Global Optimization of Nonconvex Problems with Differential-Algebraic Constraints

William R. Esposito and Christodoulos A. Floudas
Department of Chemical Engineering, Princeton University, Princeton, N.J. 08544-5263,USA

Differential-algebraic systems of constraints, in particular, initial value ordinary differential equations, appear in numerous optimization problems in the chemical engineering field. A difficulty in the solution of this formulation which has not been throughly addressed, is the problem of multiple local minima. In this paper, a novel deterministic global optimization method using a sequential approach will be presented.

1. Introduction

The solution of optimization problems involving differential-algebraic constraints arises often in the chemical engineering field. Examples include the optimal control of batch and semi-batch processes as well as the determination of kinetic constants from time series data. Difficulties arise from both numerical and optimization points of view.

Two different approaches exist for the solution of the DAE system within the optimization problem. The first approach, referred to as a simultaneous method, involves the complete discretization of the dynamic system. The resulting formulation is algebraic in nature and can be solved using known nonlinear programming methods. A second approach, referred to as a sequential method, involves the solution of the DAE system through an integration routine at each iteration of the local solver. Control parameterization is also used when necessary. In each case, do to the nonconvex nature of formulation, multiple local minima arise.

In this paper a deterministic global optimization method using the sequential approach will be presented for the solution of dynamic optimization problems. The method is based on the αBB (Adjiman et al., 1998b,a), a branch and bound algorithm for the determination of the global minimum for twice continuously differentiable NLPs.

1.1. Formulation

$$\begin{aligned}
\min_{\boldsymbol{x}\,\boldsymbol{v}} \quad & f(\boldsymbol{x}, \boldsymbol{v}) \qquad (1)\\
\text{s.t.} \quad & \dot{\boldsymbol{z}}_j = \mathbf{g}(\boldsymbol{z}, \boldsymbol{v}, t) \quad j \in J\\
& 0 = \mathbf{h}(\boldsymbol{z}, \boldsymbol{v}, t)\\
& \boldsymbol{z}(t_o) = \boldsymbol{z}_o \quad t \in [t_o, t_f]\\
& c_{m,\mu}(\boldsymbol{x}) + z_m(t_\mu) = 0 \;\; m \in M \;\; \mu \in P
\end{aligned}$$

where $\boldsymbol{z}$ are the state variables, J are the set of states whose derivatives appear explicitly in the system, and M is the set of states which appear in the set of point constants P. $\boldsymbol{v} \in [\boldsymbol{v}^L, \boldsymbol{v}^U]$ are time invariant parameters which appear in the dynamic system, $\boldsymbol{x} \in [\boldsymbol{x}^L, \boldsymbol{x}^U]$ are a set algebraic variables which do not appear in the dynamic system. $f(\boldsymbol{x}, \boldsymbol{v})$ and $\mathbf{c}(\boldsymbol{x})$ are twice continuously differentiable functions, $\mathbf{g}(\boldsymbol{z}, \boldsymbol{v}, t)$ and $\mathbf{h}(\boldsymbol{z}, \boldsymbol{v}, t)$ are twice continuously differential functions with respect to the states, $\boldsymbol{z}$ and the parameters, $\boldsymbol{v}$. Within (1), the control (if it exists) has already been parameterized by, $u = \mathcal{U}(\boldsymbol{v}, t)$.

2. Global Optimization Approach

The proposed approach is based on the αBB (Adjiman et al., 1998b,a), for twice continuously differentiable problems. Within this branch and bound framework, a sequence of upper bounds on the global solution is obtained by solving the full nonconvex problem to local optimality from multiple starting points. A lower bound is determined by solving a valid convex underestimation of original formulation. ϵ-convergence is obtained by successive subdivision of the original region at each level of the branch and bound tree. A through treatment of deterministic global optimization methods and applications can be found in the recent book by Floudas (2000).

2.1. Convex underestimation

The key to the success of the approach is in the ability to generate a valid convex underestimation. The algebraic functions within the formulation ($c_{m,\mu}(\boldsymbol{x})$ and $f(\boldsymbol{x},\boldsymbol{v})$) are underestimated using techniques shown by Adjiman et al. (1998b,a). To underestimate the dynamic part, consider the differential-algebraic system of equations as a simple input-output map.

$$\boldsymbol{v} \longrightarrow \boxed{\begin{array}{rl} \dot{\boldsymbol{z}}_j & = \mathbf{g}(\boldsymbol{z},\boldsymbol{v},t) \quad j \in J \\ \mathbf{0} & = \mathbf{h}(\boldsymbol{z},\boldsymbol{v},t) \\ \boldsymbol{z}_j(t_0) & = \boldsymbol{z}_0 \;\; t \in [t_0, t_f] \end{array}} \longrightarrow \boldsymbol{z}(t) \tag{2}$$

Pontryagin (1962) showed that under the assumptions given previously, this map is continuous and twice differentiable with respect to the parameters, $\boldsymbol{v}$. Therefore, the states, $\boldsymbol{z}$, at given time points, t_μ, can be written as twice continuously differentiable functions:

$$\boldsymbol{z}(t_\mu) \;=\; \boldsymbol{\mathcal{F}}(t_\mu,\boldsymbol{v}) \;\equiv\; \boldsymbol{\mathcal{F}}_\mu(\boldsymbol{v}) \;. \tag{3}$$

Substituting (3) into (1) results in:

$$\begin{aligned} &\min_{\boldsymbol{x},\boldsymbol{v}} \;\; f(\boldsymbol{x},\boldsymbol{v}) \\ &\text{s.t.} \;\; c_{m,\mu}(\boldsymbol{x}) + \mathcal{F}_{m,\mu}(\boldsymbol{v}) \;=\; 0 \;\; m \in M \;\; \mu \in P \end{aligned} \tag{4}$$

The underestimator of this term is generated by adding a quadratic function in $\boldsymbol{v}$

$$\mathcal{F}_{m,\mu}(\boldsymbol{v}) + \sum_{i \in I} \beta_{m,\mu,i} \left(v_i^U - v_i\right)\left(v_i^L - v_i\right) \tag{5}$$

where I is the set of $\boldsymbol{v}$ variables, and $\mathcal{L}_{\mathcal{F}}$ represents the underestimator of the function $\mathcal{F}$. A simplification of (5) has $\beta_{m,\mu,i}$ equal $\forall i \in I$. The value of these β parameters needs to be large enough to ensure convexity, but not too large as to overly underestimate the original function. The full convex underestimator takes the form:

$$\begin{aligned} &\min_{\boldsymbol{x},\boldsymbol{v}} \;\; \mathcal{L}_f(\boldsymbol{x},\boldsymbol{v}) \\ &\text{s.t.} \;\; \mathcal{L}^+_{c_{m,\mu}}(\boldsymbol{x}) \;+\; \mathcal{L}^+_{\boldsymbol{\mathcal{F}}_{m,\mu}}(\boldsymbol{v}) \;\le\; 0, \;\; \mathcal{L}^-_{c_{m,\mu}}(\boldsymbol{x}) \;+\; \mathcal{L}^-_{\boldsymbol{\mathcal{F}}_{m,\mu}}(\boldsymbol{v}) \;\le\; 0 \end{aligned} \tag{6}$$

where $\mathcal{L}^+_{c_{m,\mu}}$ is the underestimator of the function $c_{m,\mu}$, $\mathcal{L}^-_{c_{m,\mu}}$ is the underestimator of the function $-c_{m,\mu}$ and $\mathcal{L}^+_{\boldsymbol{\mathcal{F}}_{m,\mu}}$, $\mathcal{L}^-_{\boldsymbol{\mathcal{F}}_{m,\mu}}$, and $\mathcal{L}_f$ are all similarly defined. It is necessary to split the equality point constraints into a positively and a negatively signed inequality, each being underestimated separately.

2.2. Determination of β Parameters

The β parameters are positive quantities calculated from the Hessian matrix of the function $\mathcal{F}_{m,\mu}(\boldsymbol{v})$. This matrix is generated using the second order sensitivities of the state z_m with respect to the parameters $\boldsymbol{v}$,

$$\mathcal{H}_{m,\mu} \equiv \frac{\partial^2 z_m}{\partial \boldsymbol{v}^2}(t_\mu) \ . \tag{7}$$

In the case when the β parameters are equal for each variable, their values can be calculated by: $\beta_{m,\mu} \geq -1/2 \min_{\boldsymbol{v}} \lambda^{min}_{m,\mu}(\boldsymbol{v})$ (Maranas and Floudas, 1994), where $\lambda^{min}_{m,\mu}$ is the minimum eigenvalue of the Hessian matrix $\mathcal{H}_{m,\mu}$. The difficulty arises from the fact that $\mathcal{H}_{m,\mu}$ can not be written as an analytical function of $\boldsymbol{v}$. The elements of the matrix, however, can be determined through an integration of the augmented system at given values of $\boldsymbol{v}$. As a result of this, three different methods for the determination of β values have been developed.

Constant or Semi-Constant Values: The values used are preselected and can either be a constant or a function of tree level.

Sampling Approach: The values of the elements of $\mathcal{H}$ are determinable at given values of the parameters $\boldsymbol{v}$. In each region, a number of random points are selected to evaluate the Hessian matrix and calculate the eigenvalues of these matrices. The minimum of these eigenvalues is then used to calculate the value of β.

Sampling with Interval Calculations: In this approach, the values of each element of $\mathcal{H}$ are determined at given values of the parameters, but the eigenvalues of the matrices are not directly determined. Instead, an interval Hessian matrix is generated by determining the minimum and maximum of every element over the sampled set. A valid lower bound on the minimum eigenvalue of this matrix can then by determined using methods presented by Adjiman and Floudas (1996); Adjiman et al. (1998b).

2.3. Illustrative Example

In order to illustrate the above concepts, consider the system:

$$\dot{z} = u - z^3, \quad z(t_o) = 9 \quad t \in [0,1] \tag{8}$$

where the control is bounded by $u \in [-5,5]$, Consider the value of the state, z, at $t = 1$, which is an implicit function of the variable u, $\mathcal{F}(u)$. The first order sensitivity of the state with respect to the parameter, $\frac{\partial z}{\partial u}$, is determined by integrating an additional equation with the one given above (Vassiliadis et al., 1994). The second order sensitivity, $\frac{\partial^2 z}{\partial u^2}$, is calculated using a finite difference approximation. The minimum of the second order sensitivity is found to be -0.2214. Therefore a β value of 0.1107 is needed to generate a convex relaxation of this function. The original function and the underestimator are plotted in Figure 1. Notice the continuity of the original function, and the convexity of the underestimator.

3. Algorithmic Outline

Step 1: Initialize the problem:

Set the relative, ϵ^{rel}, or the absolute, ϵ^{abs} convergence tolerance. If a sampling method is being used to calculate the β values, set the number of points to use at the first iteration, $p^{initial}$, and the minimum number at each subsequent iteration, p^{every}. $iter = 0$.

Step 2: Calculate initial β values:

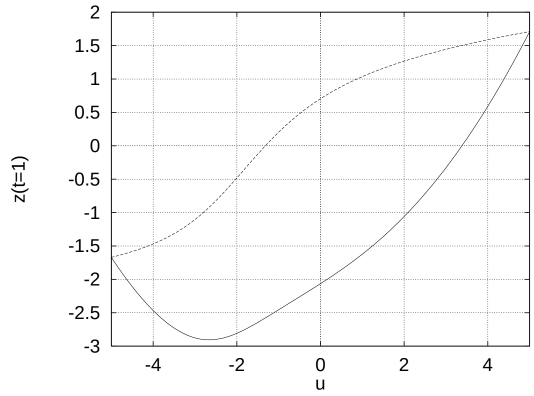

Figure 1: Function and Underestimator for the Illustrative Example.

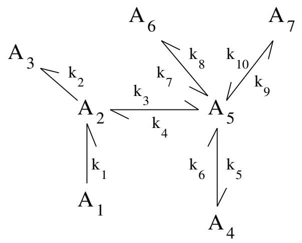

Figure 2: Bifunctional Catalyst Reaction Scheme.

Integrate, with second order sensitivity evaluations, the system at $p^{initial}$ randomly selected values for the $\boldsymbol{v}$ variables. Save this results for latter use. Using the selected method, calculate the values of the β parameters.

Step 3: Initial lower and upper bounds:
Solve the convex relaxation from a randomly chosen starting point. Using the solution as a starting point, solve the original nonconvex problem to determine an upper bound (UB) on the global solution.

Step 4: Check for convergence
If the UB and LB are within either tolerance, then terminate with the UB as the global solution, otherwise $iter = iter + 1$.

Step 5: Update Lower Bound:
Select the region with the lowest relaxed solution to be explored next.

Step 6: Branch the region:
Select the variable which has the largest effect on the underestimation and bisect the region on that variable.

Step 7: Update β values in each region:
Search a list of saved integration points to determine the number which fall in the given region, $p^{current}$. If $p^{current} < p^{every}$ select additional points randomly to make up the difference. Calculate the values of the β parameters.

Step 8: Determine bounds in each region:
Solve the convex relaxation from a randomly chosen starting point. Fathom the region if there is no feasible solution less than the current UB and go to Step 4. Solve the upper problem in this region, updating the UB if necessary, and go to Step 4.

4. Example problems

The solution of an optimal control problem and a parameter estimation problem are presented to illustration the performance of the algorithm.

4.1. Bifunctional Catalyst Problem

This example concerns the optimization of a bifunctional catalyst in converting methylcyclopentane to benzene. The catalyst contains a hydrogenation component and an isomerization component. The objective is to determine the mixture of the two along the length of the reactor which maximizes the concentration of the desired product, A_7, in the reaction scheme given in Figure 2.

The formulation for this problem using a piecewise constant control profile on 10 equally spaced intervals can be found in the in Floudas et al. (1999) and Esposito and Floudas (2000a). The problem exhibits over 300 distinct local solutions. Using 1000 random starting points, only one resulted in the determination of the global solution to the problem. Using 1000 sampling points, the β value need for convexity was determined to be 0.0698. The problem was solved using a β value which started at 0.0698 and was reduced at each level of the tree. A relative convergence tolerance of 0.1% was used. The global solution determined had an objective value of 10.095×10^{-3} with a control profile of: $\boldsymbol{v}$ = [0.66595, 0.67352, 0.67500, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9].

This example illustrates an interesting characteristic of the approach. The algorithm acts as a very effective search to determine the global solution. Initial points for the solution of the original problem are determined by solving a convex relaxation. This relaxation acts to smooth the nonconvex nature of the original formulation, thus supplying starting points which are relatively close to the global solution. Consider in this problem that using 1000 randomly chosen starting points only results in the global solution being identified once. In the proposed approach the global solution was identified in at most 169 iterations and more often less than 100 iterations. Each iteration produces two different starting points for the solution of the upper problem. Therefore, it took no more that 200 points (on average) to find the global solution. This is 5 times better than simply choosing random points.

4.2. Lotka-Voltera Problem

This problem involves the estimation of parameters in the predator-prey model used in ecology. The model is described by two differential equations:

$$\frac{dz_1}{dt} = \theta_1\, z_1\,(1 \;-\; z_2), \quad \frac{dz_2}{dt} = \theta_2\, z_2\,(z_1 \;-\; 1), \quad \boldsymbol{z}_0 = [1.2,\, 1.1], \quad t \in [0,\, 10] \tag{9}$$

where z_1 represents the population of the prey, and z_2 the population of the predator. The solutions to these equations are cyclic in nature and out of phase with each other. The data used in the study was generated with parameter values of $\boldsymbol{\theta} = [3,\, 1]$ at 10 equally spaced time points with a small amount of normally distributed random error with $\sigma = 0.01$ and zero mean added. The problem is formulated as an error-in-variables estimation problem. and the full formulation appears in Floudas et al. (1999) and Esposito and Floudas (2000b). This problem has been shown to have at least 20 local solutions, with the most prevalent solution being the third best (Esposito and Floudas, 2000a).

This problem was solved to global optimality using both of the sampling based approaches. The average and standard deviation of three runs with different numbers of sampled points are shown in Table 1 Notice how increasing the number of sampling points results in less variation within the algorithm. Also, the use of interval methods have been shown to result in looser underestimators. This gives slower convergence to global solution, which is clearly evident in the results. In every case, in spite of the large number of local minima, the global solution was obtained.

5. Conclusions

In this paper a deterministic global optimization approach has been presented to address nonconvex problems with differential-algebraic constraints. The proposed method is based on

No. of Points	Sampled		Sampled/Interval	
$p^{initial}/p^{every}$	Inter.	CPU sec.	Inter.	CPU sec.
25/10	123 ± 14	331.04 ± 56.54	168 ± 4	396.25 ± 20.00
50/25	145 ± 6	540.04 ± 41.50	195 ± 3	699.73 ± 13.24
100/50	152 ± 3	884.14 ± 49.50	199 ± 2	1119.38 ± 10.16

Table 1: Results for the Lotka-Voltera problem.

a branch-and-bound framework in which the solution of a convex relaxation is used to generate a valid lower bound on the global solution. Two examples, a parameter estimation problem and an optimal control problem, were presented to illustrate the computation and theoretical aspects.

6. Acknowledgments

Financial support from the National Science Foundation is gratefully acknowledged.

References

Adjiman C.S., Androulakis I.P., and Floudas C.A., 1998a, A global optimization method, αBB, for general twice–differentiable NLPs – II. Implementation and computational results. *Comp. Chem. Engng.* **22**, 1159–1178.

Adjiman C.S., Androulakis I.P., Floudas C.A., and Neumaier A., 1998b, A global optimization method, αBB, for general twice–differentiable NLPs – I. Theoretical advances. *Comp. Chem. Engng.* **22**, 1137–1158.

Adjiman C.S. and Floudas C.A., 1996, Rigorous convex underestimators for general twice–differentiable problems. *Journal of Global Optimization* **9**, 23–40.

Esposito W.R. and Floudas C.A., 2000a, Deterministic global optimization in nonlinear optimal control problems. *Journal of Global Optimization*, Accepted for Publication.

Esposito W.R. and Floudas C.A., 2000b, Global optimization for the parameter estimation of differential–algebraic systems. *I&EC Res.*, Accepted for Publication.

Floudas C.A., 2000, *Deterministic Global Optimization: Theory, Methods and Applications*. Nonconvex Optimization and its Applications, Kluwer Academic Publishers.

Floudas C.A., Pardalos P.M., Adjiman C.S., Esposito W.R., Gümüş Z., Harding S., Klepeis J., Meyer C., and Schweiger C., 1999, *Handbook of Test Problems in Local and Global Optimization*. Kluwer Academic Publishers.

Maranas C.D. and Floudas C., 1994, Global minimum potential energy conformations of small molecules. *Journal of Global Optimization* **4**, 135.

Pontryagin L.S., 1962, *Ordinary Differential Equations*. Addison-Wesley Publishing Co., Translated from Russian by L. Kacinskas and W. B. Counts.

Vassiliadis V.S., Sargent R.W.H., and Pantelides C.C., 1994, Solution of a class of multistage dynamic optimization problems 1. Problems without path constraints. *I&EC Res.* **33**, 2111–2122.