Deterministic Global Optimization for Dynamic Systems Using Interval Analysis

Youdong Lin and Mark A. Stadtherr
Department of Chemical and Biomolecular Engineering
University of Notre Dame, Notre Dame, IN, USA

12th GAMM–IMACS International Symposium on Scientific Computing, Computer Arithmetic and Validated Numerics
Duisburg, Germany, 26–29 September 2006

Outline

- Background
- Tools
  - Interval Analysis
  - Taylor Models
  - Validated Solutions of Parametric ODE Systems
- Algorithm Summary
- Computational Studies
- Concluding Remarks

Background

- Many practically important physical systems are modeled by ODE systems.
- Optimization problems involving these models may be stated as

$$
\begin{aligned}
\min_{\boldsymbol{\theta}, \boldsymbol{x}_\mu} \quad & \phi\left[\boldsymbol{x}_\mu(\boldsymbol{\theta}), \boldsymbol{\theta};\ \mu = 0, 1, \ldots, r\right] \\
\text{s.}t. \quad & \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}) \\
& \boldsymbol{x}_0 = \boldsymbol{x}_0(\boldsymbol{\theta}) \\
& \boldsymbol{x}_\mu(\boldsymbol{\theta}) = \boldsymbol{x}(t_\mu, \boldsymbol{\theta}) \\
& t \in [t_0, t_r] \\
& \boldsymbol{\theta} \in \boldsymbol{\Theta}
\end{aligned}
$$

- Sequential approach: Eliminate $\boldsymbol{x}_\mu$ using parametric ODE solver, obtaining an unconstrained problem in $\boldsymbol{\theta}$
- May be multiple local solutions – need for global optimization

Deterministic Global Optimization of Dynamic Systems

Much recent interest, mostly combining branch-and-bound and relaxation techniques, e.g.,

- Esposito and Floudas (2000): α-BB approach
  - Rigorous values of α not used: no theoretical guarantees
- Chachuat and Latifi (2003): Theoretical guarantee of ϵ-global optimality
- Papamichail and Adjiman (2002, 2004): Theoretical guarantee of ϵ-global optimality
- Singer and Barton (2006): Theoretical guarantee of ϵ-global optimality
  - Use convex underestimators and concave overestimators to construct two bounding IVPs, which are then solved to obtain lower and upper bounds on the state trajectories.
  - Bounding IVPs are not solved rigorously, so state bounds are not computationally guaranteed.

Deterministic Global Optimization of Dynamic Systems

Our approach: Branch-and-reduce algorithm based on interval analysis and using Taylor models

- Basic ideas
  - Use local optimizations to obtain an upper bound $\widehat{\phi}$ on the global minimum
  - Compute Taylor models of the state variables using a new validated solver for parametric ODEs (VSPODE) (Lin and Stadtherr, 2006)
  - Compute the Taylor model T_ϕ of the objective function
  - Perform constraint propagation procedure using $T_\phi \leq \widehat{\phi}$, to reduce the parameter (decision variable) space $\mathbf{\Theta}$
  - Use branch-and-bound
- Can implement to obtain either an ϵ-global minimum, or (using interval Newton approach) the exact ($\epsilon = 0$) global minimum

Interval Analysis

- A real interval $X = [a, b] = \{x \in \Re \mid a \leq x \leq b\}$ is a segment on the real number line.

- An interval vector $\boldsymbol{X} = (X_1, X_2, ..., X_n)^T$ is an n-dimensional rectangle.

- Basic interval arithmetic for $X = [a, b]$ and $Y = [c, d]$ is

$$X \text{ op } Y = \{x \text{ op } y \mid x \in X,\ y \in Y\}$$

- Interval elementary functions (e.g., $\exp(X)$, $\sin(X)$) are also available.

- The *interval extension* $F(\boldsymbol{X})$ encloses all values of $f(\boldsymbol{x})$ for $\boldsymbol{x} \in \boldsymbol{X}$; that is, $\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in \boldsymbol{X}\} \subseteq F(\boldsymbol{X})$

- Interval extensions computed using interval arithmetic may lead to overestimation of function range (the interval “dependency” problem).

Taylor Models

- Taylor Model $T_f = (p_f, R_f)$: Bounds a function $f(\boldsymbol{x})$ over $\boldsymbol{X}$ using a q-th order Taylor polynomial p_f and an interval remainder bound R_f.
- Could obtain T_f using a truncated Taylor series:

$$p_f = \sum_{i=0}^{q} \frac{1}{i!} \left[(\boldsymbol{x} - \boldsymbol{x}_0) \cdot \triangledown \right]^i f(\boldsymbol{x}_0)$$

$$R_f = \frac{1}{(q+1)!} \left[(\boldsymbol{x} - \boldsymbol{x}_0) \cdot \triangledown \right]^{q+1} F\left[\boldsymbol{x}_0 + (\boldsymbol{x} - \boldsymbol{x}_0)\zeta \right]$$

where,

$$\boldsymbol{x}_0 \in \boldsymbol{X}; \quad \zeta \in [0, 1]$$

$$[\boldsymbol{g} \cdot \triangledown]^k = \sum_{\substack{j_1 + \cdots + j_m = k \\ 0 \le j_1, \cdots, j_m \le k}} \frac{k!}{j_1! \cdots j_m!} g_1^{j_1} \cdots g_m^{j_m} \frac{\partial^k}{\partial x_1^{j_1} \cdots \partial x_m^{j_m}}$$

- Can also compute Taylor models by using Taylor model operations (Makino and Berz, 1996)

Taylor Model Operations

- Let T_f and T_g be the Taylor models of the functions $f(x)$ and $g(x)$, respectively, over the interval $x \in X$.
- Addition: $T_{f \pm g} = (p_{f \pm g}, R_{f \pm g}) = (p_f \pm p_g, R_f \pm R_g)$
- Multiplication: $T_{f \times g} = (p_{f \times g}, R_{f \times g})$ with $p_{f \times g} = p_f \times p_g - p_e$ and $R_{f \times g} = B(p_e) + B(p_f) \times R_g + B(p_g) \times R_f + R_f \times R_g$
- $B(p)$ indicates an interval bound on the function p.
- Reciprocal operation and intrinsic functions can also be defined.
- Store and operate on coefficients of p_f only. Floating point errors are accumulated in R_f.
- Beginning with Taylor models of simple functions, Taylor models of very complicated functions can be computed.
- Compared to other rigorous bounding methods (e.g., interval arithmetic), Taylor models often yield sharper bounds for modest to complicated functional dependencies (Makino and Berz, 1999).

Taylor Models – Range Bounding

- Exact range bounding of the interval polynomials – NP hard
- Direct evaluation of the interval polynomials – overestimation
- Focus on bounding the dominant part (1st and 2nd order terms)
- Exact range bounding of a general interval quadratic – also worst-case exponential complexity
- A compromise approach – Exact bounding of 1st order and diagonal 2nd order terms

$$B(p) = \sum_{i=1}^{m} \left[a_i \left(X_i - x_{i0} \right)^2 + b_i (X_i - x_{i0}) \right] + S$$

$$= \sum_{i=1}^{m} \left[a_i \left(X_i - x_{i0} + \frac{b_i}{2a_i} \right)^2 - \frac{b_i^2}{4a_i} \right] + S,$$

where S is the interval bound of other terms by direct evaluation

Taylor Models – Constraint Propagation

- Consider constraint $c(\boldsymbol{x}) \leq 0$, $\boldsymbol{x} \in \boldsymbol{X}$. Goal – Eliminate parts of $\boldsymbol{X}$ in which constraint cannot be satisfied
- For each $i = 1, 2 \cdots, m$, shrink X_i using:

$$B(T_c) = B(p_c) + R_c = a_i \left(X_i - x_{i0} + \frac{b_i}{2a_i} \right)^2 - \frac{b_i^2}{4a_i} + S_i \leq 0$$

$$\Longrightarrow \quad a_i U_i^2 \leq V_i, \quad \text{with } U_i = X_i - x_{i0} + \frac{b_i}{2a_i} \text{ and } V_i = \frac{b_i^2}{4a_i} - S_i$$

$$\Longrightarrow \quad U_i = \begin{cases} \emptyset & \text{if } a_i > 0 \text{ and } \overline{V_i} < 0 \\ \left[-\sqrt{\frac{\overline{V_i}}{a_i}}, \sqrt{\frac{\overline{V_i}}{a_i}} \right] & \text{if } a_i > 0 \text{ and } \overline{V_i} \geq 0 \\ [-\infty, \infty] & \text{if } a_i < 0 \text{ and } \overline{V_i} \geq 0 \\ \left[-\infty, -\sqrt{\frac{\overline{V_i}}{a_i}} \right] \cup \left[\sqrt{\frac{\overline{V_i}}{a_i}}, \infty \right] & \text{if } a_i < 0 \text{ and } \overline{V_i} < 0 \end{cases}$$

$$\Longrightarrow \quad X_i = X_i \cap \left(U_i + x_{i0} - \frac{b_i}{2a_i} \right)$$

Validated Solution of Parametric ODE Systems

- Consider the parametric ODE system

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta})$$

$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0 \in \boldsymbol{X}_0$$

$$\boldsymbol{\theta} \in \boldsymbol{\Theta}$$

- Validated methods:
  - Guarantee there exists a unique solution $\boldsymbol{x}(t)$ in $[t_0, t_f]$, for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\boldsymbol{x}_0 \in \boldsymbol{X}_0$
  - Compute an interval $\boldsymbol{X}_j$ that encloses all solutions of the ODE system at t_j for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\boldsymbol{x}_0 \in \boldsymbol{X}_0$
- Tools are available – VNODE, COSY VI, AWA, etc.
  - May need to treat parameters as additional state variables with zero derivative
- New tool – VSPODE (Lin and Stadtherr, 2006): Deals directly with interval-valued parameters (and also interval-valued initial states)

New Method for Parametric ODEs

- Use interval Taylor series to represent dependence on time. Use Taylor models to represent dependence on uncertain quantities (parameters and initial states).
- Assuming $\boldsymbol{X}_j$ is known, then
  - Phase 1: Same as “standard” approach (e.g., VNODE). Compute a coarse enclosure $\tilde{\boldsymbol{X}}_j$ and prove existence and uniqueness. Use fixed point iteration with Picard operator using high-order interval Taylor series.
  - Phase 2: Refine the coarse enclosure to obtain $\boldsymbol{X}_{j+1}$. Use Taylor models in terms of the uncertain parameters and initial states.
- Implemented in VSPODE (Validating Solver for Parametric ODEs) (Lin and Stadtherr, 2006).

Method for Phase 2

- Represent uncertain initial states and parameters using Taylor models $\boldsymbol{T}_{\boldsymbol{x}_0}$ and $\boldsymbol{T}_{\boldsymbol{\theta}}$, with components

$$T_{x_{i0}} = (m(X_{i0}) + (x_{i0} - m(X_{i0})), [0,0]), \quad i = 1, \cdots, m$$

$$T_{\theta_i} = (m(\Theta_i) + (\theta_i - m(\Theta_i)), [0,0]), \quad i = 1, \cdots, p.$$

- Bound the interval Taylor series coefficients $\boldsymbol{f}^{[i]}$ by computing Taylor models $\boldsymbol{T}_{\boldsymbol{f}^{[i]}}$.
  - Use mean value theorem.
  - Evaluate using Taylor model operations.
- Reduce "wrapping effect" by using a new type of Taylor model involving a parallelpiped remainder bound.
- This results in a Taylor model $\boldsymbol{T}_{\boldsymbol{x}_{j+1}}$ in terms of the initial states $\boldsymbol{x}_0$ and parameters $\boldsymbol{\theta}$.
- Compute the enclosure $\boldsymbol{X}_{j+1} = B(\boldsymbol{T}_{\boldsymbol{x}_{j+1}})$ by bounding over $\boldsymbol{X}_0$ and $\boldsymbol{\Theta}$.

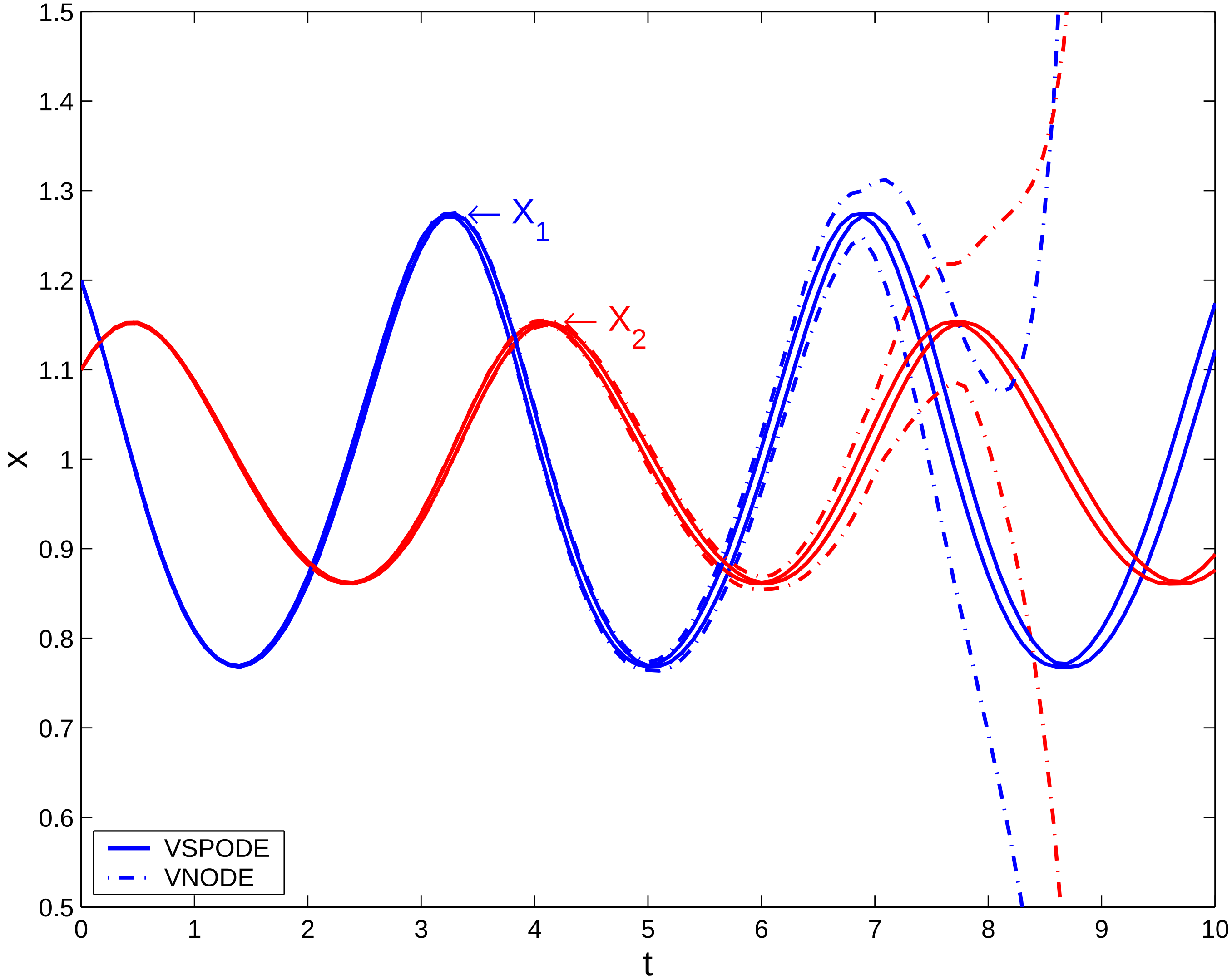
VSPODE Example

- Lotka-Volterra Problem

$$\dot{x}_1 = \theta_1 x_1 (1 - x_2)$$
$$\dot{x}_2 = \theta_2 x_2 (x_1 - 1)$$
$$\boldsymbol{x}_0 = (1.2, 1.1)^{\mathrm{T}}$$
$$\theta_1 \in [2.99, 3.01]$$
$$\theta_2 \in [0.99, 1.01]$$

- Integrate from $t_0 = 0$ to $t_N = 10$.
- VSPODE run using $q = 5$ (order of Taylor model), $k = 17$ (order of interval Taylor series) and QR for wrapping.
- For comparison, VNODE was used, with interval parameters treated as additional state variables, and run using $k = 17$ order interval Hermite-Obreschkoff and QR for wrapping.
- Constant step size of $h = 0.1$ used in both VSPODE and VNODE (step size may be reduced if necessary in Phase 1).

Lotka-Volterra Problem



(Eventual breakdown of VSPODE at $t = 31.8$)

Summary of Global Optimization Algorithm

Beginning with initial parameter interval $\mathbf{\Theta}^{(0)}$

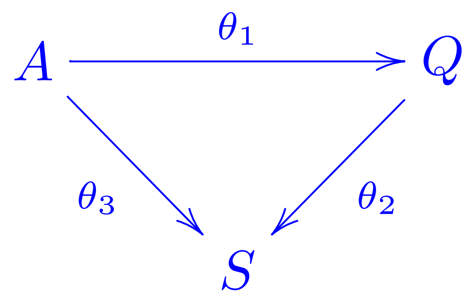
- Establish $\widehat{\phi}$, the upper bound on global minimum using p^2 local minimizations, where p is the number of parameters (decision variables)
- Iterate: for subinterval $\mathbf{\Theta}^{(k)}$
  1. Compute Taylor models of the states using VSPODE, and then obtain T_ϕ
  2. Perform constraint propagation using $T_\phi \leq \widehat{\phi}$ to reduce $\mathbf{\Theta}^{(k)}$
  3. If $\mathbf{\Theta}^{(k)} = \emptyset$, go to next subinterval
  4. If $(\widehat{\phi} - \underline{B(T_\phi)})/|\widehat{\phi}| \leq \epsilon^{\mathrm{rel}}$, or $\widehat{\phi} - \underline{B(T_\phi)} \leq \epsilon^{\mathrm{abs}}$, discard $\mathbf{\Theta}^{(k)}$ and go to next subinterval
  5. If $\overline{B(T_\phi)} < \widehat{\phi}$, update $\widehat{\phi}$ with local minimization, go to step 2
  6. If $\mathbf{\Theta}^{(k)}$ is sufficiently reduced, go to step 1
  7. Otherwise, bisect $\mathbf{\Theta}^{(k)}$ and go to next subinterval
- This finds an ϵ-global optimum with either a relative tolerance (ϵ^{rel}) or absolute tolerance (ϵ^{abs})

Global Optimization Algorithm (cont'd)

- By incorporating an interval-Newton approach, this can also be implemented as an exact algorithm ($\epsilon = 0$).
- This requires the validated integration of the first- and second-order sensitivity equations. VSPODE was used.
- Interval-Newton steps are applied only after reaching an appropriate depth in the branch-and-bound tree
- We have implemented the exact algorithm only for the case of parameter estimation problems with least squares objective
- The exact algorithm may require more or less computation time than the ϵ-global algorithm

Computational Studies – Example 1

- Parameter estimation – Catalytic cracking of gas oil (Tjoa and Biegler 1991)



- The problem is:

$$\min_{\boldsymbol{\theta}} \phi \quad = \quad \sum_{\mu=1}^{20} \sum_{i=1}^{2} \left(\widehat{x}_{\mu,i} - x_{\mu,i}\right)^2$$

$$\begin{aligned} \text{s.}t. \qquad \dot{x}_1 &= -(\theta_1 + \theta_3) x_1^2 \\ \dot{x}_2 &= \theta_1 x_1^2 - \theta_2 x_2 \\ \boldsymbol{x}_0 &= (1,0)^{\mathrm{T}}; \qquad \boldsymbol{x}_\mu = \boldsymbol{x}(t_\mu) \\ \boldsymbol{\theta} &\in [0,20] \times [0,20] \times [0,20] \end{aligned}$$

where $\widehat{\boldsymbol{x}}_\mu$ is given (experimental data).

Example 1 (Cont'd)

- Solution: $\boldsymbol{\theta}^* = (12.2139, 7.9798, 2.2217)^{\mathrm{T}}$ and $\phi^* = 2.6557 \times 10^{-3}$
- Comparisons:

Method	CPU time (s)	
	Reported	Adjusted*
This work (exact global optimum: $\epsilon = 0$) (Intel P4 3.2GHz)	11.5	11.5
This work ($\epsilon^{\mathrm{rel}} = 10^{-3}$) (Intel P4 3.2GHz)	14.3	14.3
Papamichail and Adjiman (2002) ($\epsilon^{\mathrm{rel}} = 10^{-3}$) (SUN UltraSPARC-II 360MHz)	35478	4541
Chachuat and Latifi (2003) ($\epsilon^{\mathrm{rel}} = 10^{-3}$) (Machine not reported)	10400	—
Singer and Barton (2006) ($\epsilon^{\mathrm{abs}} = 10^{-3}$) (AMD Athlon 2000XP+ 1.667GHz)	5.78	2.89

*Adjusted = Approximate CPU time after adjustment for machine used (based on SPEC benchmark)

Computational Studies – Example 2

- Singular control problem (Luus, 1990)
- The original problem is stated as:

$$\begin{aligned}
\min_{\theta(t)} \phi \quad &= \int_{t_0}^{t_f} \left[x_1^2 + x_2^2 + 0.0005(x_2 + 16t - 8 - 0.1x_3\theta^2)^2 dt\right] \\
s.t. \quad \dot{x}_1 &= x_2 \\
\dot{x}_2 &= -x_3\theta + 16t - 8 \\
\dot{x}_3 &= \theta \\
\boldsymbol{x}_0 &= (0, -1, -\sqrt{5})^{\mathrm{T}} \\
t &\in [t_0, t_f] = [0, 1] \\
\theta &\in [-4, 10]
\end{aligned}$$

- $\theta(t)$ is parameterized as a piecewise constant control profile with n pieces (equal time intervals)

Example 2 (Cont'd)

- Reformulate problem to be autonomous, and introduce quadrature variable:

$$
\begin{aligned}
\min_{\theta(t)} \phi \quad &= \quad x_5(t_f) \\
s.t. \quad \dot{x}_1 \quad &= x_2 \\
\dot{x}_2 \quad &= -x_3\theta + 16x_4 - 8 \\
\dot{x}_3 \quad &= \theta \\
\dot{x}_4 \quad &= 1 \\
\dot{x}_5 \quad &= x_1^2 + x_2^2 + 0.0005(x_2 + 16x_4 - 8 - 0.1x_3\theta^2)^2 \\
\boldsymbol{x}_0 \quad &= (0, -1, -\sqrt{5}, 0, 0)^{\mathrm{T}} \\
t \quad &\in [t_0, t_f] = [0, 1] \\
\theta \quad &\in [-4, 10]
\end{aligned}
$$

- Do ϵ-global optimization with $\epsilon^{\mathrm{abs}} = 10^{-3}$

Example 2 (Cont'd)

			CPU time (s)	
n	ϕ^*	$\boldsymbol{\theta}^*$	This work	Singer and Barton (2006)*
1	0.4965	(4.071)	0.02	0.9
2	0.2771	(5.575, -4.000)	0.32	11.3
3	0.1475	(8.001, -1.944, 6.042)	10.88	270.3
4	0.1237	(9.789, -1.200, 1.257, 6.256)	369.0	—
5	0.1236	(10.00, 1.494, -0.814, 3.354, 6.151)	8580.6	—

* Approximate CPU time after adjustment for machine used

Computational Studies – Example 3

- Oil shale pyrolysis problem (Luus, 1990)
- The original problem is stated as:

$$
\begin{aligned}
\min_{\theta(t)} \quad & \phi && = -x_2(t_f) \\
s.t. \quad & \dot{x}_1 && = -k_1 x_1 - (k_3 + k_4 + k_5) x_1 x_2 \\
& \dot{x}_2 && = k_1 x_1 - k_2 x_2 + k_3 x_1 x_2 \\
& k_i && = a_i \exp\left(\frac{-b_i/R}{\theta}\right), i = 1, \ldots, 5 \\
& \boldsymbol{x}_0 && = (1, 0)^{\mathrm{T}} \\
& t && \in [t_0, t_f] = [0, 10] \\
& \theta && \in [698.15, 748.15].
\end{aligned}
$$

- $\theta(t)$ is paramaterized as a piecewise constant control profile with n pieces (equal time intervals)

Example 3 (Cont'd)

- Transformation of control variable

$$\bar{\theta} = \frac{698.15}{\theta}$$

- The reformulated problem is stated as:

$$\begin{aligned}
\min_{\bar{\theta}(t)} \quad & \phi && = -x_2(t_f) \\
s.t. \quad & \dot{x}_1 && = -k_1 x_1 - (k_3 + k_4 + k_5) x_1 x_2 \\
& \dot{x}_2 && = k_1 x_1 - k_2 x_2 + k_3 x_1 x_2 \\
& k_i && = a_i \exp\left(-\bar{\theta} b_i / R\right), i = 1, \cdots, 5 \\
& \boldsymbol{x}_0 && = (1, 0)^{\mathrm{T}} \\
& t && \in [t_0, t_f] = [0, 10] \\
& \bar{\theta} && \in [698.15/748.15,\ 1].
\end{aligned}$$

- Do ϵ-global optimization with $\epsilon^{\text{abs}} = 10^{-3}$

Example 3 (Cont'd)

			CPU time (s)	
n	ϕ^*	$\bar{\boldsymbol{\theta}}^*$	This work	Singer and Barton (2006)*
1	-0.3479	(0.984)	3.2	13.1
2	-0.3510	(0.970, 1.000)	26.8	798.7
3	-0.3517	(1.000, 0.963, 1.000)	251.6	-
4	-0.3523	(1.000, 0.955, 1.000, 1.000)	2443.5	-

* Approximate CPU time after adjustment for machine used

Concluding Remarks

- An approach has been described for deterministic global optimization of dynamic systems using interval analysis and Taylor models
  - A validated solver for parametric ODEs is used to construct bounds on the states of dynamic systems
  - An efficient constraint propagation procedure is used to reduce the incompatible parameter domain
- Can be combined with the interval-Newton method (Lin and Stadtherr, 2006)
  - True global optimum instead of ϵ-convergence
  - May or may not reduce CPU time required

Acknowledgments

- Indiana 21st Century Research & Technology Fund
- U. S. Department of Energy