order to make these examples coercive. The first four problems stem from small, singular problems modified from the literature, whereas the last two are larger problems related to dynamic process optimization.

Catalyst mixing problem

In this problem, the reactions $A \Longleftrightarrow B \rightarrow C$ take place in a tubular reactor at constant temperature. The first reaction is reversible and is catalyzed by Catalyst I, whereas the second irreversible reaction is catalyzed by Catalyst II. The goal of this problem is to determine the optimal mixture of catalysts along the length t of the reactor in order to maximize the amount of product C. The modified optimal catalyst mixing problem, can be stated as

$$\min \quad a(t_f) + b(t_f) - a_0 + \hat{\varepsilon} \int_0^{t_f} u^2 dt \tag{35a}$$

$$\text{s.t.} \quad \frac{da(t)}{dt} = -u(k_1 a(t) - k_2 b(t)) \tag{35b}$$

$$\frac{db(t)}{dt} = u(k_1 a(t) - k_2 b(t)) - (1-u) k_3 b(t) \tag{35c}$$

$$a(0) = 1, b(0) = 0, u(t) \in [0, 1] \tag{35d}$$

with $t_f = 4, k_1 = k_3 = 1$, and $k_2 = 10$.

For the singular problem with $\hat{\varepsilon} = 0$, the initial number of finite elements for Step 1 is set to 20 and the final number of finite elements remains at 20 with the bilevel approach. This requires 10 solutions of the outer problem and a total of 1734 IPOPT iterations for (16). The total CPU time for solving the outer problems is 638.9 CPU seconds. The value of $\bar{H}$ from (26) is 0.04410395200951 and the optimal objective has a value of -0.19181435553249. The final element distribution along with the state, control, and Hamiltonian profiles are shown in Figure 7. As with the analytical solution, the control has two breakpoints and a bang-singular-bang structure. The breakpoints determined by the two-stage algorithm are 0.135021 and 3.72526. Compared with the analytical breakpoints (see Ref. 30) these values differ by 0.00128 and 2.99×10^{-5}, respectively.

For $\hat{\varepsilon} = 0.1$, the initial number of finite elements for Step 1 is also 20 and the bilevel approach increases this number to 21. This requires only one solution of the outer problem and a total of 26 IPOPT iterations for (16). The total CPU time for

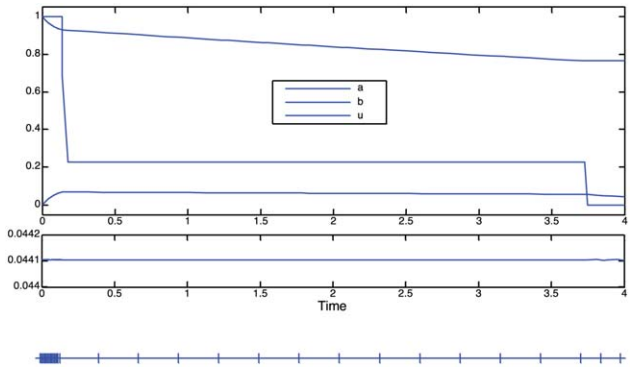

Figure 7. State, control, and Hamiltonian profiles for singular catalyst mixing problem.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

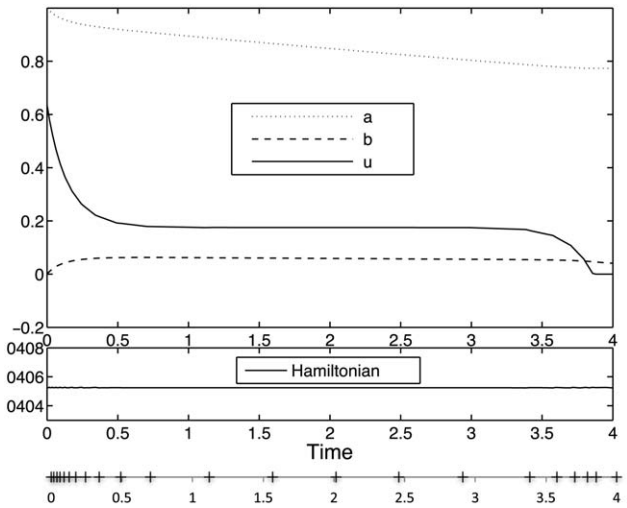

Figure 8. State, control, and Hamiltonian profiles for modified catalyst mixing problem.

solving the outer problems is 8.3 CPU seconds. The value of $\bar{H}$ from (26) is 0.04052572 and the optimal objective has a value of −0.170906354727335. The final element distribution along with the state, control, and Hamiltonian profiles are shown in Figure 8.

Aly–Chan singular problem

The modified singular optimal control problem[31] is given by

$$\min_u z_3\left(\frac{\pi}{2}\right) \tag{36a}$$

$$\text{s.t.}\ \frac{dz_1}{dt} = z_2;\ z_1(0) = 0 \tag{36b}$$

$$\frac{dz_2}{dt} = u;\ z_2(0) = 1 \tag{36c}$$

$$\frac{dz_3}{dt} = \frac{1}{2}z_2^2 - \frac{1}{2}z_1^2 + \hat{\varepsilon}u^2;\ z_3(0) = 0 \tag{36d}$$

$$-1 \le u \le 1 \tag{36e}$$

with $\hat{\varepsilon} = 0.1$. The initial number of finite elements for Step 1 is 20 and the final number remains at 20. The bilevel approach requires only one solution of the outer problem and a total of five IPOPT iterations for (16). The total CPU time for solving the outer problems is 1.6 CPU seconds. The value of $\bar{H}$ from (26) is 0.4802907 and the optimal objective has a value of 0.03836184. The final element distribution along with the state, control, and Hamiltonian profiles are shown in Figure 9.

Rayleigh problem

In this example, we find $u(t)$ that solves

$$\min \int_0^{t_f=2.5} x_1^2 + \hat{\varepsilon}u^2 dt \tag{37a}$$

$$\text{s.t.}\ \frac{dx_1}{dt} = x_2 \tag{37b}$$

$$\frac{dx_2}{dt} = -x_1 + (1.4 - 0.14x_2^2)x_2 + 4u \tag{37c}$$

$$x_1(0) = -5, x_2(0) = -5 \tag{37d}$$

with $\hat{\varepsilon} = 1$. The initial number of finite elements for Step 1 is 20 and the bilevel approach increases this number to 208. This requires four solutions of the outer problem and a total of 428 IPOPT iterations for (16). The total CPU time for solving the outer problems is 1788.4 CPU seconds. The value of $\bar{H}$ from (26) is −2.019174 and the optimal objective has a value of 29.37608. The final element distribution along with the state, control, and Hamiltonian profiles are shown in Figure 10.

Modified stirred tank reactor

The modified stirred tank reactor optimal control problem[32] is given by

$$\min_u \int_0^{t_f} (x_1^2 + x_2^2 + \hat{\varepsilon}u^2) dt \tag{38a}$$

$$\text{s.t.}\ \frac{dx_1}{dt} = 2a_1 + a_4 - a_1 u \quad x_1(0) = 0.05 \tag{38b}$$

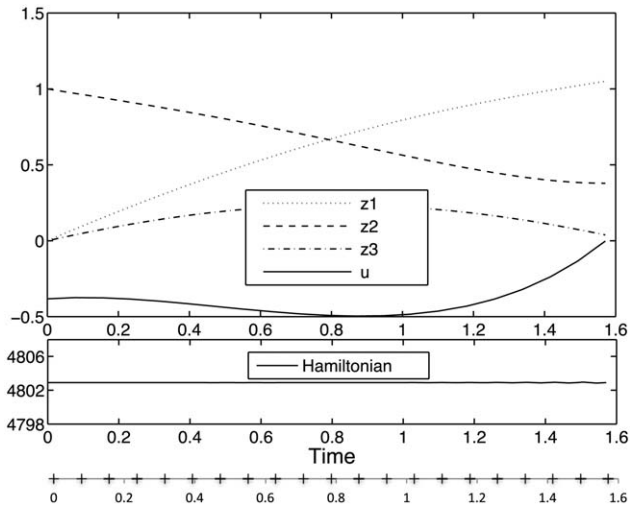

Figure 9. State, control, and Hamiltonian profiles for modified Aly–Chan problem.

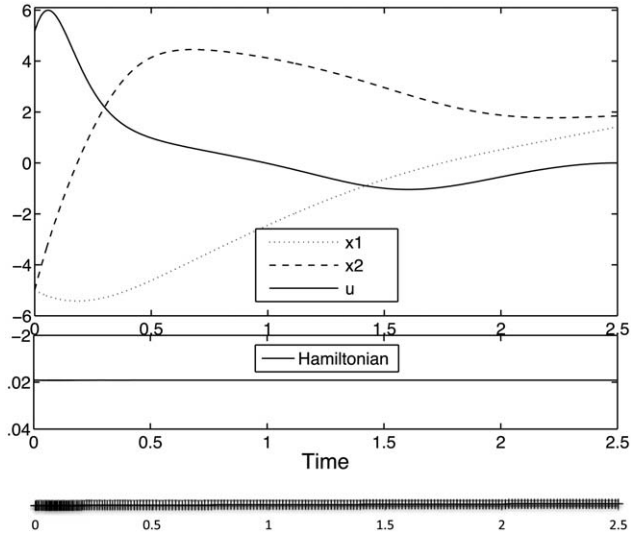

Figure 10. State, control, and Hamiltonian profiles for Rayleigh problem.

$$\frac{dx_2}{dt} = 0.5 - x_2 + a_4; x_2(0) = 0 \tag{38c}$$

$$x_1(t_f) = x_2(t_f) = 0 \tag{38d}$$

with $\hat{\varepsilon} = 1$ and $t_f = 0.78$. The initial number of finite elements from Step 1 was set to 20. The two-step algorithm is executed and terminates with the ℓ_1 penalty term at zero. The number of finite elements increases to 40. This requires two solutions of the outer problem and a total of 39 IPOPT iterations for (16). The total time cost for solving the outer problems is 16.1 CPU seconds. The value of $\bar{H}$ from (26) is 0.009716093 and the optimal objective has a value of 0.0167028. The final element distribution along with the state, control, and Hamiltonian profiles are shown in Figure 11.

Modified Lee–Ramirez problem

We now consider the modified bioreactor control problem by Lee and Ramirez.[10] Here, two control profiles (glucose and inducer feeding rate) are determined to maximize the volume of product at $t_f = 10$ and the reactor model is an index-1 DAE. This problem can be stated as

$$\min \quad -x_4(t_f)x_1(t_f) + \bar{\varepsilon}\int_0^{t_f}(u_1^2 + u_2^2)dt$$

$$\text{s.t.} \quad \frac{dx_1}{dt} = u_1 + u_2,\ x_1(0) = 1$$

$$\frac{dx_2}{dt} = g_1 x_1 x_2 - (u_1 + u_2)x_2,\ x_2(0) = 0.1$$

$$\frac{dx_3}{dt} = u_1 c_1 - (u_1 + u_2)x_3 - g_1 x_1 x_2 / c_2,\ x_3(0) = 40$$

$$\frac{dx_4}{dt} = g_2 x_2 - (u_1 + u_2)x_4,\ x_4(0) = 0$$

$$\frac{dx_5}{dt} = u_2 c_3 - (u_1 + u_2)x_5,\ x_5(0) = 0$$

$$\frac{dx_6}{dt} = -g_3 x_6,\ x_6(0) = 1$$

$$\frac{dx_7}{dt} = g_3(1 - x_7),\ x_7(0) = 0$$

$$t_1 = 14.35 + x_3 + (x_3^2/111.5),\ t_2 = 0.22 + x_5,\ t_3 = 0.22 x_7 x_1 / t_2$$

$$g_1 = x_3(x_6 + 0.22 x_7/t_2)/t_1$$

$$g_2 = 0.233 x_3((0.0005 + x_5)/(0.022 + x_5))/t_1$$

$$g_3 = 0.09 x_5/(0.034 + x_5)$$

$$0 \le u_1(t), u_2(t) \le 1$$

where $\bar{\varepsilon} = 0.1, c_1 = 100, c_2 = 0.51$, and $c_3 = 4$. The initial number of finite elements for Step 1 is 20 and the bilevel approach increases this number to 44. This requires two solutions of the outer problem and a total of 39 IPOPT iterations for (16). The total time cost for solving the outer problems is 1241.38 CPU seconds. The value of $\bar{H}$ from (26) is 3.54156591 and the optimal objective has a value of -6.0995216123. The final element distribution along with the state, control, and Hamiltonian profiles are shown in Figure 12.

Distillation optimal control column

Finally, we determine an optimal product profile in order to achieve desired product purity from a binary distillation column, which separates components A and B. The dynamic model is based on an equimolar overflow assumption (i.e., vapor V and liquid L flow rates do not vary with tray index j), with vapor–liquid equilibrium described by the Wilson equation. The index-1 DAE model has 32 differential state variables (due to 30 trays, a reboiler and condenser), 32 algebraic variables and the distillate flow as the control variable. This problem is originally described in Ref. 33 and is paraphrased below

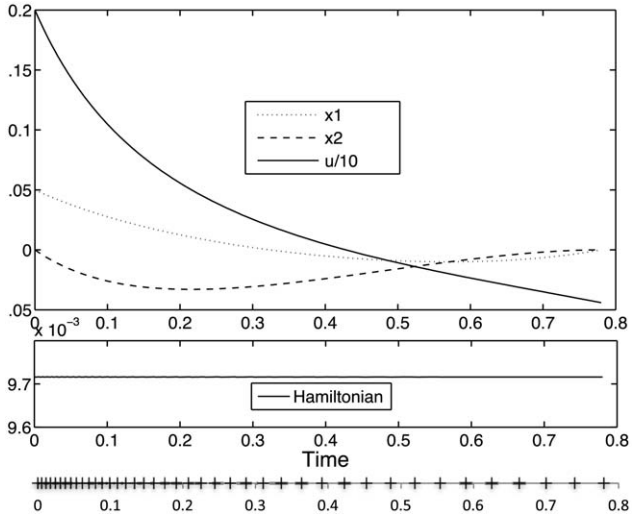

Figure 11. State, control, and Hamiltonian profiles for stirred tank reactor problem.

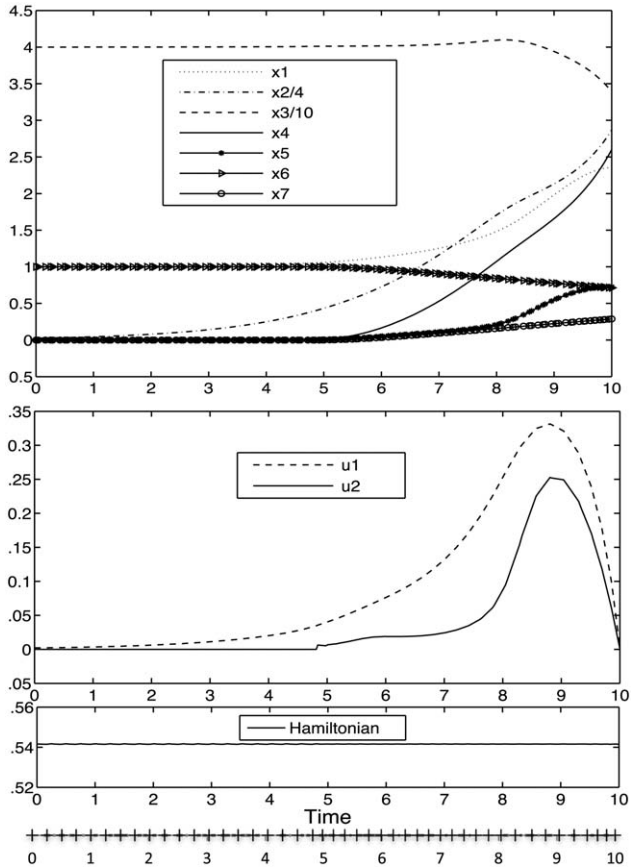

Figure 12. State, control, and Hamiltonian profiles for modified Lee–Ramirez problem.

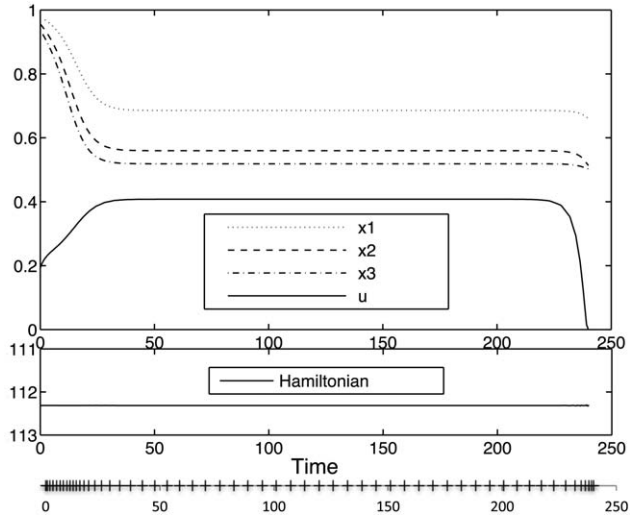

Figure 13. State, control, and Hamiltonian profiles for modified distillation problem.

$$\min \int_0^{t_f=240} (x_1-0.995)^2+0.1u^2 dt$$

$$\text{s.t.} \quad \frac{dx_1}{dt}=\frac{1}{M_{\text{cond}}}V(y_2-x_1),\ x_1(0)=x_{1,0}$$

$$\frac{dx_j}{dt}=\frac{1}{M_{\text{tray}}}L(x_{j-1}-x_j)-V(y_j-y_{j+1}),\ x_j(0)=x_{j,0}\, j=2,\ldots 16$$

$$\frac{dx_{17}}{dt}=\frac{1}{M_{\text{tray}}}(Fx_f+Lx_{16}-(F+L)x_{17}-V(y_{17}-y_{18}),\ x_{17}(0)=x_{17,0}$$

$$\frac{dx_j}{dt}=\frac{1}{M_{\text{tray}}}(F+L)(x_{j-1}-x_j)-V(y_j-y_{j+1}),\ x_j(0)=x_{j,0}\, j=18,\ldots,31$$

$$\frac{dx_{32}}{dt}=\frac{1}{M_{\text{reb}}}(F+L)x_{31}-(F-u_1(t))x_{32}-Vy_{32},\ x_{32}(0)=x_{32,0}$$

$$P=x_j\gamma_{A,j}P^{\text{sat}}_{A,j}+(1-x_j)\gamma_{B,j}P^{\text{sat}}_{B,j}$$

$$y_jP=x_j\gamma_{A,j}P^{\text{sat}}_{A,j}$$

$$P^{\text{sat}}_{A,j}=P^{\text{sat}}_A(T_j),\quad P^{\text{sat}}_{B,i}=P^{\text{sat}}_B(T_j)$$

$$\gamma_{A,j}=\gamma_A(x_j,T_j),\quad \gamma_{B,j}=\gamma_B(x_j,T_j)$$

where (for tray j) M is molar holdup, V and L are vapor and liquid flow rates, respectively, y and x are vapor and liquid mole fractions, respectively, F is the feed flow rate, γ refers to activity coefficients defined by the Wilson equation, P^{sat} is the vapor pressure, and P and T are pressure and temperature, respectively. Additional problem data and expressions for γ_A, γ_B, P^{sat}_A, and P^{sat}_B can be found in http://www.hedengren.net/research/models.htm.

The initial number of finite elements for Step 1 is set to 20 and the bilevel approach increases this number to 56. This requires two solutions of the outer problem and a total of 64 IPOPT iterations for (16). The total time for solving the outer problems is 887.5 CPU seconds. The value of $\bar{H}$ from (26) is -0.1123137 and the optimal objective has a value of 24.82098. The final element distribution along with the state, control, and Hamiltonian profiles are shown in Figure 13.

Conclusions and Future Directions

This study develops a mesh refinement strategy for a class of nonsingular optimal control problems based on simultaneous collocation. Based on moving finite elements embedded within an NLP formulation, this bilevel approach determines numerically accurate solutions for this problem class. Novel features of the algorithm include the direct location of breakpoints for control profiles, constraints on approximation error from the inner, fixed-mesh problem and a termination criterion based on a constant Hamiltonian profile, which can be embedded directly into the outer problem.

The above solution strategy has significant advantages over the full-space approach in Ref. 16 because it creates two manageable problems from a larger, ill-conditioned one. In the inner problem, a large-scale, but well-defined fixed-mesh problem is solved. In the outer problem, a much smaller problem adjusts the mesh to improve the solution. Our bilevel approach links these two problems through a large-scale NLP solver and NLP sensitivity algorithm. Applied to 11 optimal control problems ranging from 2 to 64 state equations, this simultaneous collocation approach converges to tight error tolerances and constant Hamiltonian solutions with only a few outer problem interactions. Moreover, we have demonstrated the bilevel approach is a systematic strategy for a general class of optimal control problems.

Nevertheless, a number of open issues need to be discussed for future work including the following areas.

- In the outer problem, the objective function and active error constraints are insensitive to most of the finite elements in the mesh. As a result, there may be no unique solution to the finite element mesh and the insensitive elements need to be regularized or specified in some fashion. Fortunately, this feature is confined to the outer problem, but a suitable damping/regularization strategy may need to be implemented to prevent unstable performance for the outer problem.
- Singular control problems lead to loss of sufficient second-order conditions and nonunique solutions. Although not addressed here, it may be possible to tackle them by introducing additional regularizing constraints into the outer problem.
- Although the inner problem (16) is less difficult to solve, careful attention may still be needed for initialization strategies (e.g., Steps 1 and 2 in the solution strategy) as well as constraint relaxation schemes. The latter are likely to be problem dependent.
- More general classes of optimal control problems need to be considered that include state variable constraints. Here, it may be fruitful to apply barrier functions to state constraints and assess the convergence properties of this approach.

Finally, while the mesh refinement strategy is grounded on properties of the simultaneous collocation formulation, the convergence analysis to infinite dimensional spaces is still open and needs to be addressed in future work. Moreover, a number of additional heuristics could be considered to accelerate the mesh refinement strategy within an integrated optimization platform. These additional features should lead to an even more efficient, automated strategy for dynamic process optimization.

Literature Cited

1. Pontryagin VV, Boltyanskii Y, Gamkrelidze R, Mishchenko E. The Mathematical Theory of Optimal Processes. New York, NY: Interscience Publishers Inc., 1962.
2. Li P, Arellano-Garcia H, Wozny G. Optimization of a semibatch optimization of a semibatch distillation process with model validation on the industrial site. *Ind Eng Chem Res.* 1998;37:1341–1350.

3. Logsdon JS, Biegler LT. Accurate determination of optimal reflux policies for the maximum distillate problem in batch distillation. *Ind Eng Chem Res*. 1993;32(4):692–701.
4. Oldenburg J, Marquardt W, Heinz D, Leineweber B. Mixed-logic dynamic optimization applied to batch distillation process design. *AIChE J*. 2003;49(11):2900–2917.
5. Lang Y-D, Cervantes A, Biegler LT. Dynamic optimization of a batch crystallization process. *Ind Eng Chem Res*. 1998;38(4):1469–1478.
6. Hermanto MW, Chiu M, Woo X, Braatz RD. Robust optimal control of polymorphic transformation in batch crystallization. *AIChE J*. 2007;53:2643–2650.
7. Corriou J-P, Rohani S. A new look at optimal control of a batch crystallizer. *AIChE J*. 2008;54:3188–3206.
8. Hahn DR, Fan LT, Hwang CL. Optimal startup control of a jacketed tubular reactor. *AIChE J*. 1971;17:1394–1401.
9. Jackson R. Optimal use of mixed catalysts for two successive chemical reactions. *J Optim Theory Appl*. 1968;2(1):27–39.
10. Lee J, Ramirez WF. Optimal fed-batch control of induced. Foreign protein production by recombinant bacteria. *AIChE J*. 1994;40:899–907.
11. Cuthrell J, Biegler L. Simultaneous optimization and solution methods for batch reactor control profiles. *Comput Chem Eng*. 1989;13:49–62.
12. Kiparissides C, Seferlis P, Mourikas G, Morris AJ. Online optimizing control of molecular weight properties in batch free-radical polymerization reactors. *Ind Eng Chem Res*. 2002;41:6120–6131.
13. Betts J. Practical Methods for Optimal Control and Estimation Using Nonlinear Programming. SIAM Series on Advances in Design and Control 19. SIAM, Philadelphia, PA, 2010.
14. Bartl M, Li P, Biegler LT. Improvement of state profile accuracy in nonlinear dynamic optimization with the quasi-sequential approach. *AIChE J*. 2011;57(8):2185–2197.
15. Biegler L, Cervantes A, Wächter A. Advances in simultaneous strategies for dynamic process optimization. *Chem Eng Sci*. 2002;57:575–593.
16. Chen WF, Wang KX, Shao ZJ, Biegler LT. Moving finite elements for dynamic optimization with direct transcription formulations. In: Biegler L, Mehrmann V, Campbell S, editors. Control and Optimization with Differential-Algebraic Constraints. Philadelphia, PA: SIAM, 2012:241–259.
17. Tanartkit P, Biegler LT. A nested, simultaneous approach for dynamic optimization problems. *Comput Chem Eng*. 1996;20:735–744.
18. Tanartkit P, Biegler LT. A nested simultaneous approach for dynamic optimization problems II: the outer problem. *Comput Chem Eng*. 1997;21:1365–1376.
19. Vasantharajan S, Biegler L. Simultaneous strategies for parameter optimization and accurate solution of differential-algebraic systems. *Comput Chem Eng*. 1990;14:1083–1095.
20. Stengel RF. Optimal Control and Estimation. New York: Dover Publications Inc., 1994.
21. Kameswaram S, Biegler LT. Convergence rates for direct transcription of optimal control problems using collocation at Radau points. *Comput Optim Appl*. 2008;41:81–126.
22. Ascher U, Mattheij R, Russell R. Numerical Solution of Boundary Value Problems for Ordinary Differential Equations, 2nd ed. Philadelphia, PA: SIAM Publications, 1995.
23. Russell RD, Christiansen J. Adaptive mesh selection strategies for solving boundary value problems. *SIAM J Numer Anal*. 1978;15:59–80.
24. Wächter A, Biegler L. On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. *Math Program*. 2006;106(1):25–57.
25. Pirnay H, Lopez-Negrete R, Biegler LT. Optimal sensitivity based on IPOPT. *Math Program Comput*. 2012;4:307–331.
26. Forsgren A, Gill P, Wright M. Interior methods for nonlinear optimization. *SIAM Rev*. 2002;44:525–597.
27. Fiacco AV. Introduction to Sensitivity and Stability Analysis in Nonlinear Programming. New York: Academic Press, 1983.
28. Williams T, Otto R. A generalized chemical processing. Model for the investigation of computer control. *AIEE Trans*. 1960;79:458–467.
29. Hager W. Runge-Kutta methods in optimal control and the transformed adjoint system. *Numer Math*. 2000;87:247–282.
30. Biegler LT. Nonlinear Programming: Concepts, Algorithms and Applications to Chemical Processes. Philadelphia, PA: SIAM, 2010.
31. Aly G, Chan W. Application of a modified quasilinearization technique to totally singular optimal problems. *Int J Control*. 1973;17:809–815.
32. Kirk DE. Optimal Control Theory: An Introduction. Englewood Cliffs, NJ: Prentice-Hall, 1970.
33. Hahn J, Edgar T. An improved method for nonlinear model reduction using balancing of empirical gramians. *Comput Chem Eng*. 2002;26:1379–1397.

Manuscript received Oct. 9, 2013, and revision received Dec. 14, 2013.