Journal of Global Optimization **24:** 1–33, 2002.

A Rigorous Global Optimization Algorithm for Problems with Ordinary Differential Equations

IOANNIS PAPAMICHAIL and CLAIRE S. ADJIMAN
Centre for Process Systems Engineering, Department of Chemical Engineering and Chemical Technology, Imperial College of Science, Technology and Medicine, London SW7 2BY, UK (Corresponding author: c.adjiman@ic.ac.uk)

Abstract. The optimization of systems which are described by ordinary differential equations (ODEs) is often complicated by the presence of nonconvexities. A deterministic spatial branch and bound global optimization algorithm is presented in this paper for systems with ODEs in the constraints. Upper bounds for the global optimum are produced using the sequential approach for the solution of the dynamic optimization problem. The required convex relaxation of the algebraic functions is carried out using well-known global optimization techniques. A convex relaxation of the time dependent information is obtained using the concept of differential inequalities in order to construct bounds on the space of solutions of parameter dependent ODEs as well as on their second-order sensitivities. This information is then incorporated in the convex lower bounding NLP problem. The global optimization algorithm is illustrated by applying it to four case studies. These include parameter estimation problems and simple optimal control problems. The application of different underestimation schemes and branching strategies is discussed.

Key words: Global optimization, Ordinary differential equations, Convex underestimation, Differential inequalities

1. Introduction

Many systems can be described by ordinary differential equations (ODEs). These include, for instance, physical systems (molecular dynamics (Allen and Tildesley, 1987)), chemical systems (chemical kinetics (Smith, 1981)), economic or other dynamic processes (Banks, 1994). The application of dynamic optimization to such systems allows the determination of their optimal performance under transient conditions. In most cases, numerous local solutions exist for problems of this type. Thus, Luus and Cormack (1972) showed that this is true even for rather simple problems. For a bifunctional catalyst example Luus et al. (1992) identified 25 local optima using 100 random starting points and Esposito and Floudas (2000a) identified over 300 local optima using 1000 random starting points. Due to the presence of nonconvexities, current numerical methods may fail to identify a solution for a feasible problem. Furthermore, if they succeed in finding a solution they can only guarantee that it is a local one. There is therefore a need to develop global optimization algorithms which can address these issues and guarantee optimal performance. In this section existing approaches to dynamic optimization and recent developments in global optimization of dynamic systems are reviewed briefly.

1.1. DYNAMIC OPTIMIZATION

Several methods can be applied for the numerical solution of the dynamic optimization problem. One class of approaches uses variable discretization in order to transform the problem to a finite dimensional nonlinear programming (NLP) problem. The discretization can be applied to the controls (control parameterization) or to both the state variables and the controls (complete discretization). Due to the nonconvexity of these formulations, only local solutions can be identified by most current NLP solvers.

In complete discretization (known as the simultaneous approach) the solution is carried out in the full space of variables. Tsang et al. (1975) used collocation to discretize the system. Oh and Luus (1977) proposed the use of orthogonal collocation (Villadsen and Stewart, 1967), while Biegler (1984) applied global orthogonal collocation and Lagrange polynomials for the approximation of the continuous variables. Renfro et al. (1987) divided the time horizon into finite elements and used global spline collocation on the state variables and piecewise constant approximation for the controls. Orthogonal collocation was also used by Cuthrell and Biegler (1987), but this time on finite elements. The stability and the error properties of implicit Runge-Kutta methods for systems which are described by differential and algebraic equations (DAEs) were considered by Logsdon and Biegler (1989). They enforced appropriate error constraints in a collocation based NLP formulation. However, all these methods result in an NLP with a large number of variables and nonlinear equality constraints.

In control parameterization (known as the sequential approach) only the controls are discretized. The dynamic system is decoupled from the optimization stage and is integrated using well established techniques in order to evaluate the objective function and the constraints. Various techniques can be used to calculate the derivatives of the objective function and the constraints. In contrast to complete discretization, a good approximation of the state variables can be obtained without affecting the size of the NLP problem. A number of schemes have been proposed for control parameterization. Pollard and Sargent (1970) used piecewise constant controls and variable switching times. Sargent and Sullivan (1978) presented a general formulation for ODEs with constraints. They developed an optimal control package using piecewise constant controls, variable switching times and adjoint integration to obtain the gradients with respect to the decision variables. Goh and Teo (1988) used control parameterization for problems with general constraints obtaining the gradients from the adjoint equations. Vassiliadis et al. (1994a,b) considered a class of multistage dynamic optimization problems with general constraints. They applied control parameterization using Lagrange polynomials and variable-length intervals. Gradients were obtained by the integration of the sensitivity equations.

Finally, another approach, which uses stochastic features, is dynamic programming. Dynamic programming utilizing grids was used by Luus (1990a,b). It was also applied to problems with final state constraints using penalty functions by Luus and Rosen (1991). Luus (1993) extended the originally piecewise constant control to piecewise linear continuous control policies. Bojkov and Luus (1993) and Luus and Bojkov (1994) studied the effect of the various parameters used in iterative dynamic programming. Dadebo and McAuley (1995) used dynamic programming and absolute error penalty functions in handling constrained problems.

The reader is referred to Sargent (2000) for a description of other approaches to dynamic optimization.

1.2. GLOBAL OPTIMIZATION OF DYNAMIC PROBLEMS

The algorithms that have been proposed to date for the solution of a dynamic problem to global optimality fall into two broad categories, the stochastic and the deterministic approaches. The general principles of such approaches as applied to NLPs are reviewed in Boender and Romeijn (1995) for stochastic techniques and Horst and Tuy (1996) and Floudas (1999) for deterministic techniques.

Several researchers have used stochastic optimization to address dynamic optimization problems. Rosen and Luus (1992) applied control parameterization and used line search to determine initial points for the NLP solver. Banga and Seider (1996) applied a stochastic algorithm for the optimal control of chemical engineering systems. Control parameterization with piecewise linear functions and variable-length intervals was used. The DAE system was integrated for different values of the parameters chosen using a stochastic procedure and the value of the objective function was calculated for each feasible set of parameters until convergence was achieved. Carrasco and Banga (1997) applied this algorithm and a modified version to the dynamic optimization of chemical batch reactors.

Recently, deterministic algorithms have been considered for dynamic problems. Smith and Pantelides (1996) applied their symbolic manipulation and spatial BB algorithm to the solution of a dynamic optimization problem with complete discretization. Esposito and Floudas (2000b) used the αBB method (Maranas and Floudas, 1994; Androulakis et al., 1995; Adjiman and Floudas, 1996; Adjiman et al., 1998a; Adjiman et al., 1998b) for the solution of the NLP problem that arises from the use of the complete discretization approach for the solution of the dynamic optimization problem. They found that for systems which are nonlinear in the state variables this approach performs poorly, sometimes even failing to achieve convergence. The αBB method was also used for the solution of the NLP problem that arises from the use of the control parameterization (Esposito and Floudas, 2000a–c). A theoretical guarantee of attaining the global solution of the problem is offered as long as rigorous values for the parameters needed or rigorous bounds on these parameters are obtained. Several methods were proposed to calculate these parameters and good results were produced. However, the issue of the theoretical guarantee of global optimality remains open and will be discussed in more detail in Section 4.

1.3. OUTLINE

The approach used in the present work is based on existing deterministic global optimization BB techniques for NLP problems. A new convex underestimating procedure is developed for systems that can be described by ODEs.

Section 2 gives the mathematical statement of the dynamic optimization problem and describes the sequential approach used for its solution. Section 3 presents differential inequalities and their application to the overestimation of the solution space of parameter dependent ODEs. This is used to develop a convex relaxation strategy which is incorporated within a global optimization algorithm as described in Section 4. An implementation of the algorithm and four case studies are discussed in Section 5. The case studies include two parameter estimation problems in chemical kinetics modeling and two optimal control problems.

2. Dynamic optimization

The formulation of the dynamic optimization problem studied is given by:

$$
\begin{aligned}
&\min_p \; J(x(t_i, p), p\,;\; i = 0, 1, \ldots, NS)\\
&\text{s.t.}\\
&\quad \dot{x} = f(t, x, p) \;\; \forall t \in [t_0, t_{NS}]\\
&\quad x(t_0, p) = x_0(p)\\
&\quad g_i(x(t_i, p), p) \leqslant 0\,,\;\; i = 0, 1, \ldots, NS\\
&\quad p^L \leqslant p \leqslant p^U
\end{aligned}
\tag{1}
$$

where $t \in \Re$ is time, t_0 and t_{NS} are the initial and final time, respectively, $t_i \in [t_0, t_{NS}]$, x and $\dot{x} \in \Re^n$ are the state variables and their time derivatives, respectively, and $p \in \Re^r$ are the time-invariant parameters. The functions J, f, x_0 and g_i, $i = 0, 1, \ldots, NS$, are such that $J : \Re^{n\cdot(NS+1)} \times \Re^r \mapsto \Re$, $f : [t_0, t_{NS}] \times \Re^n \times \Re^r \mapsto \Re^n$, $x_0 : \Re^r \mapsto \Re^n$ and $g_i : \Re^n \times \Re^r \mapsto \Re^{s_i}$.

Systems with time dependent controls can be transformed to this form using control parameterization (Vassiliadis et al., 1994a).

REMARK 2.1. The following assumptions are made:

- $J(x(t_i, p), p\,;\; i = 0, 1, \ldots, NS)$ is twice continuously differentiable with respect to $x(t_i, p)$, $i = 0, 1, \ldots, NS$ and p on $\Re^{n\cdot(NS+1)} \times \Re^r$.
- Each element of $g_i(x(t_i, p), p)$ is twice continuously differentiable with respect to $x(t_i, p)$ and p, $i = 0, 1, \ldots, NS$ on $\Re^n \times \Re^r$.
- Each element of $f(t, x, p)$ is twice continuously differentiable with respect to the states x and the parameters p on $[t_0, t_{NS}] \times \Re^n \times \Re^r$.

- Each element of $x_0(p)$ is twice continuously differentiable with respect to the parameters p on $\Re^r$.

The sequential approach is used for the solution of this NLP problem and provides an upper bound for the global optimum solution. Given values for the parameters, p, the system can be integrated from t_0 to t_{NS}. After reaching t_{NS}, the objective function and the constraints can be evaluated. The evaluation of their gradients requires the solution of the sensitivity equations, which are derived by differentiating the differential equations with respect to the parameters, p:

$$\dot{x}_p(t,p) = \frac{\partial f}{\partial x} x_p(t,p) + \frac{\partial f}{\partial p} \quad \forall t \in [t_0, t_{NS}] \tag{2}$$

where:

$$x_p(t,p) = \frac{\partial x}{\partial p} \tag{3}$$

and

$$\dot{x}_p(t,p) = \frac{\partial}{\partial t}\left(\frac{\partial x}{\partial p}\right). \tag{4}$$

The initial condition for the sensitivity equations is found by differentiating the initial condition of the original system with respect to the parameters, p:

$$x_p(t_0,p) = \frac{\partial x_0}{\partial p}. \tag{5}$$

Some theorems on continuity and differentiability of the solution of ODE systems are presented in Appendix A.

In the context of global optimization, second-order information can play a significant role. The following remarks, to be used in the development of convex relaxations, show that this information exists and that it can be derived in a manner analogous to first-order information.

REMARK 2.2. Based on Remark 2.1 and Appendix A the solution $x(t_i, p)$ of the ODE system using the initial condition specified is twice continuously differentiable with respect to the parameters p on $\Re^{\mathrm{r}}$.

REMARK 2.3. If the system of first-order sensitivity equations (2) and (5) is differentiated once again with respect to the parameters, p, then the second-order sensitivity equations are produced.

3. Differential inequalities

The dependence of convex relaxations on variable bounds is a common feature of deterministic global optimization algorithms. Since the state variables appear

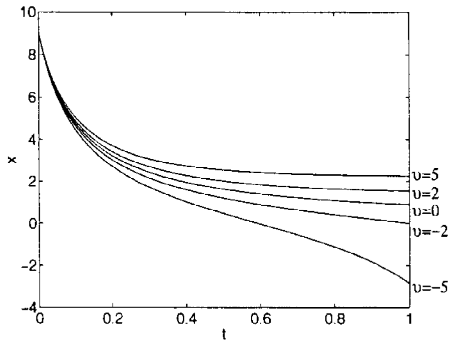


Figure 1. Trajectories of the state variable for different values of v.

in the nonconvex objective function and constraints, a method for the derivation of rigorous bounds on these variables at time $t_i, \ i = 0, 1, \ldots, NS$, is needed. This issue can be resolved by generating an overestimation of the solution space of ODEs.

The following ODE is considered as an example:

$$\dot{x}(t) = -x(t)^2 + v \ , \ \forall t \in [0, 1] \tag{6}$$

$$x(0) = 9 \tag{7}$$

If the parameter v can take values in the following interval:

$$-5 \leqslant v \leqslant 5, \tag{8}$$

then the right hand side of ODE (6)–(7) represents a set of functions rather than a single function. Solutions of this ODE for different values of v are shown in Figure 1.

It can be observed that the solutions for the upper bound and the lower bound of v give bounds on the trajectories. The aim of this section is to propose a systematic approach for the derivation of such bounds, applicable to general ODEs.

Validated (or interval) methods produce an enclosure of the solution of an ODE system. Surveys of interval methods can be found in Moore (1984), Nickel (1986) and Corliss (1995). Differential inequalities (Lakshmikantham and Leela, 1969; Walter, 1970) can be used to construct sub- and superfunctions. Nickel (1978, 1983) and Adams (1980) characterized systems of ODEs for which the theory of differential inequalities can be used to compute inclusions which are as tight as possible. Finite difference approximations combined with interval calculations can also be applied to produce interval solutions. Moore, (1966, 1979) applied them so as to overestimate the remainder term of a Taylor series expansions. Differentiation arithmetic was used to generate all the series terms. Nedialkov et al. (1999), provided a common framework for the Taylor series methods. They compared different schemes (Moore, 1966, 1978; Krückeberg, 1969; Lohner, 1987; Rihm, 1994) and identified their difficulties.

Neumaier (1994) used logarithmic norms and differential inequalities for the enclosure of solutions to initial value problems for ODEs. Recently, Berz et al. (2001) used high-order Taylor polynomials with remainder bound for the verified integration of ODEs.

The next subsection presents the basic definitions and theorems on how to identify bounding trajectories for the solutions of ODEs, using differential inequalities. The extension of these theorems to parameter dependent ODEs is discussed in Section 3.2.

3.1. BOUNDING THE SOLUTIONS OF ODES

The following ODE is studied:

$$\dot{x}(t) = f(t, x(t)), \ \forall t \in (0, T] \tag{9}$$

$$x(0) = x_0 \tag{10}$$

where $x(t)$ and $\dot{x}(t) \in \Re^n$ and $f : (0, T] \times \Re^n \mapsto \Re^n$. The next theorems deal with lower bounds (subfunctions) $\underline{x}(t)$ and upper bounds (superfunctions) $\overline{x}(t)$ for the solution of this ODE, $x(t)$.

DEFINITION 3.1. Let $0 < T \in \Re$, $\mathcal{I} = [0, T]$, $\mathcal{I}_0 = (0, T]$ and the class $\mathcal{X}$ be defined as the set of all functions $x : \mathcal{I} \mapsto \Re^n$, continuous on $\mathcal{I}$ and differentiable on $\mathcal{I}_0$. Let also $x = (x_1, x_2, \dots, x_n)^T$ and $x_{k^-} = (x_1, x_2, \dots, x_{k-1}, x_{k+1}, \dots, x_n)^T$. The notation $f(t, x) = f(t, x_k, x_{k^-})$ is used.

THEOREM 3.1 (12.IVa Walter (1970)). *Let f be continuous and satisfy a uniqueness condition on $\mathcal{I}_0 \times \Re^n$. If $\underline{x}(t), \overline{x}(t) \in \mathcal{X}$ satisfy the following inequalities:*

$$\begin{aligned}
&\underline{x}(0) \leqslant x_0 \leqslant \overline{x}(0)\\
&\dot{\underline{x}}_k(t) \leqslant f_k(t, \underline{x}_k(t), [\underline{x}_{k^-}(t), \overline{x}_{k^-}(t)])\\
&\dot{\overline{x}}_k(t) \geqslant f_k(t, \overline{x}_k(t), [\underline{x}_{k^-}(t), \overline{x}_{k^-}(t)])\\
&\quad \forall t \in \mathcal{I}_0 \text{ and } k = 1, 2, \dots, n
\end{aligned}$$

then $\underline{x}(t)$ is a subfunction and $\overline{x}(t)$ is a superfunction for the solution of ODE (9∼10) $x(t)$, i.e.,

$$\underline{x}(t) \leqslant x(t) \leqslant \overline{x}(t)\ , \ \forall t \in \mathcal{I},$$

where the inequalities are understood component-wise.

DEFINITION 3.2 Let $g(x)$ be a mapping $g : \mathcal{D} \mapsto \Re$ with $\mathcal{D} \subseteq \Re^n$. Again the notation $g(x) = g(x_k, x_{k^-})$ is used. The function g is called unconditionally partially isotone (antitone) on $\mathcal{D}$ with respect to the variable x_k if

$$g(x_k, x_{k^-}) \leqslant g(\tilde{x}_k, x_{k^-}) \text{ for } x_k \leqslant \tilde{x}_k \ \ (x_k \geqslant \tilde{x}_k)$$

and for all $(x_k, x_{k^-}),\ (\tilde{x}_k, x_{k^-}) \in \mathcal{D}$.

DEFINITION 3.3 Let $f(t, x) = (f_1(t, x), \dots, f_2(t, x))^T$ and each element $f_k(t, x_k, x_{k^-})$ be unconditionally partially isotone on $\mathcal{I}_0 \times \Re \times \Re^{n-1}$ with respect to any component of x_{k^-}, but not necessarily with respect to x_k. Then f is quasi-monotone increasing on $\mathcal{I}_0 \times \Re^n$ with respect to x. (This name is due to Walter (1970).)

It is interesting to notice that if $n = 1$, then any function $f(t, x)$ is quasi-monotone increasing.

THEOREM 3.2 (12.VIa Walter (1970)). *Let f be continuous, satisfy a uniqueness condition on $\mathcal{I}_0 \times \Re^n$ and be quasi-monotone increasing on $\mathcal{I}_0 \times \Re^n$ with respect to x. Further suppose $\underline{x}(t)$, $\overline{x}(t) \in \mathcal{X}$. Then $\underline{x}(t)$ is a subfunction for the solution of ODE (9∼10) $x(t)$, i.e.,*

$$\underline{x}(t) \leqslant x(t) \ , \ \forall t \in \mathcal{I}$$

if

$$\underline{x}(0) \leqslant x_0$$
$$\dot{\underline{x}}_k(t) \leqslant f_k(t, \underline{x}(t)) \ , \ \forall t \in \mathcal{I}_0 \text{ and } k = 1, 2, \dots, n$$

and $\overline{x}(t)$ is a superfunction, i.e.,

$$\overline{x}(t) \geqslant x(t) \ , \ \forall t \in \mathcal{I}$$

if

$$\overline{x}(0) \geqslant x_0$$
$$\dot{\overline{x}}_k(t) \geqslant f_k(t, \overline{x}(t)) \ , \ \forall t \in \mathcal{I}_0 \text{ and } k = 1, 2, \dots, n$$

3.2. BOUNDING THE SOLUTIONS OF PARAMETER DEPENDENT ODES

In order to address problems of type (1), ODE (9∼10) must be replaced by:

$$\dot{x}(t) = f(t, x(t), p), \ \forall t \in (0, T] \tag{11}$$
$$x(0) = x_0(p) \tag{12}$$

where f is a function of parameters, $p \in [p^L, p^U] \subset \Re^r$, and can be considered as a set of functions $\{f(t, x(t), p)\}$. The same is true for the initial value x_0 which is usually a function of p and is considered as a set $\{x_0(p)\}$. Let $\{x\}$ be the set of solutions of (11)–(12). Again lower and upper bounds must be determined such that $\underline{x}(t) \leqslant x(t, p) \leqslant \overline{x}(t) \ , \ \forall p \in [p^L, p^U] \ , \ \forall t \in \mathcal{I}$. This can be written as $\{x\} \subseteq [\underline{x}, \overline{x}]$.

DEFINITION 3.4. If $[\underline{x}, \overline{x}]$ is the interval hull of $\{x\}$, which means that $\underline{x} = \inf\{x\}$ and $\overline{x} = \sup\{x\}$, then the bounds are called optimal. If $\underline{x}, \overline{x} \in \{x\} \subseteq [\underline{x}, \overline{x}]$ then this implies that $[\underline{x}, \overline{x}]$ is the interval hull of $\{x\}$ and so again the bounds are optimal. (Nickel, 1983)

THEOREM 3.3. *Let f be continuous and satisfy a uniqueness condition on $\mathcal{I}_0 \times \Re^n \times [p^L, p^U]$. If $\underline{x}(t), \overline{x}(t) \in \mathcal{X}$ satisfy the following inequalities:*

$$
\begin{aligned}
&\underline{x}(0) \leqslant x_0([p^L, p^U]) \leqslant \overline{x}(0) \\
&\dot{\underline{x}}_k(t) \leqslant f_k(t, \underline{x}_k(t), [\underline{x}_{k^-}(t), \overline{x}_{k^-}(t)], [p^L, p^U]) \\
&\dot{\overline{x}}_k(t) \geqslant f_k(t, \overline{x}_k(t), [\underline{x}_{k^-}(t), \overline{x}_{k^-}(t)], [p^L, p^U]) \\
&\quad \forall t \in \mathcal{I}_0 \ \textit{and} \ k = 1, 2, \ldots, n
\end{aligned}
$$

then $\underline{x}(t)$ is a subfunction and $\overline{x}(t)$ is a superfunction for the set $\{x\}$ of solutions of ODE (11∼12), i.e.,

$$\underline{x}(t) \leqslant x(t, p) \leqslant \overline{x}(t)\ , \ \forall p \in [p^L, p^U]\ , \ \forall t \in \mathcal{I}.$$

Proof. The proof is presented in Appendix B.

REMARK 3.1. If f is continuous and satisfies a uniqueness condition on $\mathcal{I}_0 \times \Re^n \times [p^L, p^U]$ then the solution of the following ODE system satisfies Theorem 3.3:

$$
\begin{aligned}
&\dot{\underline{x}}_k(t) = \inf f_k(t, \underline{x}_k(t), [\underline{x}_{k^-}(t), \overline{x}_{k^-}(t)], [p^L, p^U]) \\
&\dot{\overline{x}}_k(t) = \sup f_k(t, \overline{x}_k(t), [\underline{x}_{k^-}(t), \overline{x}_{k^-}(t)], [p^L, p^U]) \\
&\quad \forall t \in \mathcal{I}_0 \ \text{and} \ k = 1, 2, \ldots, n
\end{aligned} \tag{13}
$$

$$
\begin{aligned}
&\underline{x}(0) = \inf x_0([p^L, p^U]) \\
&\overline{x}(0) = \sup x_0([p^L, p^U])
\end{aligned} \tag{14}
$$

System (13)–(14) provides a practical procedure to construct bounding trajectories for any ODE system which satisfies the appropriate continuity and uniqueness conditions. Natural interval extensions are used as inclusion functions due to their ease of implementation and directed outward rounding is applied to the calculations in the system (13)–(14) .

REMARK 3.2. For the case of functions $f(t, x, p)$ that are quasi-monotone increasing on $\mathcal{I}_0 \times \Re^n \times [p^L, p^U]$ with respect to x the ODE system (13)–(14) is decoupled. If that is not the case, then the so-called *wrapping effect* (the phenomenon that appears when a set, which is not representable exactly by an interval vector, has to be enclosed in an interval vector) produces poor enclosures. Moore (1966) noticed that effect and proposed a coordinate transformation. Stewart (1971) applied a heuristic to reduce the tendency of the widths of the inclusion to grow more rapidly than widths of the optimal inclusion. Jackson (1975) gave a definition of wrapping and identified classes of problems on which Moore's and Krückeberg's algorithms compute good bounds. Nickel (1985) examined the wrapping effect and identified systems where it can be eliminated. Parallelotope enclosures (Lohner, 1987), ellipsoids (Neumaier, 1993) and high order zonotope enclosures (Kühn, 1998) have been used for the reduction of the wrapping effect.

REMARK 3.3. If the functions on the right-hand side of system (13) belong to the space of functions represented by system (11) and the initial conditions (14) belong to the space of initial conditions represented by (12), then the solutions of system (13)–(14), $\underline{x}$ and $\overline{x}$, are also solutions of system (11)–(12). This means that if the following are true:

$$\inf f_k(t, \underline{x}_k(t), [\underline{x}_{k^-}(t), \overline{x}_{k^-}(t)], [p^L, p^U]) \in \{f_k(t, x(t), p)\}$$
$$\sup f_k(t, \overline{x}_k(t), [\underline{x}_{k^-}(t), \overline{x}_{k^-}(t)], [p^L, p^U]) \in \{f_k(t, x(t), p)\}$$
$$\inf x_0([p^L, p^U]) \in \{x_0(p)\}$$
$$\sup x_0([p^L, p^U]) \in \{x_0(p)\}$$

then $\underline{x}, \overline{x} \in \{x\} \subseteq [\underline{x}, \overline{x}]$ and so the inclusion is optimal.

3.3. EXAMPLE

The example discussed at the beginning of this section is reconsidered here. The function, $f(t, x(t), v) = -x(t)^2 + v$, on the right-hand side of ODE (6)–(7) is quasi-monotone increasing on $[0, 1] \times \Re \times [-5, 5]$ with respect to x and, based on Remark 3.2, the bounding system is decoupled. The subfunction is given by:

$$\dot{\underline{x}}(t) = -\underline{x}(t)^2 - 5 \,, \ \forall t \in [0, 1] \tag{15}$$
$$\underline{x}(0) = 9 \tag{16}$$

and the superfunction is given by:

$$\dot{\overline{x}}(t) = -\overline{x}(t)^2 + 5 \,, \ \forall t \in [0, 1] \tag{17}$$
$$\overline{x}(0) = 9 \tag{18}$$

The functions on the right-hand side of these ODEs belong to the space of functions of the original ODE and, based on Remark 3.3, the interval bounds are optimal. The solutions of these bounding ODEs are shown in Figure 2. They enclose all the solutions of the original parameter dependent ODE.

4. Global optimization algorithm

A deterministic spatial BB global optimization algorithm for problems with ODEs in the constraints is presented. The convex relaxed problem is first formulated

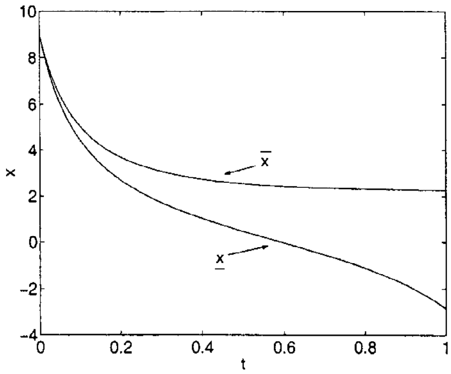


Figure 2. Bounds on the state variable as a function of time.

based on well-known techniques and the ideas introduced in Section 3. Each step of the algorithm is then presented.

4.1. FORMULATION OF THE CONVEX RELAXATION

The dynamic optimization problem has been formulated as an NLP in Section 2. Its solution gives an upper bound for the global optimization problem. A convex relaxation is now needed to underestimate the global optimum. A reformulation of the NLP problem (1) is given by:

$$\begin{aligned} &\min_{\hat{x},p} \; J(\hat{x}, p) \\ &\text{s.t.} \\ &\quad g_i(\hat{x}_i, p) \leqslant 0\,, \; i = 0, 1, \ldots, NS \\ &\quad \hat{x}_i = x(t_i, p)\,, \; i = 0, 1, \ldots, NS \\ &\quad p \in [p^L, p^U] \end{aligned} \tag{19}$$

where the values of $x(t_i, p)$, $i = 0, 1, \ldots, NS$ are obtained by solving the ODE system:

$$\dot{x}(t, p) = f(t, x(t, p), p) \;\; \forall t \in [t_0, t_{NS}] \tag{20}$$

$$x(t_0, p) = x_0(p) \tag{21}$$

4.1.1. *Bounds on $\hat{x}_i$*

As previously noted, it is essential to have bounds on all the variables participating in a nonconvex manner. In problem (19), bounds for p are user-specified. The bounds on $\hat{x}_i$ depend on the parameters bounds and must be derived automatically.

Based on Remark 3.1 bounds can be constructed for the solutions of ODE system (20)–(21). These bounds are also valid for the variable vectors $\hat{x}_i$ that have been introduced in the reformulated NLP problem:

$$\underline{x}(t_i) \leqslant \hat{x}_i \leqslant \overline{x}(t_i)\ ,\ i = 0, 1, \ldots, NS. \tag{22}$$

4.1.2. *Convex relaxation of J and g_i*

It is assumed that the functions J and g_{ij}, $i = 0, 1, \ldots, NS$, $j = 1, 2, \ldots, s_i$ can be decomposed into a sum of terms, where each term may be classified as convex, bilinear, univariate concave or general nonconvex twice continuously differentiable. Convex terms do not require any transformation. Bilinear terms can be underestimated using the convex envelope proposed by McCormick (1976). Each bilinear term is replaced by a new variable and linear constraints are inserted for this variable. Univariate concave terms can be underestimated using the secant underestimator. For a general nonconvex twice continuously differentiable term $f_{NT}(z)$ the α-based underestimator (Maranas and Floudas, 1994; Androulakis et al., 1995) can be used over the domain $[z^L, z^U] \subset \Re^m$. The addition of a negative separable convex quadratic term overpowers the nonconvexity of the original function:

$$f_{NT}(z) + \sum_{i=1}^{m} \alpha_i (z_i^L - z_i)(z_i^U - z_i) \tag{23}$$

where the values for the non-negative α_i parameters are calculated using the scaled Gerschgorin method proposed by Adjiman et al. (1998b). This method requires the use of a symmetric interval matrix $[H_{f_{NT}}] = ([\underline{h}_{ij}, \overline{h}_{ij}])$ such that $[H_{f_{NT}}] \ni H_{f_{NT}}(z) = \nabla^2 f_{NT}(z)\ ,\ \forall z \in [z^L, z^U]$. α_i can be calculated by the following formula:

$$\alpha_i = \max\left\{0, -\frac{1}{2}\left(\underline{h}_{ii} - \sum_{j \neq i} |h|_{ij}\right)\right\} \tag{24}$$

where $|h|_{ij} = \max\{|\underline{h}_{ij}|, |\overline{h}_{ij}|\}$. These values for the α_i parameters guarantee the convexity of the underestimator. The interval matrix $[H_{f_{NT}}]$ is calculated by applying natural interval extensions to the analytical expression for each second-order derivative of f_{NT} and is given by $[H_{f_{NT}}] = H_{f_{NT}}([z^L, z^U])$.

An overall convex underestimator is given by the summation of the convex underestimators for each term in the function and the introduction of additional constraints required for the bilinear terms.

4.1.3. *Convex relaxation of the set of equality constraints*

The set of equalities can be written as two sets of inequalities:

$$\hat{x}_i - x(t_i, p) \leqslant 0\ ,\ i = 0, 1, \ldots, NS \tag{25}$$

$$x(t_i, p) - \hat{x}_i \leqslant 0\ ,\ i = 0, 1, \ldots, NS \tag{26}$$

Their relaxation is given by:

$$\hat{x}_i + \breve{x}^-(t_i, p) \leqslant 0\ ,\ \ i = 0, 1, \dots, NS \tag{27}$$

$$\breve{x}(t_i, p) - \hat{x}_i \leqslant 0\ ,\ \ i = 0, 1, \dots, NS \tag{28}$$

where the $\breve{}$ superscript denotes the convex underestimator of the specified function and $x^-(t_i, p) = -x(t_i, p)$. The function $\breve{x}(t_i, p)$ is a convex underestimator of $x(t_i, p)$ and the function $-\breve{x}^-(t_i, p)$ is a concave overestimator of $x(t_i, p)$. Two strategies have been developed to derive these over and underestimators.

4.1.3.1. *Constant bounds* The constant bounds given by inequalities (22) are valid convex underestimators and concave overestimators for $x(t_i, p)$. This means that inequalities (27)–(28) can be replaced by inequalities (22). These bounds do not depend on the parameters p themselves, but do depend on the bounds on p.

4.1.3.2. *α-based bounds* An alternative way to generate the underestimators needed has been proposed by Esposito and Floudas (2000a–c). Based on Remark 2.2, $x(t_i, p)$ is a twice continuously differentiable function of the parameters p on $\Re^r$. This means that the α-based underestimators can be used for the convex underestimation of $x(t_i, p)$ and $x^-(t_i, p)$ over the domain $[p^L, p^U] \subset \Re^r$:

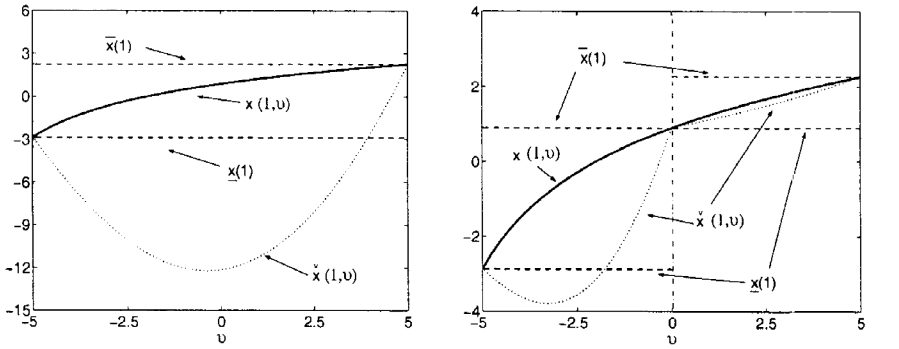
$$\breve{x}_k(t_i, p) = x_k(t_i, p) + \sum_{j=1}^{r} \alpha^+_{kij}(p^L_j - p_j)(p^U_j - p_j)$$
$$i = 0, 1, \dots, NS\ ,\ \ k = 1, 2, \dots, n \tag{29}$$

$$\breve{x}^-_k(t_i, p) = x^-_k(t_i, p) + \sum_{j=1}^{r} \alpha^-_{kij}(p^L_j - p_j)(p^U_j - p_j)$$
$$i = 0, 1, \dots, NS\ ,\ \ k = 1, 2, \dots, n \tag{30}$$

The difficulty in this approach is the calculation of the non-negative α^+_{kij} and α^-_{kij} parameters. There is no functional form available for the Hessian matrices in order to use interval calculations directly, as was done in Section 4.1.2. This hinders the calculation of the required interval matrices $[H_{x_k(t_i)}] \ni H_{x_k(t_i)}(p) = \nabla^2 x_k(t_i, p)$, $\forall p \in [p^L, p^U]$ and $[H_{x^-_k(t_i)}] = -[H_{x_k(t_i)}]$.

Esposito and Floudas (2000a–c) proposed three methods based on sampling. Using this local information, the authors show that the ability of the algorithm to identify the global solution depends on the value of the parameters, which must be large enough for the lower bounding problem to have a unique solution. As a result, they find that the number of sample points used affects the convexity of the underestimator. Thus, the method using interval calculations produces an interval matrix, $[H^*]$, that may be an underestimation of the space of the Hessian matrices. This means that there may exist $p \in [p^L, p^U] : \nabla^2 x_k(t_i, p) = H_{x_k(t_i)}(p) \notin [H^*]$.

A rigorous procedure is proposed in the present work for the calculation of the α^+_{kij} and α^-_{kij} parameters. The scaled Gerschgorin method put forward by Adjiman



a. One region: $v \in [-5, 5]$. b. Two regions: $v \in [-5, 0]$ and $v \in [0, 5]$.

Figure 3. Over and underestimators for the solution of example (6)–(7) at time $t = 1$. The solution of the ODE, $x(1, v)$, (——), which is also the α-based overestimator, the constant bounds (- - -) and the α-based underestimator ($\cdots$) are shown.

et al. (1998b) and formulas similar to Eq. (24) can be utilized again. The difficulties associated with the computation of valid Hessian matrices are resolved by constructing bounds based on Remark 3.1 for the ODE system that is generated when the first and the second-order sensitivity equations are coupled with the original ODE system (20)–(21). These bounds on the second-order derivatives can then be used to construct each element of the interval Hessian matrices needed.

4.1.3.3. *Comparison of the two strategies* The solution, $x(1, v)$, of example (6)–(7) at time $t = 1$ is a concave function of the parameter v, as shown in Figure 3a. The methods proposed in Sections 4.1.3.1 and 4.1.3.2 can be applied to construct a valid convex relaxation. The constant bounds for the whole range of parameters are given by the dashed lines. The application of the α-based underestimator gives the dotted line. An exact α-based overestimator is produced because the analysis of $[H_{x^-(1,v)}]$ reveals that the function is concave and α^- is set to zero. However, the α-based underestimator ($\alpha^+ = 0.5212$) is worse than the constant lower bound for most of the range of v. In Figure 3b the domain of v is divided into two subdomains and both strategies are again applied. For the subdomain $[0, 5]$ the α-based underestimator ($\alpha^+ = 0.0303$) is tighter than the constant bound for the whole range of the variable. This is often the case for small ranges. For the subdomain $[-5, 0]$ the value for the α^+ parameter is $\alpha^+ = 0.5212$. It is thus not possible to identify a strategy which is inherently tighter than the other.

The derivation of the α-based underestimators requires bounds on the second-order sensitivities. To obtain these, bounds on the original system (the constant bounds) and on the first-order sensitivities must also be calculated. Since the constant bounds are generated at no extra cost when α-based underestimators are used, the relaxation strategies used in practice always involve the constant bounds, with or without the α-based bounds.

4.1.4. *Convex relaxation of the NLP*

After relaxing (underestimating) the objective function and relaxing (overestimating) the feasible region, the convex relaxation of the NLP (19) is given by:

$$\begin{aligned}
\min_{\hat{x},p,w} \quad & \breve{J}(\hat{x}, p, w) \\
\text{s.t.} \quad & \\
& \breve{g}_i(\hat{x}_i, p, w) \leqslant 0\ ,\ \ i = 0, 1, \ldots, NS \\
& \underline{x}(t_i) \leqslant \hat{x}_i \leqslant \overline{x}(t_i)\ ,\ \ i = 0, 1, \ldots, NS \\
& \mathcal{C}(\hat{x}, p, w) \leqslant 0 \\
& p \in [p^L, p^U]
\end{aligned} \tag{31}$$

where the $\breve{}$ superscript denotes the convex underestimator of the specified function, $\mathcal{C}$ denotes the set of additional constraints arising from the convex relaxation of bilinear terms and w denotes the vector of new variables introduced by this relaxation. If the α-based method is also used for the convex relaxation of the set of equality constraints then the following constraints can be added to the above formulation:

$$\begin{aligned}
& \hat{x}_i + \breve{x}^-(t_i, p) \leqslant 0\ ,\ \ i = 0, 1, \ldots, NS \\
& \breve{x}(t_i, p) - \hat{x}_i \leqslant 0\ ,\ \ i = 0, 1, \ldots, NS
\end{aligned} \tag{32}$$

4.2. SPATIAL BB ALGORITHM

After constructing the convex relaxation of the original NLP problem, a spatial BB algorithm (Horst and Tuy, 1996) can be used in order to obtain the global minimum within an optimality margin, ϵ. This algorithm is described in the present subsection. Some steps are then analyzed further.

4.2.1. *Structure of BB algorithm*

Given a relative optimality margin, ϵ_r, and a maximum number of iterations, $MaxIter$:

Step 1. Initialization

Set the upper bound on the objective function: $J^u := +\infty$.
Initialize the iteration counter: $Iter := 0$.
Initialize a list of subregions $\mathcal{L}$ to an empty list: $\mathcal{L} := \emptyset$.
Initialize a region $\mathcal{R}$ to the region covering the full domain of variables p: $\mathcal{R} := [p^L, p^U]$.

Step 2. Upper bound

Solve the original NLP with bounds on p given by $\mathcal{R}$.
If a feasible solution $p_{\mathcal{R}}$ is obtained with objective function $J^u_{\mathcal{R}}$, then set the best feasible solution $p^* := p_{\mathcal{R}}$ and $J^u := J^u_{\mathcal{R}}$.

Step 3. Lower bound

Obtain bounds on the differential variables.
If the α-based relaxation is additionally used for the overestimation of the equality constraints then obtain bounds on the second-order sensitivities.
Form the relaxed problem for $\mathcal{R}$ and solve it.
If a feasible solution $p^*_{\mathcal{R}}$ is obtained for $\mathcal{R}$ with objective function $J^\ell_{\mathcal{R}}$, then add $\mathcal{R}$ to the list $\mathcal{L}$ together with $J^\ell_{\mathcal{R}}$ and $p^*_{\mathcal{R}}$.

Step 4. Subregion selection

If the list is empty, then the problem is infeasible. Terminate.
Otherwise set the region $\mathcal{R}$ to the region from the list $\mathcal{L}$ with the lowest lower bound: $\mathcal{R} := \arg\min_{\mathcal{L}_i \in \mathcal{L}} J^\ell_{\mathcal{L}_i}$.
Remove $\mathcal{R}$ from the list $\mathcal{L}$.

Step 5. Checking for convergence

If $(J^u - J^\ell_{\mathcal{R}})/|J^\ell_{\mathcal{R}}| \leqslant \epsilon_r$, then the solution is p^* with an objective function J^u. Terminate.
If $Iter = MaxIter$, then terminate and report $(J^u - J^\ell_{\mathcal{R}})/|J^\ell_{\mathcal{R}}|$.
Otherwise increase the iteration counter by one: $Iter := Iter + 1$.

Step 6. Branching within $\mathcal{R}$

Apply a branching rule on subregion $\mathcal{R}$ to choose a variable on which to branch and generate two new subregions, $\mathcal{R}_1, \mathcal{R}_2$ which are a partition of $\mathcal{R}$.

Step 7. Upper bound for each region

For $i = 1, 2$, solve the original NLP with bounds on p given by $\mathcal{R}_i$.
If a feasible solution $p_{\mathcal{R}_i}$ is obtained with objective function $J^u_{\mathcal{R}_i} < J^u$, then update the best feasible solution found so far $p^* := p_{\mathcal{R}_i}$, set $J^u := J^u_{\mathcal{R}_i}$ and remove from the list $\mathcal{L}$ all subregions $\mathcal{R}'$ such that $J^\ell_{\mathcal{R}'} > J^u$.

Step 8. Lower bound for each region

Obtain bounds on the differential variables.
If the α-based relaxation is additionally used for the overestimation of the equality constraints then obtain bounds on the second-order sensitivities.
Form the relaxed problem for each subregion $\mathcal{R}_1, \mathcal{R}_2$ and solve it.
For $i = 1, 2$, if a feasible solution $p^*_{\mathcal{R}_i}$ is obtained for $\mathcal{R}_i$ with objective function $J^\ell_{\mathcal{R}_i}$ then:

- If $J^\ell_{\mathcal{R}_i} < J^\ell_{\mathcal{R}}$, then set $J^\ell_{\mathcal{R}_i} := J^\ell_{\mathcal{R}}$.
- If $J^\ell_{\mathcal{R}_i} \leqslant J^u$, then add $\mathcal{R}_i$ to the list $\mathcal{L}$ together with $J^\ell_{\mathcal{R}_i}$ and $p^*_{\mathcal{R}_i}$.

Go to step 4.

4.2.2. *Step 6: Branching*

In this step the subregion $\mathcal{R}$ is partitioned into two new subregions, $\mathcal{R}_1, \mathcal{R}_2$. The variable on which to branch is selected via one of the following strategies:

4.2.2.1. *Strategy 1* The well-known least reduced axis rule is applied (e.g., Adjiman et al., 1998a). The ratio of the current range to the initial range is calculated for each variable. The variable with the largest ratio is selected for branching. In the case of more than one equal maxima, the variable with the smallest index is selected.

4.2.2.2. *Strategy 2* The overall influence of each variable on the tightness of the convex underestimating problem is considered. This overall contribution is made up of two components:

$$\mu_{j,\mathcal{R}} = \mu_{alg}^{j,\mathcal{R}} + \mu_{dyn}^{j,\mathcal{R}}$$

where $\mu_{alg}^{j,\mathcal{R}}$ is the contribution of the variable p_j calculated from the algebraic part of the formulation (i.e., the objective function and the inequality constraints) and $\mu_{dyn}^{j,\mathcal{R}}$ is the contribution of the variable p_j calculated from the dynamic part of the formulation (i.e., the equality constraints).

The measure for the variable p_j from the algebraic part is a summation of the contributions of all the terms in which variable p_j participates. The separation distance between each term and its underestimator is first calculated at the optimum solution of the lower bounding problem on $\mathcal{R}$ (Adjiman et al., 1998a).

The measure for the variable p_j from the dynamic part is given by (Esposito and Floudas, 2000b):

$$\begin{aligned}\mu_{dyn}^{j,\mathcal{R}} = &- \sum_{k=1}^{n}\sum_{i=0}^{NS} \alpha_{kij}^{+}(p_{j,\mathcal{R}}^{L} - p_j^*)(p_{j,\mathcal{R}}^{U} - p_j^*) \\ &- \sum_{k=1}^{n}\sum_{i=0}^{NS} \alpha_{kij}^{-}(p_{j,\mathcal{R}}^{L} - p_j^*)(p_{j,\mathcal{R}}^{U} - p_j^*)\end{aligned}$$

where the * superscript denotes the value of the variable at the solution of the lower bounding problem on $\mathcal{R}$.

The variable with the largest $\mu_{j,\mathcal{R}}$ is selected for branching. In the case of more than one equal maxima, the variable with the smallest index is selected.

4.2.3. *Step 7: Upper bound calculation*

To reduce the computational expense arising from the repeated solution of local dynamic optimization problems, the upper bound generation does not have to be applied at every iteration of the algorithm. This does not affect the ability of the algorithm to identify the global solution.

4.2.4. *Step 8: Lower bound calculation*

If the relaxed problem is feasible then it has to be as tight as the relaxation at its parent node to ensure that the bounding operation is improving (Horst and Tuy, 1996). This is enforced by applying the following test: If $J^{\ell}_{\mathcal{R}_i} < J^{\ell}_{\mathcal{R}}$, then set $J^{\ell}_{\mathcal{R}_i} := J^{\ell}_{\mathcal{R}}$.

5. Implementation and case studies

The global optimization algorithm presented in the previous section was implemented using MATLAB 5.3 (The MathWorks, Inc., 1999).

The upper and lower bound generation steps require the solution of NLP problems. The function fmincon was used, which is an implementation of a general NLP solver provided by the Optimization Toolbox 2.0 (Coleman et al., 1999) for MATLAB. This solver uses either a subspace trust region method based on the interior-reflective Newton method, or a sequential quadratic programming method.

The solution of the upper and lower bounding problems also requires the integration of the ODE system. The sensitivity equations are integrated together with the original ODE. In order to formulate the lower bounding problem the equality constraints can be overestimated using the constant bounds given by inequalities (22). Additionally the α-based underestimators given by Eqs. (29) and (30) can be used. The integration of another ODE, generated based on Remark 3.1, is needed for the production of these bounds on the state variables and their second-order sensitivities. The MATLAB function ode45 (Shampine and Reichelt, 1997) was used in all cases. It is an implementation of a Runge-Kutta method based on the Dormand-Prince pair. The interval calculations needed at each function evaluation of the latter ODE system were implemented explicitly where possible. Otherwise, an interval arithmetic package for MATLAB, called INTLAB (Rump, 1999a,b), was used. It is worth noting that this prototype implementation is not optimized. CPU times are therefore reported only for the purpose of comparing different bounding and branching strategies within the present implementation.

In the next subsections four examples are studied. The first one is a simple optimal control problem. The next two are parameter estimation problems in chemical kinetics modeling. The last example is a well-known optimal control problem. All the case studies were solved on an Ultra-60 workstation (2 $\times$ 360 MHz UltraSPARC-II CPU, 512MB RAM).

5.1. CASE STUDY 1: A SIMPLE OPTIMAL CONTROL PROBLEM

This example is an optimal control problem with one constant control. The differential system (6)–(8) considered in Section 3 is included in the constraints. This problem has at least two local minima. Its formulation is given by:

Table 1. Case study 1: Global optimization results

Underestimation scheme	Branching strategy	ϵ_r	Iter.	CPU time (s)
Constant	1	1.00e-07	2	7
Constant and α-based	1	1.00e-07	1	25

$$\begin{aligned} \min_{v} \quad & -x(1)^2 \\ \text{s.t.} \quad & \dot{x}(t) = -x(t)^2 + v \quad \forall t \in [0,1] \\ & x(0) = 9 \\ & -5 \leqslant v \leqslant 5 \end{aligned} \tag{33}$$

The procedure followed to form the lower bounding problem is presented analytically in Appendix C. The global optimum parameter is $v = -5$ and the value of the objective function for the global optimum parameter is equal to -8.23262. The results are presented in Table 1. The upper bound calculation was performed only once. Since there is only one parameter in this problem, it is selected for branching at every iteration.

When both the constant and the α-based underestimation schemes are used, although the number of iterations is halved, the CPU time is much larger because of the time spent on the integration of the ODE that produces the bounds on the state variables and their second-order sensitivities.

5.2. CASE STUDY 2: A FIRST-ORDER IRREVERSIBLE LIQUID-PHASE REACTION

The second example is a parameter estimation problem with two parameters and two differential equations in the constraints. It appears in Tjoa and Biegler (1991) as well as in Floudas et al. (1999) and Esposito and Floudas (2000b). It involves a first-order irreversible isothermal liquid-phase chain reaction:

$$A \xrightarrow{k_1} B \xrightarrow{k_2} C$$

The problem can be formulated as follows:

$$\begin{aligned} \min_{k_1,k_2} \quad & \sum_{j=1}^{10}\sum_{i=1}^{2}(x_i(t_j) - x_i^{exp}(t_j))^2 \\ \text{s.t.} \quad & \dot{x}_1(t) = -k_1 x_1(t) \qquad \forall t \in [0,1] \\ & \dot{x}_2(t) = k_1 x_1(t) - k_2 x_2(t) \\ & x_1(0) = 1 \\ & x_2(0) = 0 \\ & 0 \leqslant k_1 \leqslant 10 \\ & 0 \leqslant k_2 \leqslant 10 \end{aligned} \tag{34}$$

Table 2. Case study 2: Global optimization results

Underestimation scheme	Branching strategy	ϵ_r	Iter.	CPU time (s)
Constant	1	1.00e-02	3,501	2,828
Constant	1	1.00e-03	34,508	22,959
Constant and α-based	1	1.00e-02	31	396
Constant and α-based	1	1.00e-03	35	420
Constant and α-based	2	1.00e-02	27	366
Constant and α-based	2	1.00e-03	31	407

where x_1 and x_2 are the mole fractions of components A and B, respectively. k_1 and k_2 are the rate constants of the first and second reaction, respectively. $x_i^{exp}(t_j)$ is the experimental point for the state variable i at time t_j. The points used in the present work are taken from Floudas et al. (1999).

The global optimum parameters are $k_1 = 5.0035$ and $k_2 = 1.0000$ and the value of the objective function for the global optimum parameters is equal to 1.18562e-06. The results are presented in Table 2. The upper bound calculation was performed once every 100 iterations.

While using only the constant bounds for the overestimation of the equality constraints, the second branching strategy cannot be applied because the lower bounding problem does not depend on the parameter. When the α-based underestimation scheme is used additionally to the constant bounds, the number of iterations needed is decreased by several orders of magnitude. For the same type of underestimation and for both branching strategies only four iterations more are needed if the optimality margin, ϵ_r, is decreased from 1.00e-02 to 1.00e-03. Both optimality margins require four iterations less when branching strategy 2 is used instead of branching strategy 1.

It can be easily shown from the analytical solution of the first equation of the ODE system that the function $x_1(t)$ does not depend on k_2 and that it is a convex function of k_1. The use of second-order information identifies this fact automatically and sets the values of all the elements of the vectors $\alpha^+_{1\cdot 2}$, $\alpha^-_{1\cdot 2}$ and $\alpha^+_{1\cdot 1}$ to zero. The values of the elements of the vector $\alpha^-_{1\cdot 1}$ at the root node of the BB tree range from 0.005 to 0.5. When branching strategy 2 is used, they range from 0.003 to 0.01 after five iterations (level 5). The range is the same for all the levels of the BB tree greater than 5. At the root node the values of all the other parameters $\alpha^+_{2\cdot\cdot}$ and $\alpha^-_{2\cdot\cdot}$ range from 0.002 to 2.5. When branching strategy 2 is used, they range from 0 to 0.8 after five iterations (level 5). The maximum value decreases to 0.15 as branching occurs. The magnitude of the values is small and is reduced from level to level, quickly driving the system to convergence.

5.3. CASE STUDY 3: CATALYTIC CRACKING OF GAS OIL

This example is a parameter estimation problem with three parameters and two differential equations in the constraints. It appears in Tjoa and Biegler (1991) as well as in Floudas et al. (1999) and Esposito and Floudas (2000b). It involves an overall reaction of catalytic cracking of gas oil (A) to gasoline (Q) and other products (S):

$$\begin{array}{ccc} A & \xrightarrow{k_1} & Q \\ k_3 \searrow & & \swarrow k_2 \\ & S & \end{array}$$

The problem can be formulated as follows:

$$\begin{aligned} \min_{k_1,k_2,k_3} \quad & \sum_{j=1}^{20} \sum_{i=1}^{2} (x_i(t_j) - x_i^{\exp}(t_j))^2 \\ \text{s.t.} \quad & \dot{x}_1(t) = -(k_1 + k_3)x_1(t)^2 \qquad \forall t \in [0, 0.95] \\ & \dot{x}_2(t) = k_1 x_1(t)^2 - k_2 x_2(t) \\ & x_1(0) = 1 \\ & x_2(0) = 0 \\ & 0 \leqslant k_1 \leqslant 20 \\ & 0 \leqslant k_2 \leqslant 20 \\ & 0 \leqslant k_3 \leqslant 20 \end{aligned} \tag{35}$$

where x_1 and x_2 are the mole fractions of components A and Q, respectively. k_1, k_2 and k_3 are the rate constants of the respective reactions. $x_i^{\exp}(t_j)$ is the experimental point for the state variable i at time t_j. The points used in the present work are again taken from Floudas et al. (1999).

The global optimum parameters are $k_1 = 12.2141$, $k_2 = 7.9799$ and $k_3 = 2.2215$ and the value of the objective function for the global optimum parameters is equal to 2.65567e-03. The results are presented in Table 3. The upper bound calculation was performed once every 100 iterations.

While using only the constant bounds, a maximum number of iterations was set. This number was reached and the algorithm was terminated. The relative optimality obtained is reported. When the α-based underestimation scheme is used additionally to the constant bounds, the number of iterations needed for even smaller optimality margins is decreased by several orders of magnitude. Again, less iterations are needed when branching strategy 2 is applied.

Table 3. Case study 3: Global optimization results

Underestimation scheme	Branching strategy	ϵ_r	Iter.	CPU time (s)
Constant	1	6.41e-02	10,000	16,729
Constant	1	1.33e-02	100,000	152,816
Constant and α-based	1	1.00e-02	73	11,415
Constant and α-based	1	1.00e-03	88	13,524
Constant and α-based	2	1.00e-02	65	10,116
Constant and α-based	2	1.00e-03	81	12,300

5.4. CASE STUDY 4: OPTIMAL CONTROL WITH AN END-POINT CONSTRAINT

This example is an optimal control problem with one control, two differential equations and one end-point constraint. It appears in Goh and Teo (1988) as well as in Dadebo and McAuley (1995). The formulation of the problem is:

$$\begin{aligned}
\min_{u(t),\,\forall t\in[0,1]} \quad & x_2(1) \\
\text{s.t.} \quad & \dot{x}_1(t) = u(t) \qquad \forall t \in [0,1] \\
& \dot{x}_2(t) = x_1(t)^2 + u(t)^2 \\
& x_1(0) = 1 \\
& x_2(0) = 0 \\
& x_1(1) = 1
\end{aligned} \tag{36}$$

Control parameterization is used in order to transform the dynamic optimization problem from infinite dimensional to a finite optimization problem. One finite element is used and the control is approximated by a linear function of t. The problem is then transformed to the following:

$$\begin{aligned}
\min_{u_1,u_2} \quad & x_2(1) \\
\text{s.t.} \quad & \dot{x}_1(t) = u_1(1-t) + u_2 t \qquad \forall t \in [0,1] \\
& \dot{x}_2(t) = x_1(t)^2 + (u_1(1-t) + u_2 t)^2 \\
& x_1(0) = 1 \\
& x_2(0) = 0 \\
& x_1(1) = 1 \\
& -1 \leqslant u_1 \leqslant 1 \\
& -1 \leqslant u_2 \leqslant 1
\end{aligned} \tag{37}$$

where bounds are imposed on the parameters.

The global optimum parameters are $u_1 = -0.4545$ and $u_2 = 0.4545$ and the value of the objective function for the global optimum parameters is equal

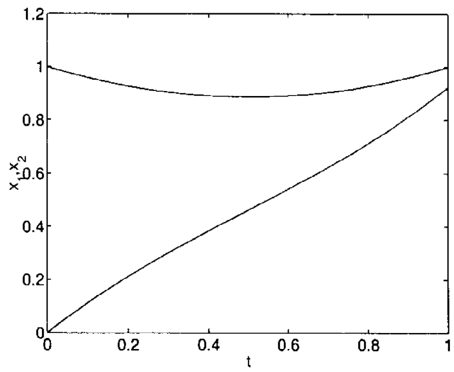


Figure 4. Case study 4: The state variable trajectories for the global optimum parameters.

to 9.24242e-01. The state variable trajectories for the global optimum parameters are shown in Figure 4. The results are presented in Table 4. The upper bound calculation was performed once every 100 iterations.

While using only the constant bounds, a large number of iterations is needed for convergence to the relative optimality margin that has been set. When the α-based underestimating scheme is additionally used, the first lower bound calculated is equal to the upper bound, achieving the relative optimality reported. It can be easily shown that the solution of the ODE system results in an affine function with respect to the parameters for $x_1(1)$ and a convex function for $x_2(1)$. The use of second-order information identifies this fact automatically. A significant advantage of the general α-based underestimators proposed by Androulakis et al. (1995) is that, when a function is automatically identified as convex through the analysis of its second-order information, no convex relaxation is applied. This can improve convergence dramatically as is the case here.

Dadebo and McAuley (1995) presented the analytical solution of problem (36) which is given by a feedback control. For this analytical solution the objective function is again equal to 0.92424. The analytically calculated control and the linear one are presented in Figure 5. Little difference between them can be observed.

Table 4. Case study 4: Global optimization results

Underestimation scheme	Branching strategy	ϵ_r	Iter.	CPU time (s)
Constant	1	1.00e-02	302	317
Constant	1	1.00e-03	1,062	1,106
Constant and α-based	1 or 2	1.12e-13	0	8

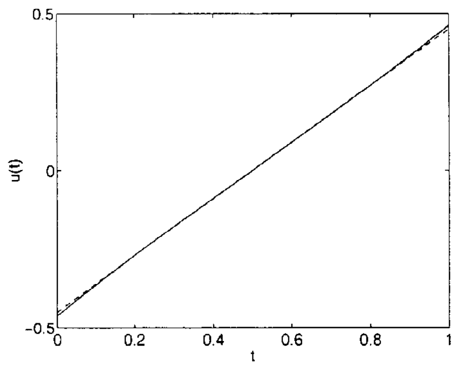


Figure 5. Case study 5: The analytically calculated control (—) and the linear one (– –) calculated using the global optimization algorithm.

6. Conclusions

Existing algorithms for the optimization of systems which are described by ODEs can produce only local optimum solutions that can be used as an upper bound for the global optimum. A new deterministic spatial BB algorithm has been developed for the global optimization of such systems.

The sequential approach was used for the local solution of the dynamic optimization problem. The ODE system was decoupled from the NLP formulation and was integrated using well established techniques.

The global optimization of the resulting NLP problem requires the formulation of its convex relaxation. Well known techniques have been used in order to underestimate specific kinds of terms that may appear in the formulation of the problem. The equality constraints have been overestimated using constant bounds and convex α-based underestimators. A new approach has been developed for the calculation of these constant bounds and the values of the α parameters. This approach is based on the overestimation of the space of solutions of the parameter dependent ODE system that results when the original system is coupled with the first and the second-order sensitivity equations. The concept of differential inequalities has been used.

The structure of the algorithm has also been presented and analyzed. A prototype implementation was used to solve four case studies. When α-based underestimators are used additionally to the constant bounds there is an increase in the size of the ODE systems that are integrated. This latter increase is due to the information needed for the bounds on the second-order sensitivities. However, the results suggest that the algorithm converges in far fewer iterations than in the case of using the constant bounds alone, and usually requires less CPU time. Branching strategy 2, which quantifies the influence of the variables on the tightness of the underestimators, results in less iterations than branching strategy 1. To further investigate these performance related issues current work is focusing on larger systems.

Appendix

A. Theorems on continuity and differentiability of the solution of ODE systems

An ODE system whose right-hand side depends on certain parameters is considered. Theorems on continuity and differentiability of the solution of this system are presented taken from the book of Pontryagin (1962). The final remark makes use of these theorems to present extended continuity and differentiability results.

The system that is considered is given by:

$$\dot{x} = f(t, x, p) \tag{38}$$

where $t \in \Re$, $x(t)$ and $\dot{x}(t) \in \Re^n$, $p \in \Re^r$ and $f : \Re \times \Re^n \times \Re^r \mapsto \Re^n$. It is assumed that the right-hand side

$$f(t, x, p)$$

and its partial derivatives

$$\frac{\partial f(t, x, p)}{\partial x}$$

are defined and are continuous in some domain Γ.

THEOREM A.1 (Pontryagin (1962)). *If (t_0, x_0, p_0) is an arbitrary point of the domain Γ, there exist positive numbers ρ_1 and ρ_2 such that for*

$$|p - p_0| < \rho_1$$

the solution

$$x = \varphi(t, p)$$

of Eq. (38), which satisfies the initial condition

$$\varphi(t_0, p) = x_0,$$

is defined on the interval $|t - t_0| < \rho_2$ and is a continuous function of all the variables t and p on which it depends.

THEOREM A.2 (Pontryagin (1962)). *Let the partial derivatives*

$$\frac{\partial f(t, x, p)}{\partial p}$$

of the right-hand side of system (38) exist and be continuous in the domain Γ. Let (t_0, x_0, p_0) be some point of Γ. Then there exist positive numbers ρ_1' and ρ_2' such *that for* $|t - t_0| < \rho_2'$, $|p - p_0| < \rho_1'$ *the solution* $\varphi(t, p)$ *of equation (38), which satisfies the initial condition*

$$\varphi(t_0, p) = x_0,$$

has continuous partial derivatives

$$\frac{\partial \varphi(t, p)}{\partial p}.$$

COROLLARY A.1 (Pontryagin (1962)). *If all the partial derivatives of the right-hand side of system (38) with respect to the variables* x *and* p *up to the* m*th-order inclusive exist and are continuous, then the function* $\varphi(t, p)$ *forming the solution of equation (38) and satisfying the initial condition* $\varphi(t_0, p) = x_0$*, also has continuous partial derivatives with respect to the parameters* p *up to the* m*th-order inclusive.*

REMARK A.1 (Adapted from Pontryagin (1962)) If the initial value is actually a function of the parameters $x_0(p)$ then the solution of system (38) depends also on the initial values and is written in the form:

$$x = \varphi(t, p, x_0(p)) \tag{39}$$

Let $(t_0, x_0(p_0), p_0)$ be an arbitrary point of the domain Γ. The following variable transformation can be applied:

$$x = x_0(p) + y \tag{40}$$

Then system (38) can be written:

$$\dot{y} = f(t, x_0(p) + y, p) \tag{41}$$

Since the function $f(t, x, p)$ of the variables t, x, p is defined in Γ, the function

$$g(t, y, p, x_0(p)) = f(t, x_0(p) + y, p) \tag{42}$$

of the variables $t, y, p, x_0(p)$ is defined under the condition that the point $(t, x_0(p)+ y, p)$ belongs to Γ. This condition distinguishes a certain domain Γ^* in the space of variables $t, y, p, x_0(p)$, and in this domain the function $g(t, y, p, x_0(p))$ is continuous and has continuous partial derivatives with respect to y and $x_0(p)$. Let

$$y = \psi(t, p, x_0(p))$$

be a solution of Eq. (41) which satisfies the initial condition

$$\psi(t_0, p, x_0(p)) = 0.$$

Then using the variable transformation (40) and Eq. (39) the solution of (38) is given by:

$$x = \varphi(t, p, x_0(p)) = x_0(p) + \psi(t, p, x_0(p)) \tag{43}$$

which obviously satisfies the initial condition

$$\varphi(t_0, p, x_0(p)) = x_0(p). \tag{44}$$

By applying Theorem A.1 to Eq. (41) it can be deduced that solution (43) is continuous with respect to t and p if $x_0(p)$ is continuous with respect to p. By applying Theorem A.2 it can also be deduced that solution (43) is continuously differentiable with respect to p if $x_0(p)$ is continuously differentiable with respect to p. Finally, by applying Corollary A.1 to Eq. (41) it can be shown that solution (43) has continuous partial derivatives with respect to the parameters, p, up to the mth-order if $x_0(p)$ is mth-order continuously differentiable with respect to p.

B. Proof of Theorem 3.3

Proof. Using the inclusion isotonicity property of interval operations the following are true $\forall p \in [p^L, p^U]$:

$$\begin{aligned} &x_0(p) \in x_0([p^L, p^U]) \Rightarrow \\ &\inf x_0([p^L, p^U]) \leqslant x_0(p) \leqslant \sup x_0([p^L, p^U]) \end{aligned} \tag{45}$$

If $\underline{x}(t), \overline{x}(t) \in \mathcal{X}$ satisfy the following inequalities:

$$\underline{x}(0) \leqslant x_0([p^L, p^U]) \leqslant \overline{x}(0)$$

then

$$\underline{x}(0) \leqslant \inf x_0([p^L, p^U]) \leqslant \sup x_0([p^L, p^U]) \leqslant \overline{x}(0)$$

and using inequality (45):

$$\underline{x}(0) \leqslant x_0(p) \leqslant \overline{x}(0), \ \forall p \in [p^L, p^U] \tag{46}$$

The following are also true $\forall p \in [p^L, p^U]$, $\forall t \in \mathcal{I}_0$ and $k = 1, 2, \ldots, n$:

$$\begin{aligned} &f_k(t, \underline{x}_k(t), X_{k^-}(t), p) \in f_k(t, \underline{x}_k(t), X_{k^-}(t), [p^L, p^U]) \\ &\quad \Rightarrow \inf f_k(t, \underline{x}_k(t), X_{k^-}(t), [p^L, p^U]) \leqslant f_k(t, \underline{x}_k(t), X_{k^-}(t), p) \end{aligned} \tag{47}$$

and

$$\begin{aligned} &f_k(t, \overline{x}_k(t), X_{k^-}(t), p) \in f_k(t, \overline{x}_k(t), X_{k^-}(t), [p^L, p^U]) \\ &\Rightarrow \sup f_k(t, \overline{x}_k(t), X_{k^-}(t), [p^L, p^U]) \geqslant f_k(t, \overline{x}_k(t), X_{k^-}(t), p) \end{aligned} \tag{48}$$

where $X_{k^-}(t) = [\underline{x}_{k^-}(t), \overline{x}_{k^-}(t)]$.

If $\underline{x}(t), \overline{x}(t) \in \mathcal{X}$ satisfy the following inequalities:

$$\begin{aligned} &\dot{\underline{x}}_k(t) \leqslant f_k(t, \underline{x}_k(t), [\underline{x}_{k^-}(t), \overline{x}_{k^-}(t)], [p^L, p^U]) \\ &\dot{\overline{x}}_k(t) \geqslant f_k(t, \overline{x}_k(t), [\underline{x}_{k^-}(t), \overline{x}_{k^-}(t)], [p^L, p^U]) \\ &\quad \forall t \in \mathcal{I}_0 \text{ and } k = 1, 2, \ldots, n \end{aligned}$$

then

$$\begin{aligned} &\dot{\underline{x}}_k(t) \leqslant \inf f_k(t, \underline{x}_k(t), [\underline{x}_{k^-}(t), \overline{x}_{k^-}(t)], [p^L, p^U]) \\ &\dot{\overline{x}}_k(t) \geqslant \sup f_k(t, \overline{x}_k(t), [\underline{x}_{k^-}(t), \overline{x}_{k^-}(t)], [p^L, p^U]) \\ &\quad \forall t \in \mathcal{I}_0 \text{ and } k = 1, 2, \ldots, n \end{aligned}$$

and using inequalities (47)–(48):

$$\begin{aligned} &\dot{\underline{x}}_k(t) \leqslant f_k(t, \underline{x}_k(t), [\underline{x}_{k^-}(t), \overline{x}_{k^-}(t)], p) \\ &\dot{\overline{x}}_k(t) \geqslant f_k(t, \overline{x}_k(t), [\underline{x}_{k^-}(t), \overline{x}_{k^-}(t)], p) \\ &\quad \forall p \in [p^L, p^U], \ \ \forall t \in \mathcal{I}_0 \text{ and } k = 1, 2, \ldots, n \end{aligned} \tag{49}$$

Based on Theorem 3.1 and inequalities (46) and (49) $\underline{x}(t)$ is a subfunction and $\overline{x}(t)$ is a superfunction for the solution of ODE (11)–(12) $\forall p \in [p^L, p^U]$, i.e.,

$$\underline{x}(t) \leqslant x(t, p) \leqslant \overline{x}(t) \ , \ \ \forall p \in [p^L, p^U] \ , \ \ \forall t \in \mathcal{I}.$$

C. Lower bounding problem formulation for case study 1

The procedure followed in order to construct the lower bounding problem for the first case study is presented. The formulation of the problem for the region $\mathcal{R} = [v^L, v^U]$ is given by:

$$\min_{v} \ -x(1)^2 \tag{50}$$

$$\text{s.t.} \ \ \dot{x}(t) = -x(t)^2 + v \ \ \forall t \in [0, 1] \tag{51}$$

$$x(0) = 9 \tag{52}$$

$$v^L \leqslant v \leqslant v^U \tag{53}$$

This problem is equivalent to:

$$\min_{\hat{x}, v} \ -\hat{x}^2 \tag{54}$$

$$\text{s.t.} \ \ \hat{x} = x_1(1) \tag{55}$$

$$v^L \leqslant v \leqslant v^U \tag{56}$$

where the value of $x_1(1)$ is obtained by solving ODE (51)–(52). When this ODE is differentiated with respect to the parameter, v, the following first-order sensitivity equations are produced:

$$\dot{x}_2(t) = -2x_1(t)x_2(t) + 1\ ,\ \ \forall t \in [0, 1] \tag{57}$$

$$x_2(0) = 0 \tag{58}$$

where

$$x_2(t) = \frac{\partial x_1}{\partial v} \tag{59}$$

and

$$\dot{x}_2(t) = \frac{\partial}{\partial t}\left(\frac{\partial x_1}{\partial v}\right) \tag{60}$$

The second-order sensitivity equations are produced when the system of first-order sensitivity Eqs. (57)–(58) is differentiated once more with respect to the parameter, v:

$$\dot{x}_3(t) \ = \ -2x_2(t)^2 - 2x_1(t)x_3(t)\ ,\ \ \forall t \in [0, 1] \tag{61}$$

$$x_3(0) \ = \ 0 \tag{62}$$

where

$$x_3(t) = \frac{\partial x_2}{\partial v} \tag{63}$$

and

$$\dot{x}_3(t) = \frac{\partial}{\partial t}\left(\frac{\partial x_2}{\partial v}\right) \tag{64}$$

Based on Remark 3.1 the following ODE system can be constructed:

$$\left.\begin{aligned}
\dot{\underline{x}}_1(t) &= \inf(-\underline{x}_1(t)^2 + [v^L, v^U]) \\
\dot{\underline{x}}_2(t) &= \inf(-2[\underline{x}_1(t), \overline{x}_1(t)]\underline{x}_2(t) + 1) \\
\dot{\underline{x}}_3(t) &= \inf(-2[\underline{x}_2(t), \overline{x}_2(t)]^2 - 2[\underline{x}_1(t), \overline{x}_1(t)]\underline{x}_3(t)) \\
\dot{\overline{x}}_1(t) &= \sup(-\overline{x}_1(t)^2 + [v^L, v^U]) \\
\dot{\overline{x}}_2(t) &= \sup(-2[\underline{x}_1(t), \overline{x}_1(t)]\overline{x}_2(t) + 1) \\
\dot{\overline{x}}_3(t) &= \sup(-2[\underline{x}_2(t), \overline{x}_2(t)]^2 - 2[\underline{x}_1(t), \overline{x}_1(t)]\overline{x}_3(t))
\end{aligned}\right\} \quad \forall t \in [0, 1] \tag{65}$$

$$\begin{aligned}
\underline{x}_1(0) &= 9 \\
\underline{x}_2(0) &= 0 \\
\underline{x}_3(0) &= 0 \\
\overline{x}_1(0) &= 9 \\
\overline{x}_2(0) &= 0 \\
\overline{x}_3(0) &= 0
\end{aligned} \tag{66}$$

The solution of this system gives bounds for the set of solutions of the system consisting of the original system (51)–(52), the first and the second-order sensitivity Eqs. (57)–(58) and (61)–(62):

$$\underline{x}_i(t) \leqslant x_i(t,v) \leqslant \overline{x}_i(t)\ ,\ \forall v \in [v^L, v^U]\ ,\ \forall t \in [0,1]\ ,\ i = 1,2,3 \qquad (67)$$

Using Eqs. (59), (63), (67) and interval arithmetic properties the following are true:

$$\nabla^2 x_1(1,v) \in [\underline{x}_3(1), \overline{x}_3(1)]\ ,\ \forall v \in [v^L, v^U] \qquad (68)$$

$$\nabla^2(-x_1)(1,v) \in [-\overline{x}_3(1), -\underline{x}_3(1)]\ ,\ \forall v \in [v^L, v^U] \qquad (69)$$

When the scaled Gerschgorin method (24) proposed by Adjiman et al. (1998b) is applied on the interval Hessians defined by (68) and (69) the values for the parameters α^+ and α^- can be calculated. Finally, the objective function is underestimated using the secant underestimator, the constant bounds on $\hat{x}$ are defined by equation (67) and the α-based parameter dependent overestimation of the equality constraint (55) is constructed based on Eqs. (25)–(30). The convex relaxation of the problem for the region $\mathcal{R} = [v^L, v^U]$ is given by:

$$\min_{\hat{x},v}\ \{-[\underline{x}_1(1) + \overline{x}_1(1)]\hat{x} + \underline{x}_1(1)\overline{x}_1(1)\} \qquad (70)$$

$$\text{s.t.}\quad \underline{x}_1(1) \leqslant \hat{x} \leqslant \overline{x}_1(1) \qquad (71)$$

$$x_1(1,v) + \alpha^+(v^L - v)(v^U - v) - \hat{x} \leqslant 0 \qquad (72)$$

$$\hat{x} - x_1(1,v) + \alpha^-(v^L - v)(v^U - v) \leqslant 0 \qquad (73)$$

$$v^L \leqslant v \leqslant v^U \qquad (74)$$

where the value of $x_1(1,v)$ is obtained by solving ODE (51)–(52).

References

Adams, E.: 1980, On methods for the construction of the boundaries of sets of solutions for differential equations or finite-dimensional approximations with input sets. In: G. Alefeld and R. D. Grigorieff (eds.): *Fundamentals of numerical computation*, Computing Supplement 2. Berlin: Springer, pp. 1–16.

Adjiman, C. S., I. P. Androulakis, and C. A. Floudas: 1998a, A global optimization method, αBB, for general twice-differentiable constrained NLPs–II. Implementation and computational results. *Comput. Chem. Eng.* 22(9), 1159–1179.

Adjiman, C. S., S. Dallwig, C. A. Floudas, and A. Neumaier: 1998b, A global optimization method, αBB, for general twice-differentiable constrained NLPs–I. Theoretical advances. *Comput. Chem. Eng.* 22(9), 1137–1158.

Adjiman, C. S. and C. A. Floudas: 1996, Rigorous convex underestimators for general twice-differentiable problems. *J. Global Opt.* 9, 23–40.

Allen, M. P. and D. J. Tildesley: 1987, *Computer simulation of liquids*. Oxford: Clarendon Press.

Androulakis, I. P., C. D. Maranas, and C. A. Floudas: 1995, αBB: A global optimization method for general constrained nonconvex problems. *J. Global Opt.* 7, 337–363.

Banga, J. R. and W. D. Seider: 1996, Global optimization of chemical processes using stochastic algorithms. In: *State of the art in global optimization*, Series in nonconvex optimization and its applications. Dordrecht: Kluwer Academic Publishers, pp. 563–583.

Banks, R. B.: 1994, *Growth and diffusion phenomena. Mathematical frameworks and applications*. Springer, Berlin.

Berz, M., K. Makino, and J. Hoefkens: 2001, Verified integration of dynamics in the solar system. *Nonlinear Analysis: Theory, Methods and Applications* 47, 179–190.

Biegler, L. T.: 1984, Solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation. *Comput. Chem. Eng.* 8(3/4), 243–248.

Boender, C. G. E. and H. E. Romeijn: 1995, Stochastic methods. In: R. Horst and P. M. Pardalos (eds.): *Handbook of global optimization*, Series in nonconvex optimization and its applications. Dordrecht: Kluwer Academic Publishers, pp. 829–869.

Bojkov, B. and R. Luus: 1993, Evaluation of the parameters used in iterative dynamic programming. *Can. J. Chem. Eng.* 71, 451–459.

Carrasco, E. F. and J. R. Banga: 1997, Dynamic optimization of batch reactors using adaptive stochastic algorithms. *Ind. Eng. Chem. Res.* 36(6), 2252–2261.

Coleman, T., M. A. Branch, and A. Grace: 1999, Optimization toolbox. For use with MATLAB. User's guide, Ver. 2. The MathWorks, Inc.

Corliss, G. F.: 1995, Guaranteed error bounds for ordinary differential equations. In: J. Levesley, W. A. Light, and M. Marletta (eds.): *Theory of numerics in ordinary and partial differential equations*. Oxford: Oxford Univ. Press, pp. 1–75.

Cuthrell, J. E. and L. T. Biegler: 1987, On the optimisation of differential-algebraic process systems. *AIChE J.* 33(8), 1257–1270.

Dadebo, S. A. and K. B. McAuley: 1995, Dynamic optimization of constrained chemical engineering problems using dynamic programming. *Comput. Chem. Eng.* 19(5), 513–525.

Esposito, W. and C. Floudas: 2000a, Deterministic global optimization in nonlinear optimal control problems. *J. Global Opt.* 17(1/4), 97–126.

Esposito, W. R. and C. A. Floudas: 2000b, Global optimization for the parameter estimation of differential-algebraic systems. *Ind. Eng. Chem. Res.* 39, 1291–1310.

Esposito, W. R. and C. A. Floudas: 2000c, Global optimization of nonconvex problems with differential-algebraic constraints. In: S. Pierucci (ed.): *European Symposium on Computer Aided Process Engineering – 10*. pp. 73–78, Elsevier.

Floudas, C. A.: 1999, *Deterministic global optimization: Theory, methods and applications*, Series in nonconvex optimization and its applications. Dordrecht: Kluwer Academic Publishers.

Floudas, C. A., P. M. Pardalos, C. S. Adjiman, W. R. Esposito, Z. H. Gümüs, S. T. Harding, J. L. Klepeis, C. A. Meyer, and C. A. Schweiger: 1999, *Handbook of test problems in local and global optimization*, Series in nonconvex optimization and its applications. Dordrecht: Kluwer Academic Publishers.

Goh, C. J. and K. L. Teo: 1988, Control parameterization: a unified approach to optimal control problems with general constraints. *Automatica* 24(1), 3–18.

Horst, R. and H. Tuy: 1996, *Global Optimization. Deterministic approaches*. Berlin: Springer-Verlag, 3rd edition.

Jackson, L.: 1975, Interval arithmetic error-bounding algorithms. *SIAM J. Numer. Anal.* 12(2), 223–238.

Krückeberg, F.: 1969, Ordinary differential equations. In: E. Hansen (ed.): *Topics in interval analysis*. Oxford: Clarendon Press, pp. 91–97.

Kühn, W.: 1998, Rigorously computed orbits of dynamical systems without the wrapping effect. *Computing* 61, 47–67.

Lakshmikantham, V. and S. Leela: 1969, *Differential and integral inequalities. Theory and applications*, Vol. 55-I of *Series in mathematics in science and engineering*. New York: Academic Press.

Logsdon, J. S. and L. T. Biegler: 1989, Accurate solution of differential algebraic optimization problems. *Ind. Eng. Chem. Res.* 28(11), 1628–1639.

Lohner, R. J.: 1987, Enclosing the solutions of ordinary initial and boundary problems. In: E. W. Kaucher, U. W. Kulisch, and C. Ullrich (eds.): *Computer arithmetic: Scientific computation and programming languages*, Series in Computer Science. Stuttgart: Wiley-Teubner, pp. 255–286.

Luus, R.: 1990a, Application of dynamic programming to high-dimensional nonlinear optimal control problems. *Int. J. Control* 52(1), 239–250.

Luus, R.: 1990b, Optimal control by dynamic programming using systematic reduction in grid size. *Int. J. Control* 51(5), 995–1013.

Luus, R.: 1993, Piecewise linear continuous optimal control by iterative dynamic programming. *Ind. Eng. Chem. Res.* 32, 859–865.

Luus, R. and B. Bojkov: 1994, Global optimization of the bifunctional catalyst problem. *Can. J. Chem. Eng.* 72, 160–163.

Luus, R. and D. E. Cormack: 1972, Multiplicity of solutions resulting from the use of variational methods in optimal control problems. *Can. J. Chem. Eng.* 50, 309–311.

Luus, R. and O. Rosen: 1991, Application of iterative dynamic programming to final state constrained optimal control problems. *Ind. Eng. Chem. Res.* 30(7), 1525–1530.

Luus, R., J. Dittrich, and F. J. Keil: 1992, Multiplicity of solutions in the optimization of a bifunctional catalyst blend in a tubular reactor. *Can. J. Chem. Eng.* 70, 780–785.

Maranas, C. D. and C. A. Floudas: 1994, Global minimum potential energy conformations of small molecules. *J. Global Opt.* 4, 135–170.

McCormick, G. P.: 1976, Computability of global solutions to factorable nonconvex programs: Part I — Convex underestimating problems. *Mathematical Programming* 10, 147–175.

Moore, R. E.: 1966, *Interval analysis*. Englewood Cliffs, NJ: Prentice-Hall.

Moore, R. E.: 1978, Bounding sets in function spaces with applications to nonlinear operator equations. *SIAM Review* 20(3), 492–512.

Moore, R. E.: 1979, *Methods and applications of interval analysis*. Philadelphia, PA: SIAM.

Moore, R. E.: 1984, A survey of interval methods for differential equations. In: *Proceedings of 23rd conference on decision and control (Las Vegas)*. pp. 1529–1535.

Nedialkov, N. S., K. R. Jackson, and G. F. Corliss: 1999, Validated solutions of initial value problems for ordinary differential equations. *Appl. Math. Comput.* 105(1), 21–68.

Neumaier, A.: 1993, The wrapping effect, ellipsoid arithmetic, stability and confidence regions. In: R. A. et al. (ed.): *Validation Numerics*, Vol. 9 of *Computing Supplementum*. Wien: Springer, pp. 175–190.

Neumaier, A.: 1994, Global, rigorous and realistic bounds for the solution of dissipative differential equations. Part I: Theory. *Computing* 52, 315–336.

Nickel, K. L.: 1978, Ein Zusammenhang zwischen Aufgaben monotoner Art und Intervall-Mathematik. In: R. Bulirsch and R. D. Grigorieff (eds.): *Numerical treatment of differential equations*, Lecture Notes in Computer Science No 631. Berlin: Springer, pp. 121–132.

Nickel, K. L.: 1983, Bounds for the set of solutions of functional-differential equations. *Annales Polonici Mathematici* 42, 241–257. MRC Technical summary report No 1782, University of Winsconsin, Madison (1977).

Nickel, K. L.: 1985, How to fight the wrapping effect. In: K. L. Nickel (ed.): *Interval Mathematics*, Lecture Notes in Computer Science No 212. Berlin: Springer, pp. 121–132.

Nickel, K. L.: 1986, Using interval methods for numerical solution of ODE's. *ZAMM* 66(11), 513–523.

Oh, S. H. and R. Luus: 1977, Use of orthogonal collocation methods in optimal control problems. *Int. J. Control* 26(5), 657–673.

Pollard, G. P. and R. W. H. Sargent: 1970, Off line computation of optimum controls for a plate distillation column. *Automatica* 6, 59–76.

Pontryagin, L. S.: 1962, *Ordinary differential equations*. London: Pergamon Press.

Renfro, J. G., A. M. Morshedi, and O. A. Asbjornsen: 1987, Simultaneous optimization and solution of systems described by differential/algebraic equations. *Comput. Chem. Eng.* 11(5), 503–517.

Rihm, R.: 1994, On a class of enclosure methods for initial value problems. *Computing* 53, 369–377.

Rosen, O. and R. Luus: 1992, Global optimization approach to nonlinear optimal control. *J. Opt. Th. Appl.* 73(3), 547–562.

Rump, S. M.: 1999a, Fast and parallel interval arithmetic. *BIT* 39(3), 534–554.

Rump, S. M.: 1999b, INTLAB — Interval laboratory. In: T. Csendes (ed.): *Developments in reliable computing*. Dordrecht: Kluwer Academic Publishers, pp. 77–104.

Sargent, R. W. H.: 2000, Optimal control. *J. Comput. Appl. Math.* 124(1/2), 361–371.

Sargent, R. W. H. and G. R. Sullivan: 1978, The development of an efficient optimal control package. In: J. Stoer (ed.): *Proc. 8th IFIP Conf. Optimisation Tech. Pt. 2*. Springer, Berlin: pp. 158–168.

Shampine, L. F. and M. W. Reichelt: 1997, The MATLAB ODE suite. *SIAM J. Sci. Comput.* 18(1), 1–22.

Smith, E. M. B. and C. C. Pantelides: 1996, Global optimisation of general process models. In: I. E. Grossmann (ed.): *Global optimization in engineering design*, Series in nonconvex optimization and its applications. Dordrecht: Kluwer Academic Publishers, Ch. 12, pp. 355–386.

Smith, J. M.: 1981, *Chemical engineering kinetics*. London: McGraw-Hill, 3rd edition.

Stewart, N. F.: 1971, A heuristic to reduce the wrapping effect in the numerical solution of $x' = f(t, x)$. *BIT* 11, 328–337.

The MathWorks, Inc.: 1999, Using MATLAB.

Tjoa, I. B. and L. T. Biegler: 1991, Simultaneous solution and optimization strategies for parameter estimation of differential-algebraic equation systems. *Ind. Eng. Chem. Res.* 30, 376–385.

Tsang, T. H., D. M. Himmelblau, and T. F. Edgar: 1975, Optimal control via collocation and non-linear programming. *Int. J. Control* 21(5), 763–768.

Vassiliadis, V. S., R. W. H. Sargent, and C. C. Pantelides: 1994a, Solution of a class of multistage dynamic optimization problems. 1. Problems without path constraints. *Ind. Eng. Chem. Res.* 33(9), 2111–2122.

Vassiliadis, V. S., R. W. H. Sargent, and C. C. Pantelides: 1994b, Solution of a class of multistage dynamic optimization problems. 2. Problems with path constraints. *Ind. Eng. Chem. Res.* 33(9), 2123–2133.

Villadsen, J. V. and W. E. Stewart: 1967, Solution of boundary-value problems by orthogonal collocation. *Chem. Eng. Sci.* 22, 1483–1501.

Walter, W.: 1970, *Differential and integral inequalities*. Berlin: Springer.