Parameter Estimation in Nonlinear Algebraic Models via Global Optimization

William R. Esposito and Christodoulos A. Floudas [1]

Department of Chemical Engineering, Princeton University, Princeton, N.J. 08544-5263,USA

Abstract: The estimation of parameters in semi-empirical nonlinear models through the error-in-variables method has been widely studied from a computational standpoint. This method involves the minimization of a quadratic objective function subject to the model equations being satisfied. Due to the nonlinear nature of these models, the resulting formulation is nonconvex in nature. The approaches to solve this problem presented so far in the literature, although computationally efficient, only offer convergence to a local solution of a model which may contain multiple minima. In this paper a global optimization approach based on a branch-and-bound framework will be presented to solve the error-in-variables formulation. Various estimation problems were solved and will be presented to illustrate the theoretical and computational aspects of the proposed method.

Introduction

Mathematical models which accurately predict physical phenomena are essential in the design, control and optimization of chemical processes. These models often contain adjustable parameters which need to be determined from the available experimental data. Due to the complexity of many processes, most models used are nonlinear in nature which adds to an already difficult problem.

Many statistical methods exist for the estimation of the parameters in both linear and nonlinear models (Bard, 1974). One particular approach, known as the error-in-variables formulation has been extensively studied in the literature. This formulation is derived using the principle of maximum likelihood through the assumption that the error in all variables can be described by a normal distribution with zero mean and know diagonal covariance matrix. Two major difficulties arise in addressing this problem. First, since error is accounted for in all variables, the optimization is performed not only over the parameters, but also over a fitted set of data variables. Secondly, and more importantly, due to the nonlinear nature of the models, the resulting optimization problem is nonconvex and may contain many local minima in the area of interest.

Different methods have been proposed to address the solution of this formulation. Bard and Lapidus (1968) provided a description of the early work in the area as it relates to the estimation of kinetic parameters in complex reaction networks. Southwell (1969) presented a method based on a gradient approach to solve models with two variables and minimally nonlinear. Schwetlick and Tiller (1985) extended this approach to include models which have more than two variables and are highly nonlinear in nature. Fariss and Law (1979) presented an approach which modifies the form of the error response to allow for the detection of gross errors in the data. Fabries and Renon (1975) developed a method specifically for thermodynamic models using a Gauss-Newton approach. Anderson et al. (1978) also provided an approach for thermodynamic models, but using two equations to fully describe the system and obtaining the solution by a successive linearization method.

Other methods presented attempt in some way to reduce the domain over which the minimization is performed. Patino-Leal and Reilly (1982) accomplished this by splitting the full problem into inner and outer optimization sub-problems. The outer problem calculates values for the parameter values using given values for the fitted data variables, while the inner problem determines the fitted data variables from given values of the parameters. Valko and Vajda (1987) presented a similar approach using linearized equations to calculated the fitted data variables in the inner problem. Rod and Hancil (1980) presented a method in which full optimization problems are solve at both levels. Dovi and Paladino (1989) presented a slightly different approach which uses constrained variation to write the change in the fitted data variables as a functions of the change in the parameters. Therefore the optimization need now only be performed in the space of the parameter variables. Tjoa and Biegler (1992) presented a similar approach based on successive quadratic programming in which the optimality conditions in terms of the fitted data variables are solved analytically and included in each quadratic subproblem.

Reviews and comparisons of the various methods are given by Ricker (1984); Stewart et al. (1992); Kim et al. (1990).

All the aforementioned approaches only offer convergence to a local solution of a formulation which may have multiple minima. There is currently no method which guarantees convergence to the global minimum and exploits the structure of the mathematical formulation. In this work, a deterministic global optimization algorithm based on a branch and bound framework is proposed to address the error-in-variables formulation.

Error-In-Variables Estimation

The model takes the form of an implicit algebraic system of equations.

$$\mathbf{f}(\theta, \mathbf{z}) = \mathbf{0} \tag{1}$$

where θ is a vector of p unknown parameters, $\mathbf{z}$ is a vector of n experimentally measured variables, and $\mathbf{f}$ represents the system of l algebraic equations.

All experimentally measured variables are affected to some extent by error. The measurements are related to the

[1] Author to whom all correspondence should be addressed

true values through:

$$\mathbf{z}_\mu \;=\; \hat{\mathbf{z}}_\mu \;+\; \mathbf{e}_\mu \quad \mu \;=\; 1 \ldots m \tag{2}$$

where $\hat{\mathbf{z}}_\mu$ is the vector of "true" values of the experimentally measured variables, $\mathbf{z}_\mu$, at the μ data point, and $\mathbf{e}_\mu$ is the vector of additive error. Through the principle of maximum likelihood estimation the error-in-variables formulation is derived using the following assumptions: i) the error can be described by a normal distribution with zero mean and known covariance, ii) the error in each experiment is independent, and iii) the covariance matrix is diagonal and equal in each experiment. For a full derivation see Bard (1974); Esposito and Floudas (1997). The approach results in the following optimization problem:

$$\min_{\hat{\mathbf{z}}_\mu \, \theta} \sum_{\mu=1}^{r} \sum_{i=1}^{m} \frac{(\hat{z}_{\mu,i} \;-\; z_{\mu,i})^2}{\sigma_i^2} \tag{3}$$

s.t.

$$\mathbf{f}(\hat{\mathbf{z}}_\mu \,,\, \theta) \;=\; \mathbf{0} \quad \mu \;=\; 1 \ldots m$$

where σ_i is the standard deviation of variable i in each experiment, and $\hat{\mathbf{z}}_\mu$ are referred to as the fitted data variables. This formulation has the following mathematical properties:

1. The objective function is convex.
2. Since the model equations enter into the formulation as equality constraints, nonlinear models result in nonconvex optimization problems.
3. Since the minimization is not only over the parameters, but also over the fitted data variables, $\hat{\mathbf{z}}_\mu$, even linear models such as $y \;=\; a\,x \;+\; b$ are nonlinear due to the bilinear term $a\,x$.
4. A group of constraints, corresponding to the model equations, are written for each data point, $\mu \;=\; 1 \ldots m$. This results in the fitted data variables, $\hat{\mathbf{z}}_\mu$, appearing in one set of the constraints, while the parameters, θ, appear in every set.
5. An unconstrained formulation is sometimes possible through substitution of the model equations into the objective function.

These properties form the basis for the development of a global optimization approach to solve the error-in-variables formulation.

Global Optimization Approach

In order to solve (3) to global optimality, a recently developed deterministic branch and bound algorithm, αBB (Androulakis et al., 1995; Adjiman et al., 1996, 1997a,b), was extensively modified. First the basic ideas of the αBB algorithm will be presented as applicable to more general problems. Then the application of this approach to the error-in-variables formulation will be discussed.

Basic Ideas of the αBB

The αBB global optimization method guarantees convergence to the global minimum for general twice continuously differentiable constrained NLPs. This is accomplished through the generation of a non-decreasing sequence of lower and a non-increasing sequence of updated upper bounds on the global solution. Finite ϵ-convergence to the global minimum is achieved through the successive subdivision of the region at each level in the branch and bound tree. The sequence of upper bounds on the global solution is obtained by solving, to local optimality, the full nonconvex problem from different starting points. The lower bounds are generated by solving a valid convex relaxation of the original problem.

The convex relaxation, or underestimator, is generated by replacing each nonconvex term in the objective function and the constraints by its convex envelope. There exist different methods of generating this convex envelope depending on the type of term involved. For a univariate concave term, $f(x)$, the tightest convex relaxation, $\mathcal{L}(x)$, is a linearization from x^L to x^U.

$$\mathcal{L}(x) \;=\; f(x^L) \;+\; \frac{f(x^U) \;-\; f(x^L)}{x^U \;-\; x^L}\,(x \;-\; x^L) \tag{4}$$

Linear underestimators for bilinear terms are based on the work of Al-Khayyal and Falk (1983); Al-Khayyal (1990); McCormick (1976). In this case, each bilinear term xy is replaced by an auxiliary variable, w, and the following four linear inequality constraints are added Maranas and Floudas (1995).

$$\begin{array}{rcl} x^l y \;+\; y^l x \;-\; x^l y^l \;-\; w & \leq & 0 \\ x^u y \;+\; y^u x \;-\; x^u y^u \;-\; w & \leq & 0 \\ -x^u y \;-\; y^l x \;+\; x^u y^l \;+\; w & \leq & 0 \\ -x^l y \;-\; y^u x \;+\; x^l y^u \;+\; w & \leq & 0 \end{array} \tag{5}$$

This relaxation can easily be extended to trilinear and fractional terms as well as general products of univariate functions (Maranas and Floudas, 1995). Terms of general nonconvex nature are relaxed using an α underestimator developed by Maranas and Floudas (1994b). For a given nonconvex term in several variables $NC(\mathbf{x})$, the underestimator $\mathcal{L}(\mathbf{x})$ would be

$$\mathcal{L}(\mathbf{x}) \;=\; NC(\mathbf{x}) \;+\; \alpha \sum_{i \,\in\, \mathcal{X}} (x_i^U \;-\; x_i)\,(x_i^L \;- x_i) \tag{6}$$

where $\mathcal{X}$ is the set of $\mathbf{x}$ variables participating in the term $NC(\mathbf{x})$. The value of α needs to be large enough to generate a convex function, but not too large as to overly underestimate the function. It is shown that

$$\alpha \;\geq\; \max_{\mathbf{x}^L \,\leq\, \mathbf{x} \,\leq\, \mathbf{x}^U} \left\{ 0, -\frac{1}{2} \min_k \lambda_k(\mathbf{x}) \right\} \tag{7}$$

where $\lambda_k(\mathbf{x})$ are the eigenvalues of the Hessian matrix of $NC(\mathbf{x})$. It is preferable to derive an analytical expression for the value of α using (7) as an equality. This will provide the tightest possible convex underestimation of $NC(\mathbf{x})$. In cases where this is not possible, several methods have been developed which provide valid lower bounds on the eigenvalues of the Hessian matrix (Adjiman and Floudas, 1996; Adjiman et al., 1997a,b).

In formulation (3) the objective function is convex, and therefore does not need to be underestimated. The equality constraints, however, are by definition nonlinear and therefore nonconvex. Since the nonlinear models are general in nature, there is no specific mathematical structure which can be identified. The relaxation will need to be derived for each type of model studied.

Modified αBB for Error-in-Variables

There are three areas which have a major effect on the convergence rate of this global optimization algorithm. These are: (i) the problem formulation, (ii) the initial bounds on the variables, and (iii) the selection of the branching variables. Each of these three areas was extensively studied so as to develop an efficient algorithm tailored specifically to the parameter estimation of nonlinear algebraic models.

Problem Formulation: The problem formulation has a much greater effect on the performance of the algorithm then one would first think. The convergence depends on the generation of tight underestimators, and the tighter the underestimation, the faster the convergence. The forms of the nonconvexities present contribute greatly to the ability to generate tight lower bounding problems. Various methods are employed to produce simpler nonconvex terms than those originally present in the model equations, such as the use of substitution variables and variable transformations. The examples will show just how much of a effect the formulation can have on the performance of the approach.

Variable bounds: The initial variable bounds effect the convergence for the same reasons as the problem formulation. Tighter initial bounds on the variables results in tighter underestimation and faster convergence. The error-in-variables formulation contains two distinct classes of variables, the fitted data set variables ($\hat{\mathbf{z}}_\mu$), and the parameter variables (θ). Each will be treated separately.

Valid bounds on the fitted data variables can be found from statistical considerations, but for most problems these bounds are not tight enough and convergence is slow at best. An iterative method is employed which generates tight variable bounds, but still reasonably insures the the global solution for the full statistical region is found. The method involves the determination of an initial interior local solution, global optimization using those bounds, bound expansions and subsequent global optimizations to insure the determinate of the true global minimum in the full statistically possible region.

Bounds on the parameter values are obtained in a different manner. A serious of minimizations and maximizations are solved to determine the feasible region for each of the parameter variables.

$$\theta_j^{l,*}/\theta_j^{u,*} = \left\{ \begin{array}{ll} \min\limits_{\theta\ \hat{\mathbf{z}}_\mu} / \max\limits_{\theta\ \hat{\mathbf{z}}_\mu} & \theta_j \\ \text{subject to} & \mathcal{L}(\theta, \hat{\mathbf{z}}_\mu) \leq 0 \\ & \hat{\mathbf{z}}_\mu^L \leq \hat{\mathbf{z}}_\mu \leq \hat{\mathbf{z}}_\mu^U \\ & \theta^L \leq \theta \leq \theta^U \end{array} \right. \tag{8}$$

where $\mathcal{L}(\theta, \hat{\mathbf{z}}_\mu)$ is the convex underestimation of the model equations. Since these problems are solved in sequence, $j = 1 \ldots p$, where p is the total number of parameters, those variable bounds solved for first are not as tight as those solved last. Therefore the series is solved repeatedly until little improvement is seen.

Branching Method: Since the parameter variables appear in each set of constraints, branching on them initially and then on the fitted data variables, appears to be the most logical approach. The method which produces this desired sequence looks at the sum of the deviations caused by each variable in all the nonconvex terms of the problem. In general this deviation, for each nonconvex term and for each variable, is defined as:

$$\delta_i^{F_k} = F_k(\mathbf{x}^{sol}) - \mathcal{L}_k(\mathbf{x}^{sol}) \tag{9}$$

where F_k is a function in which the ith variable participates in, $\mathbf{x}^{sol}$ is the value of the vector $\mathbf{x}$ at the solution to the lower bounding problem, $\mathcal{L}_k(\mathbf{x}^{sol})$ is the underestimating function. Detailed description of the forms resulting from the different types of terms can be found in Esposito and Floudas (1997). These deviations are summed for each variable and then weighted by the number of constants that variable participates in. The branching variable index i^* is selected by:

$$i^* = \arg\max_i \left\{ \sum_{k \in F_i} \delta_i^{F_k} \right\} ncon_i \tag{10}$$

where F_i is the set of terms which the ith variable participates in, and $ncon_i$ is the number of constraints in the original problem in which the ith variable appears.

These concepts were implemented in the following proposed approach.

Step 1 Initialization

1. Set the relative ϵ^{rel} or the absolute ϵ^{abs} convergence tolerance for αBB.
2. Set the outer iteration counter $iter^{out} = 1$
3. Set initial bounds on the parameter variables, $[\theta^{l,orig} : \theta^{u,orig}]$, from physical significant, prior knowledge.
4. For fitted data variables, $\hat{\mathbf{z}}_\mu$, which participate in nonconvex terms, set the bounds: $[\mathbf{z}_\mu - \sigma : \mathbf{z}_\mu + \sigma]$, where σ is the vector of standard deviations, and $\mathbf{z}_\mu$ is the observed values of the data variables. Variables not participating in nonconvex terms have there bounds set at: $[\mathbf{z}_\mu - 3\sigma : \mathbf{z}_\mu + 3\sigma]$.

Step 2 Determine initial bounds on the data set variables, $\hat{\mathbf{z}}_\mu$. This is accomplished by solving the full nonconvex problem locally. Three cases are identified:

- *The problem is infeasible* - Expand the upper and lower bounds on all the variables by a given amount, δ, and resolve.
- *Feasible, but does not have an interior solution* - Expand the active bounds by δ and resolve.

- *Has an interior Solution* - Stop solving the problem locally.

Step 3 Tighten the bounds on the parameter variables by using the following method.

1. Solve (8) for each parameter variable.

2. Calculate the total amount the bounds have changed.

$$\Delta_B = \sum_{i=1}^{p} \left(1 - \frac{\theta_i^{new,u} - \theta_i^{new,l}}{\theta_i^{old,u} - \theta_i^{old,l}} \right)$$

where the superscript old refers to the bounds determined at the previous iteration, and new refers to those determined in this iteration.

3. If $\Delta_B \leq 0.10$ then terminate and start the αBB solution. Otherwise return to 2.

Step 4 Solve the problem to global optimality using the αBB algorithm.

1. Initialize the iteration counter $iter^{\alpha \mathrm{BB}} = 1$.

2. Solve the full nonconvex problem locally to generate an upper bound (UB) on the global solution and store the variable values.

3. Solve the full convex relaxation locally to generate a lower bound (LB) on the global solution. Store the objective value and values of all variables and the α's.

4. If $UB - LB \leq \epsilon^{abs}$ or $\frac{UB - LB}{UB} \leq \epsilon^{rel}$, then terminate the αBB solution.

5. If desired, update the bounds on selected variables by solving min/max problems similar to (8) but not limited to the set of parameter variables.

6. Select the variable to branch the domain on.

7. Divide the region by bisecting along the selected branching variable.

8. If desired, update the bounds on selected variables in new each region.

9. Solve the upper and lower bounding problems to local optimality in each newly created region. Update the upper bound if a local solution in one of the new regions is less than the current upper bound. Store the solutions to the lower problems that are less than the current upper bound.

10. Select the region with the lowest lower bounding solution as the new lower bound. Remove this region from the list of stored regions.

11. Increment the iteration counter, $iter^{\alpha \mathrm{BB}}$, and return to 4.

Step 5 Check for global convergence.
If this is the first outer iteration, or the global solution has changed from the last outer iteration, then expand the fitted data variable's, $\hat{\mathbf{z}}_\mu$, bounds by 10%, reset the parameter bounds to the original values, increment the outer iteration counter, $iter^{outer}$, and return to step 2. Otherwise terminate.

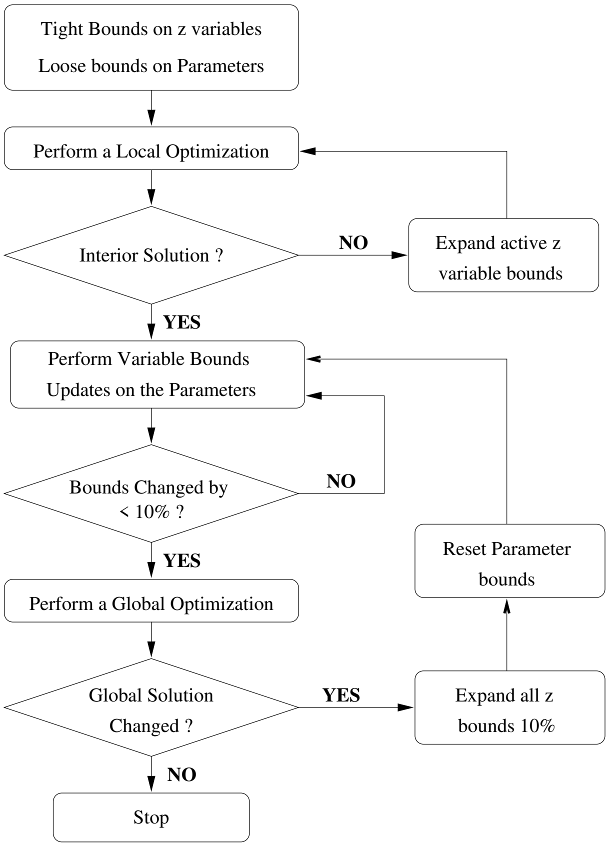

Figure 1: Algorithmic Flowsheet

This algorithm has been implemented in a C program as an extension to the original αBB. The package MINOS 5.4 (Murtagh and Saunders, 1988) is used to perform the needed local optimization.

Computational Studies

In order to illustrate the theoretical and computational aspects of the proposed approach, three examples problems from various literature sources will be considered. Extensive analytical and computational studies are reported in Esposito and Floudas (1997). All problems were solved on a Hewlett Packard C160.

Example 1: Polynomial Fit

This example is taken from Tjoa and Biegler (1992). The model takes the form of a third order polynomial.

$$z_2 = \theta_1 + \theta_2 z_1 + \theta_3 z_1^2 + \theta_4 z_1^3 \qquad (11)$$

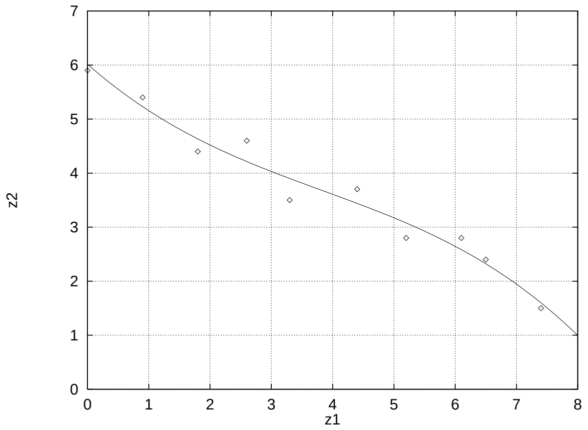

Figure 2: Data and Fitted Model for Example 1

This model results in a number of general nonconvex terms. In order to simplify these terms the following substitutions are introduced: $b_1 = z_1^2$ and $b_2 = z_1^3$. The error associated with each variable is given as equal. The resulting problem formulation is:

$$\min_{\hat{\mathbf{z}}_\mu \; \theta} \sum_{\mu=1}^{10} \sum_{i=1}^{2} (z_{\mu,i} - \hat{z}_{\mu,i})^2 \tag{12}$$

s. t.

$$\begin{array}{rcl} -\hat{z}_{\mu,2} + \theta_1 + \theta_2 \, \hat{z}_{\mu,1} + & & \\ \theta_3 \, b_{\mu,1} + \theta_4 \, b_{\mu,2} & = & 0 \\ -b_{\mu,1} + \hat{z}_{\mu,1}^2 & \leq & 0 \\ b_{\mu,1} - \hat{z}_{\mu,1}^2 & \leq & 0 \\ -b_{\mu,2} + \hat{z}_{\mu,1}^3 & \leq & 0 \\ b_{\mu,2} - \hat{z}_{\mu,1}^3 & \leq & 0 \end{array}$$

This formulation results in two very close local solutions.

Obj	θ_1	θ_2	θ_3	θ_4
0.48515	6.015	-0.9998	0.1525	-0.0132
0.48866	5.963	-0.9521	0.1404	-0.0124

The generation of the convex relaxation involves underestimating the bilinear terms by the added constraints given in (5). In addition, the term $-\hat{z}_{\mu,1}^2$ is treated as a univariate concave term, and the terms $\pm\hat{z}_{\mu,1}^3$ are treated as general nonconvex terms with analytical values of α calculated.

Two different approaches to the generation of bounds on the fitted data variables were used. In both cases, the bounds on $\hat{z}_{\mu,2}$ were held fixed at $z_{\mu,2} \pm 0.35$, since this variable only appears in linear and convex terms. Also branching is performed only on the parameters $\theta_2 \ldots \theta_4$, and bounds updating at each iteration before branching is performed on two variables randomly chosen from the set $\{\theta_2 \ldots \theta_4, \hat{z}_{\mu,1}\}$.

The first approach uses fixed bounds on the variables $\hat{z}_{\mu,1}$. It is shown that even for fairly tight bounds, the algorithm does not converge in 1000 iterations. As a second approach, the full proposed algorithm was used with the following parameters: initial bounds on $\hat{z}_{\mu,1}$ of $z_{\mu,1} \pm 0.08$ and $\delta_1 = 0.01$. The algorithm converged to the global solution: obj = 0.485152, θ = [6.0153, -0.9998, 0.15247, -0.01324], in a total of 604 αBB iterations and 465.4 CPU sec. in 2 outer iterations.

Bounds on $\hat{z}_{\mu,1}$	Iters.	ϵ^{rel}	CPU sec.
$z_{\mu,1} \pm 0.2$	1000	4.3 %	937.0
$z_{\mu,1} \pm 0.16$	1000	1.9 %	872.1
Present Method	604	1.00 %	465.4

If the sequence of parameter bounds problems is solved only once, as opposed to the iterative method presented, before each αBB optimization, then the full solution would require a total of 713 iterations and 527.6 CPU sec (an approximate 15% increase in computational effort).

Example 2: Non-Linear Fit

This example appears in Rod and Hancil (1980) and Tjoa and Biegler (1992). The model equation takes the form of:

$$z_2 = \theta_1 + \frac{1}{z_1 - \theta_2} \tag{13}$$

A data sets of 25 points was generated for this problem using values of θ_1 = 2.00 and θ_2 = 6.00. Random errors from a uniform distribution [-0.025, 0.025] were added to both z_1 and z_2.

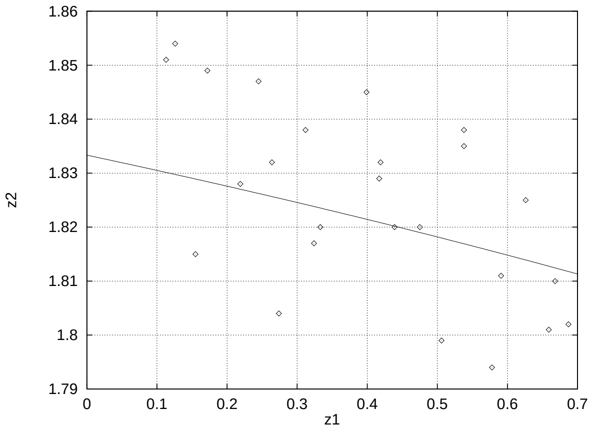

Figure 3: Data and Fitted Model for Example 2

This example illustrates well the effect of problem formulation on the performance of the algorithm since three different formulations can be generated. The first formulation simply has the model equation as a constraint in the optimization and will be referred to as the α formulation.

$$\min_{\hat{\mathbf{z}}_\mu \; \theta} \sum_{\mu=1}^{m} \sum_{i=1}^{2} (\hat{z}_{\mu,i} - z_{\mu,i})^2 \tag{14}$$

s. t.

$$\begin{array}{rcl} -\hat{z}_{\mu,2} + \theta_1 + \frac{1}{\hat{z}_{\mu,1} - \theta_2} & \leq & 0 \\ \hat{z}_{\mu,2} - \theta_1 - \frac{1}{\hat{z}_{\mu,1} - \theta_2} & \leq & 0 \end{array}$$

If the relationship $\hat{z}_{\mu,1}^u < \theta_2^l$ holds (as it does given the data sets generated), then the second constraint is convex. The term in the first constraint is treated as a general nonconvex term with analytical values of α calculated. A second constrained formulation is possible in which all the nonconvexities take the form of bilinear terms.

$$\min_{\hat{\mathbf{z}}_\mu\,\theta} \sum_{\mu=1}^{m} \sum_{i=1}^{2} \left(\hat{z}_{\mu,i} - z_{\mu,i}\right)^2 \tag{15}$$

s. t.

$$-\hat{z}_{\mu,1}\,\hat{z}_{\mu,2} + \hat{z}_{\mu,2}\,\theta_2 + \hat{z}_{\mu,1}\,\theta_1 - \theta_1\,\theta_2 = 1$$

Finally an unconstrained formulation can be generated by substituting the model equation directly into the objective function:

$$\min_{\hat{z}_{\mu,1}\,\theta} \sum_{\mu=1}^{m} \left(z_{\mu,1} - \hat{z}_{\mu,1}\right)^2 + \left(\theta_1 + \frac{1}{\hat{z}_{\mu,1} - \theta_2} - z_{\mu,2}\right)^2 \tag{16}$$

Expanding (16) and grouping the terms results in:

$$\begin{aligned}\min_{\hat{z}_{\mu,1}\,\theta} \quad & -\sum_{\mu=1}^{m} 2\,z_{\mu,2}\,\theta_1 + 2\,z_{\mu,1}\,\hat{z}_{\mu,1} \\ & + \sum_{\mu=1}^{m} z_{\mu,2}^2 + z_{\mu,1}^2 + \hat{z}_{\mu,1}^2 + \theta_1^2 + \frac{1}{(\hat{z}_{\mu,1}-\theta_2)^2} \\ & + \sum_{\mu=1}^{m} \frac{2\,\theta_1 - 2\,z_{\mu,2}}{\hat{z}_{\mu,1} - \theta_2}\end{aligned} \tag{17}$$

In formulation (17) only the last term is nonconvex and treated as a term of general structure. An analytical expression for α was determined, but due to the complexity of this term, it was not possible to determine where the maximum value occurs aprior. Therefore, within the αBB, a valid bound on the value is determined using interval arithmetic (Adjiman and Floudas, 1996; Adjiman et al., 1997a,b). The following table provides the statistics for each formulation (U refers to the upper problem and L the lower).

	Alpha		Bilinear		Unconst.	
	U	L	U	L	U	L
Total Vars.	52	52	52	128	27	27
NC Vars.	26	-	52	-	27	-
NC Terms	25	-	76	-	25	-
Total Consts.	50	50	25	329	-	-
Linear Consts.	0	0	0	329	-	-

In each attempted solution, bounds on the variables were set at: θ_1: [1:10], θ_2: [1:10] and $\hat{z}_{\mu,i}$: $z_{\mu,i} \pm 0.05$. Bounds on the parameters were initially updated using the iterative method of step 3 in the proposed approach, but bounds on the data set variables were held constant.

Form	Iters.	ϵ^{rel}	CPU sec.	sec./iter.
Alpha	21	1%	11.2	0.54
Bilinear	1000	4.65%	540.2	0.54
Unconst.	1000	>100%	282.8	0.28

The α formulation converged to a global solution of $\theta_1 = 2.068$, $\theta_2 = 4.511$ with an objective value of 4.65×10^{-3} quickly, while both the bilinear and unconstrained formulations did not achieve convergence in 1000 iterations. The unconstrained problem was the easiest to solve locally, as illustrated through the time per iteration, but the increased complexity of the nonconvex terms lead to poor underestimators and poor convergence. The bilinear formulation contains very simple nonconvex terms and the underestimation represents the convex hull of each term, but in generating this formulation the number of variables in those terms was doubled compared to the α formulation. This resulted in twice as many variables needing to be branched on to achieve convergence. These results show that not only the form of the nonconvexities matters, but the number of variables participating in them.

Example 3: Pharmacokinetic Data

The model equation takes the form of a sum of three exponential functions:

$$z_2 \;=\; \sum_{j=1}^{3} \theta_{1,j}\;\exp\left[-\theta_{2,j}\,z_1\right] \tag{18}$$

where in this case z_1 is assumed to be without error and σ_2 is proportional to z_2.

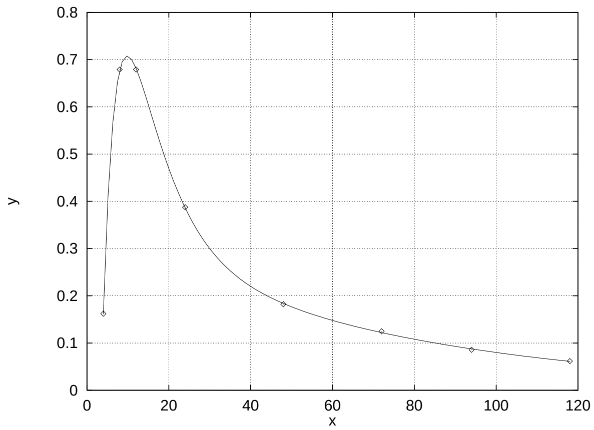

Figure 4: Data and Fitted model for Example 3

This problem has many different possible formulations. For example, the simple inclusion of the model equation as a constraint:

$$\min_{\hat{z}_{\mu,2}\,\theta} \sum_{\mu=1}^{8} \frac{\left(\hat{z}_{\mu,2} - z_{\mu,2}\right)^2}{z_{\mu,2}^2} \tag{19}$$

s. t.

$$\sum_{j=1}^{3} \theta_{1,j}\;\exp\left[-\theta_{2,j}\,z_{\mu,1}\right] \;-\; \hat{z}_{\mu,2} \;\leq\; 0$$

$$-\sum_{j=1}^{3} \theta_{1,j}\;\exp\left[-\theta_{2,j}\,z_{\mu,1}\right] \;+\; \hat{z}_{\mu,2} \;\leq\; 0$$

The exponential terms in the constraints are treated as general nonconvex terms with analytically calculated α values. Using bounds on the parameters variables θ_1 of [-10:10] and θ_2 of [0:0.5], the formulation appears to exhibit at least two local solutions (determined by using a local solver from multiple starting points: (1) Obj = 1.14×10^{-3}, $\theta_{1,j}$ = [0.3554 2.007 -4.572], $\theta_{2,j}$ = [0.01491 0.1102 0.2847]; (2) Obj = 7.86×10^{-2},$\theta_{1,j}$ = [0.5122 0.5122 -5.051], $\theta_{2,j}$ = [0.08011 0.01908 0.5]. The global optimization approach did not converge in

100,000 iterations, in fact the lower bound on the solution never becomes greater than zero, the theoretical lower limit of the objective function. Another formulation involves the use of substitution variables. The substitutions $b_{\mu,j} = \exp\left[-\theta_{2,j}\, z_{\mu,1}\right]$ were introduced and the formulation results in:

$$\min_{\hat{z}_{\mu,2}\,\theta} \sum_{\mu=1}^{8} \frac{\left(\hat{z}_{\mu,2} - z_{\mu,2}\right)^2}{z_{\mu,2}^2} \tag{20}$$

s. t.

$$\sum_{j=1}^{3} \theta_{1,j}\, b_{\mu,j} \;-\; \hat{z}_{\mu,2} \;\leq\; 0$$

$$-\sum_{j=1}^{3} \theta_{1,j}\, b_{\mu,j} \;+\; \hat{z}_{\mu,2} \;\leq\; 0$$

$$-b_{\mu,j} \;+\; \exp\left[-\theta_{2,j}\, z_{\mu,1}\right] \;\leq\; 0$$

$$b_{\mu,j} \;-\; \exp\left[-\theta_{2,j}\, z_{\mu,1}\right] \;\leq\; 0$$

The positive exponential terms are convex while the negative are treated as univariate concave. This formulation is underestimated using the best possible relaxations, but still convergence is not achieved in 100,000 iterations (the lower bound on the solution is once again zero).

We can look at the data provided to get some insight into the problem. From the observations it is clear that at least one $\theta_{1,j}$ must be positive and one negative. We then assume that the other term is also positive. Now that the signs of the terms are known, a logarithmic variable transform can be introduced for $\theta_{1,j}$:

$$\bar{\theta}_{1,1} \;=\; \ln\theta_{1,1} \quad \bar{\theta}_{1,2} \;=\; \ln\theta_{1,2} \quad \bar{\theta}_{1,3} \;=\; -\ln\theta_{1,3}$$

The bounds on these transformed variables are easily calculated from the previous expressions. The model equation now takes the form:

$$z_2 = \sum_{j=1}^{2} \exp\left[\bar{\theta}_{1,j} - \theta_{2,j}\, z_1\right] - \exp\left[\bar{\theta}_{1,3} - \theta_{2,3}\, z_1\right] \tag{21}$$

The substitution $b_{\mu,j} = \bar{\theta}_{1,j} - \theta_{2,j}\, z_{\mu,1}$ is introduced to simplify the exponentials and the resulting formulation is:

$$\min_{\hat{z}_{\mu,2}\,\theta} \sum_{\mu=1}^{8} \frac{\left(\hat{z}_{\mu,2} - z_{\mu,2}\right)^2}{z_{\mu,2}^2} \tag{22}$$

s.t.

$$\sum_{j=1}^{2} e^{b_{\mu,j}} - e^{b_{\mu,3}} - \hat{z}_{\mu,2} \leq 0$$

$$-\sum_{j=1}^{2} e^{b_{\mu,j}} + e^{b_{\mu,3}} + \hat{z}_{\mu,2} \leq 0$$

$$-b_{\mu,j} + \bar{\theta}_{1,j} - \theta_{2,j}\, z_{\mu,1} = 0$$

This formulation contains only convex (positive exponentials), and univariate concave (negative exponentials) terms. Convergence to the global solution (Obj = 1.14×10^{-3}, with 1×10^{-4} absolute convergence) was obtained. This shows that the information contained in the observed values can be invaluable in determining the solution.

Conclusions

The work presented in this paper represents the first global optimization approach tailored to the error-in-variables parameter estimation problem for nonlinear algebraic models. The approach is based upon a deterministic branch and bound global optimization algorithm. Three distinct areas: variable bounds, branching method, and problem formulation, were extensively studied and their effect on the performance of the proposed algorithm was illustrated through several computational studies. The algorithm was shown to converge with reasonable computation effort for problems of various size and complexity.

Acknowledgments

Financial support from the National Science Foundation is gratefully acknowledged. The authors also thank Drs. L. Brüll and U. Pallaske of Bayer AG for providing the third example.

References

Adjiman C.S., Androulakis I.P., and Floudas C.A., 1997b, A global optimization method, αBB, for general twice–differentiable NLPs – II. implementation and computational results. *Computers chem. Engng.*, accepted for publication.

Adjiman C.S., Androulakis I.P., Floudas C.A., and Neumaier A., 1997a, A global optimization method, αBB, for general twice–differentiable NLPs – I. theoretical advances. *Computers chem. Engng.*, accepted for publication.

Adjiman C.S., Androulakis I.P., Maranas C.D., and Floudas C.A., 1996, A global optimization method, αBB, for process design. *Computers chem. Engng, Suppl.* **20**, S419.

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

Al-Khayyal F.A., 1990, Jointly constrained bilinear programs and related problems : An overview. *Computers Math. Applic.* **19**, 53.

Al-Khayyal F.A. and Falk J.E., 1983, Jointly constrained biconvex programming. *Maths Ops Res.* **8**, 273.

Anderson T.F., Abrams D.S., and Grens II E.A., 1978, Evaluation of parameters for nonlinear thermodynamic models. *AIChE J.* **24**, 20.

Androulakis I., Maranas C.D., and Floudas C.A., 1995, αBB : A global optimization method for general constrained nonconvex problems. *Journal of Global Optimization* **7**, 337.

Bard Y., 1974, *Nonlinear Parameter Estimation*. Academic Press.

Bard Y. and Lapidus L., 1968, Kinetics analysis by ditigal parameter estimation. *Catalysis Reviews* **2**, 67.

Dovi V.G. and Paladino O., 1989, Fitting of experimental data to implicit models using a constrained variation method. *Comp. Chem. Engng.* **13**, 731.

Esposito W.R. and Floudas C.A., 1997, Global optimization in parameter estimation of nonlinear algebraic models via the error-in-variables approach. Submitted for Publication.

Fabries J. and Renon H., 1975, Method of evalutation and reduction of vapor-liquid equilbrium data of binary-mixtures. *AIChE J.* **21**, 735.

Fariss R.H. and Law V.H., 1979, An efficient computational technique for the generalized application of maximum likelihood to improve correlation of experimental data. *Comp. Chem. Engng.* **3**, 95.

Kim I., Liebman M.J., and Edgar T.F., 1990, Robust error-in-variables estimation using nonlinear programming techniques. *AIChE J.* **36**, 985.

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

Maranas C.D. and Floudas C.A., 1995, Finding all solutions of nonlinearly constrained systems of equations. *Journal of Global Optimization* **7**, 143.

McCormick G.P., 1976, Computationaly of global solutions to factorable nonconvex programs: Part I - convex underestimations problems. *Mathematical Programming* **10**, 147.

Murtagh B.A. and Saunders M.A., 1988, *MINOS 5.0 User's Guide*. Systems Optimization Laboratory, Dept. of Operations Research, Stanford University, CA.

Patino-Leal H. and Reilly P.M., 1982, Statistical estimation of parameters in vapor-liquid equilbrium. *AIChE J.* **28**, 580.

Ricker N.L., 1984, Comparison of methods for nonlinear parameter estimation. *Ind. Eng. Process Des. Dev.* **23**, 283.

Rod V. and Hancil V., 1980, Iterative estimation of model parameters when measurements of all variables are subject to error. *Comp. Chem. Engng.* **4**, 33.

Schwetlick H. and Tiller V., 1985, Numerical methods for estimating parameters in nonlinear models with errors in the variables. *Technometrics* **27**, 17.

Southwell W.H., 1969, Fitting expermental data. *J. of Comp. Physics* **4**, 465.

Stewart W.E., Caracotsios M., and Sorensen J.P., 1992, Parameter estimation from multiresponse data. *AIChE J.* **38**, 641.

Tjoa T.B. and Biegler L.T., 1992, Reduced successive quadratic programming strategy for errors-in-variables estimation. *Comp. Chem. Engng.* **16**, 523.

Valko P. and Vajda S., 1987, An extended marquardt-type procedure for fitting error-in-variables models. *Comp. Chem. Engng.* **11**, 37.