Global Optimization in Parameter Estimation of Nonlinear Algebraic Models via the Error-In-Variables Approach

William R. Esposito and Christodoulos A. Floudas*
Department of Chemical Engineering
Princeton University
Princeton, N.J. 08544-5263
Tel: (609) 258-4595
Fax: (609) 258-0211
email: floudas@titan.princeton.edu

Abstract

The estimation of parameters in nonlinear algebraic models through the *error-in-variables* method has been widely studied from a computational standpoint. The method involves the minimization of a weighted sum of squared errors subject to the model equations. Due to the nonlinear nature of the models used, the resulting formulation is nonconvex, and may contain several local minima in the region of interest. Current methods tailored for this formulation, although computationally efficient, can only attain convergence to a local solution. In this paper, a global optimization approach based on a branch and bound framework and convexification techniques for general twice differentiable nonlinear optimization problems is proposed for the parameter estimation of nonlinear algebraic models. The proposed convexification techniques exploit the mathematical properties of the formulation. Classical nonlinear estimation problems were solved and will be used to illustrate the various theoretical and computational aspects of the proposed approach.

Keywords: Global Optimization, Parameter Estimation, Maximum Likelihood, Error-in-Variables

*Author to whom all correspondence should be addressed

1 Introduction

Mathematical models which accurately predict physical phenomena are essential in many fields of engineering and the sciences. These models frequently contain adjustable parameters which need to be determined from available experimental data. In the chemical engineering field, mathematical models form the basis for the design, control, and optimization of process systems. In the drive for more economically profitable and environmentally safe processes, the models used have become more complex. Where initially a linear or linearized model was sufficiently accurate for the goal at hand, a nonlinear model is now used to reach a more stringent objective. The use of nonlinear models, many of which are implicit in nature, introduces an added level of complexity into the numerical estimation of the model parameters.

Many different statistical methods exist for the estimation of the parameters in both linear and nonlinear models, as shown by Bard[1]. In particular one method, maximum likelihood, has been presented extensively in the literature. In the maximum likelihood approach, errors in all measured variables, rather than just in a set of dependent variables, are accounted for thus leading to unbiased estimates of the true parameter values. One formulation of this approach which has become popular, assumes that the error is normally distributed with a zero mean and known diagonal covariance matrix. The resulting optimization problem is referred to as the *error-in-variables* formulation. 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 thus leading to many degrees of freedom. 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.

Over the years many different methods have been proposed to address this formulation. Bard and Lapidus[2] provided a description of the early work in the area as it relates to the estimation of kinetic parameters in complex reaction networks. They provided an overview of the various statistical and numerical methods. An emphasis is placed on gradient based approaches which are then used to solve for the kinetic constants in different reaction systems. Southwell[3] presented a method also based on a gradient approach, but through the use of iterative matrix calculations. The method is directly applicable to models which are linear in the parameters. In order to solve models which are nonlinear in the parameters, a Taylor series about the current iteration point is used in the gradient calculations. The models solved are limited to a two variable single equation explicit form, that is, where one variable can be written as a function of another. Schwetlick and Tiller[4] extended this approach to include models which have more than two variables and are highly nonlinear in nature. They accomplished this by exploiting the structure of the Jacobian matrix, and developing a different correction calculation to overcome difficulties in the Gauss-Newton correction for strongly nonlinear problems. This correction is based on a regularization proposed by Marquardt[5] for the least squares problem. Britt and Luecke[6] derived an algorithm based on the use of Lagrange multipliers for the model equations which enter into the problem as equality constraints. Their method involves the use of successive constraint linearization and closely mimics the Gauss-Newton approach for the least squares problem. Fariss and Law[7] presented an approach which modifies the form of the error response to allow for the detection of gross errors in the data. Their numerical method employs gradient based approach, but only requires analytical first derivatives. Fabries and Renon[8] developed a method specifically for thermodynamic models also using a Gauss-Newton approach. They presented a method to not only estimate the parameters in various thermodynamic models, but also to reduce the number of parameters needed through the determination of correlation coefficients. Anderson et al.[9] provided an extension of this work with incorporates two equation models to fully describe the system. They used a successive linearization approach in the determination of the model parameters.

Patino-Leal and Reilly[10] also addressed the vapor-liquid equilibrium problem, but used a very different approach than the previous authors. Through the linearization of the model equations, and integration over the fitted data sets, they formulated an optimization problem which contains only the parameters and linearization points. Both exact and approximate approaches were provided to determine these linearization points. There work can be referred to as a two step approach, where the optimization problem is decomposed into an inner and outer problem, in order to reduce the size of the variable set. 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[11] presented an almost identical approach with the exception that the outer optimization problem is reformulated so that a Gauss-Marquardt method can be directly applied. The inner problem is only a calculation of the fitted data variables using a linearization of the model equations. The approach by Rod and Hancil[12] is quite different. Instead of using a linearization of the model equations in calculating the values of the fitted data variables in the inner problem, an optimization problem is solved.

Another approach to reduce the size of the variable set is via problem decomposition that allows for the optimization to be performed only in the space of the parameters. Dovi and Paladino[13] accomplished this through the use of constrained variation in which the change in one variable can be written as a function of the change in another. This results in a system of equations which can be solved simultaneously along with the optimization problem. Tjoa and Biegler[14] proposed a method based on successive quadratic programming. At each quadratic sub-problem, the optimality conditions in terms of the fitted data variables are solved analytically and included in the optimization.

Reviews and comparisons of the various methods are given by Ricker[15], Stewart et al.[16], and Kim et al.[17].

All the aforementioned approaches address the first of the two difficulties referred to earlier. They only offer convergence to a local solution of a formulation which may have multiple minima. The result could be erroneous parameter estimates and inadequate models. There is currently no method which guarantees convergence to the global minimum and exploits the structure of the mathematical formulation. Up to this point parameter estimation problems have only been addressed as examples for general approaches. Moore et al.[18] and Csendes and Ratz[19] both provide global optimization algorithms based on interval arithmetic. As illustrations of their techniques, parameter estimation problems, in the form of nonlinear, unconstrained least squares problems were solved.

In this work, a deterministic global optimization algorithm based on a branch and bound framework is proposed to address the error-in-variables formulation with no assumptions as to the mathematical structure of the model other than that they are twice continuously differentiable. In section 2, the principle of maximum likelihood will be discussed and the error-in-variables estimation problem will be derived. Section 3 will provide a detailed description of the proposed global optimization approach. First, an overview of a recently developed deterministic algorithm, αBB, will be presented. Secondly, the extensive modifications made to the algorithm will be discussed. Finally, the detailed algorithmic approach will be presented. In Section 4, various nonlinear parameter estimation problems taken from different literature sources will be solved to illustrate the mathematical and computational properties of the proposed approach.

2 Maximum-Likelihood 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 functions.

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

$$\mathbf{z}_\mu \ = \ \zeta_\mu \ + \ \mathbf{e}_\mu \qquad \mu \ = \ 1 \ldots m \tag{2}$$

where ζ_μ is vector of unknown 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.

We now define the likelihood function (L) as the probability of the observed errors in all data points occurring given a set of statistical parameters (ψ) for the distribution used to represent those errors. The errors are functions of the chosen values for the parameters through the model equations

$$L(\theta \ , \psi) \ \equiv \ p \ (\mathbf{E}(\theta) \,|\, \psi) \tag{3}$$

with $\mathbf{E} \equiv [\mathbf{e}_1^T, \mathbf{e}_2^T, \ldots, \mathbf{e}_m^T]^T$. As it is frequently assumed, the measurement errors from different data points (experiments) are considered as uncorrelated or independent. Then, the joint probability density shown in (3) can be expressed as a product of the individual probability densities, and hence the likelihood function becomes:

$$L(\theta,\, \psi) \;=\; \prod_{\mu=1}^{m} p\;(\mathbf{e}_\mu(\theta)\,|\,\psi_\mu) \tag{4}$$

The maximum likelihood estimator seeks to determine values for θ and ψ which maximize L. For convenience it is much easier to work with $(\ln L)$. Maximization of $(\ln L)$ is equivalent to the maximization of the original function L. This results in:

$$\ln L(\theta,\, \psi) \;=\; \sum_{\mu=1}^{m} \ln p\;(\mathbf{e}_\mu(\theta)\,|\,\psi_\mu) \tag{5}$$

If we assume that the errors follow a normal distribution with zero mean and covariance matrix $\mathbf{V}$, we have:

$$p(\mathbf{e}|\mathbf{V}) \;=\; N(\mathbf{e}\,,\mathbf{V}) \;=\; \frac{\mathbf{2}\,\pi^{-\mathbf{n/2}}}{\sqrt{|\mathbf{V}|}}\;\exp\left[-\frac{\mathbf{1}}{\mathbf{2}}\;\mathbf{e^T}\;\mathbf{V}^{-\mathbf{1}}\;\mathbf{e}\right] \tag{6}$$

Substituting (6) into (5), we obtain:

$$\ln L \;=\; -\frac{n\,m}{2}\,\ln 2\,\pi \;-\; \frac{1}{2}\sum_{\mu=1}^{m} \ln|\mathbf{V}_\mu| \;-\; \frac{1}{2}\sum_{\mu=1}^{m} e_\mu^T\,\mathbf{V}_\mu^{-1}\,e_\mu \tag{7}$$

Different forms of (7) exist depending on what is known, or assumed about the covariance matrix. In the sequel, we outline three common cases:

case 1: If the covariance matrix is known completely for each experiment, then only the final term is variable. Therefore maximization of $\ln L$ is equivalent to:

$$\min\;\sum_{\mu=1}^{m} \mathbf{e}_\mu^T\,\mathbf{V}_\mu^{-1}\,\mathbf{e}_\mu \tag{8}$$

case 2: If the covariance matrix is assumed to be known and to be the same in each experiment, that is, $\mathbf{V}_1 \;=\; \mathbf{V}_2 \;=\; \cdots \;= \mathbf{V}$, then (8) reduces to:

$$\min\;\sum_{\mu=1}^{m} \mathbf{e}_\mu^T\,\mathbf{V}^{-1}\,\mathbf{e}_\mu \tag{9}$$

case 3: If it is assumed that the errors in each experiment are known, equal, and independent, that is, the covariance matrix is diagonal with elements v_i, then (9) can be written as:

$$\min \sum_{\mu=1}^{m} \sum_{i=1}^{n} e_{\mu,i}^{2}\, v_i^{-1} \tag{10}$$

where $e_{\mu,i}$ is the ith component on the vector $\mathbf{e}_\mu$, that is, the error associated with the ith variable in the μth experiment. This is the particular form of the maximum likelihood estimator which will be studied in this paper. Defining v_i as σ_i^2 where σ_i is the standard deviation assigned to the ith variable in all experiments, substituting the definition of $e_{\mu,i}$ from (2), and subjecting the minimization to the model equations, we have the following optimization problem, known as the error-in-variables formulation:

$$\min_{\hat{\mathbf{z}}_\mu\, \theta} \sum_{\mu=1}^{r} \sum_{i=1}^{m} \frac{(\hat{z}_{\mu,i} - z_{\mu,i})^2}{\sigma_i^2} \tag{11}$$

s.t.

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

The optimization variables, $\hat{\mathbf{z}}_\mu$, in (11) are referred to as the fitted data variables. These are approximations, obtained through the optimization, of the true values of the experimental data, ζ_μ. This formulation has the following mathematical properties:

1. For a known and constant experimental variance 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 models linear in the parameters, such as $z_2 = a\, z_1 + b$, are nonlinear due to the bilinear term $a\, z_1$.

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

3 Global Optimization Approach

In order to solve (11) to global optimality, a recently developed deterministic branch and bound algorithm, αBB[20;21;22;23], was extensively modified and applied to the parameter estimation of nonlinear algebraic models. In section 3.1, the basic ideas of the αBB algorithm will be presented as applicable to more general problems. In section 3.2, the modifications made to tailor the algorithm to the error-in-variables method will be discussed. Finally, in section 3.3 the detailed algorithmic approach will be presented.

3.1 Basic Ideas of the αBB

The αBB global optimization method guarantees convergence to the global minimum for general twice continuously differentiable constrained and unconstrained 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 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

$$\mathcal{L}(x) \;=\; f(x^L) \;+\; \frac{f(x^U) \;-\; f(x^L)}{x^U \;-\; x^L}\,(x \;-\; x^L) \tag{12}$$

Linear underestimators for bilinear terms are based on the work of Al-Khayyal and Falk[24], Al-Khayyal[25], and McCormick[26]. In this case, each bilinear term xy is replaced by an auxiliary variable, w, and the following four linear inequality constraints are added

$$\begin{aligned}
x^l y + y^l x - x^l y^l - w &\leq 0 \\
x^u y + y^u x - x^u y^u - w &\leq 0 \\
-x^u y - y^l x + x^u y^l + w &\leq 0 \\
-x^l y - y^u x + x^l y^u + w &\leq 0
\end{aligned} \tag{13}$$

The first two linear cuts represent the convex hull (underestimator) of the term, while the last two represent the concave hull (overestimator). The original approach presented in Maranas and Floudas[27] only included the underestimator. By also including the overestimator, the feasible region is greatly reduced which results in a tighter convex relaxation. This relaxation can easily be extended to trilinear terms, xyz, and general products of univariate terms, $f(x)g(y)$,[22;27].

Terms of a general nonconvex nature are relaxed using an α underestimator developed by Maranas and Floudas[28]. For a given nonconvex term in several variables, $NC(\mathbf{x})$, the underestimator, $\mathcal{L}(\mathbf{x})$, would be

$$\mathcal{L}(\mathbf{x}) \;=\; NC(\mathbf{x}) \;+\; \alpha \sum_{i \,\in\, \mathcal{X}} (x_i^U \;-\; x_i)\,(x_i^L \;-x_i) \tag{14}$$

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

$$\begin{aligned} \alpha \;\geq\; \quad & \max\left\{0,\, -\frac{1}{2}\,\min_k\; \lambda_k(\mathbf{x})\right\} \\ & \text{s.t.} \quad \mathbf{x}^L \;\leq\; \mathbf{x} \;\leq\; \mathbf{x}^U \end{aligned} \tag{15}$$

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 (15) as an equality. This will provide the tightest possible convex underestimation of $NC(\mathbf{x})$. As an example, consider the term $x\ \exp(y)$ which appears in the formulation of a multi-product batch design under uncertainty problem presented by Harding and Floudas[29]. The Hessian matrix of the term is written as:

$$\begin{bmatrix} 0 & \exp y \\ \exp y & x\ \exp y \end{bmatrix} \tag{16}$$

The eigenvalues of (16) are:

$$\lambda_{1,2} = \left(\frac{x \;\pm\; \sqrt{x^2 \;+\; 4}}{2}\right) \exp y \tag{17}$$

For all values of x and y, λ_2 is the minimum eigenvalue. The minimum value of λ_2 over the range $x^l \;\leq\; x \;\leq\; x^u$ and $y^l \;\leq\; y \;\leq\; y^u$ is found to occur at x^l, y^u. Therefore the value of α used to underestimate this term would be:

$$\alpha \;=\; -\frac{1}{2}\left(\frac{x^l \;-\; \sqrt{(x^l)^2 \;+\; 4}}{2}\right) \exp y^u \tag{18}$$

In cases where this type of analysis is not possible, several methods have been developed which provide valid lower bounds on the eigenvalues of the Hessian matrix[30;22;23].

In formulation (11) 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.

3.2 Important Modifications of αBB

There are three main areas which have a major effect on the convergence rate of the global optimization algorithm. These are: (i) the problem formulation, (ii) the initial bounds on the variables, and (iii) the selection of the branching variable. 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 formulation of the problem has a major effect on the quality of the convex relaxation. The goal is to produce a formulation in which the nonconvexities are in as simple form as possible which will allow for the analytically determination of tight convex relaxations. This is accomplished through the use of variable substitutions and model transformations to allow for the use of as many special terms as possible. Another consideration is the number of variables which participate in the nonconvex terms. The fewer variables that are involved in the relaxation, the fewer variables which need to be branched on to achieve convergence.

As mentioned previously, an unconstrained formulation is possible in most cases. Local optimization methods are able to solve these formulations with reasonable computational effort. However, in the framework of this global optimization method these formulations are more difficult to solve. By substituting the nonlinear model equations into an already quadratic objective function, the terms generated are much more complex and involve many variables. This leads to difficulties in obtaining tight convex relaxations for these functions, resulting in very poor performance of the global optimization algorithm.

The effects of various formulations on the convergence of the algorithm is shown extensively through the examples presented in section 4.

Variable bounds: Since the variable bounds appear explicitly in the formulation of the convex underestimators, tighter initial bounds produce tighter initial underestimators, which in turn lead to faster convergence of the algorithm. The object is to generate tight enough initial bounds as to obtain convergence to the global solution in a reasonable amount of time, but not to exclude a physically significant global minimum for the problem. This formulation contains two distinct classes of variables, the fitted data variables, $\hat{\mathbf{z}}_\mu$, and the parameter variables, θ. Since each type of variable has a different effect on the problem, alternative methods are used to generate the initial bounds.

Data set variable bounds: Valid bounds on the fitted data variables can be found from statistical considerations. For a known experimental standard deviation, σ, there is a 99.7% probability that the true value will be within $\pm 3\sigma$ of the observations. Frequently variables also have absolute values dictated be physical limitations, for example 0 to 1 for mole fractions. For some problems these bounds are reasonably tight, but for the majority, they exhibit poor convergence characteristics. Therefore a method was developed to generate the tightest possible bounds, while still reasonably ensuring that the global minimum is obtained.

The first step is to identify a set of bounds on $\hat{\mathbf{z}}_\mu$ which provide an interior local solution ($\hat{\mathbf{z}}_\mu^L < \hat{\mathbf{z}}_\mu^{sol} < \hat{\mathbf{z}}_\mu^U$) and then solve the problem to global optimality using these bounds. This does not ensure that the global solution for the full statistically possible region is identified. Therefore, after the problem is solved, the bounds on the fitted data variables are expanded and the problem is re-solved to global optimality using knowledge obtained from the previous global solution. In the limit as the bounds are expanded to the full statistical region, the absolute global solution will be identified. This method only needs to be applied to those data set variables which participate in nonconvex terms. The other variables are given bounds corresponding to the full statistical region.

Parameter variable bounds: Bounds on the parameter variables are obtained very differently. Loose initial bounds are set, for instance, from physical considerations, and prior knowledge. Upon determination of the bounds on the data variables at each outer iteration, the bounds on the parameters are then tightened by solving a series of minimization and maximization problems of the following form:

$$\theta_j^{l,*}/\theta_j^{u,*} = \begin{cases} \min\limits_{\theta\ \hat{\mathbf{z}}_\mu} / \max\limits_{\theta\ \hat{\mathbf{z}}_\mu} & \theta_j \\ \text{subject to} & \mathcal{L}(\theta, \hat{\mathbf{z}}_\mu) \leq 0 \\ & \hat{\mathbf{z}}_\mu^L \leq \hat{\mathbf{z}}_\mu \leq \hat{\mathbf{z}}_\mu^U \\ & \theta^L \leq \theta \leq \theta^U \end{cases} \tag{19}$$

where $\mathcal{L}(\theta, \hat{\mathbf{z}}_\mu)$ is the convex underestimation of the model equations.

These problems are solved in sequence, $j = 1 \ldots p$, where p is the total number of parameters. Since the bounds derived from (19) are dependent on the bounds of the other variables, those solved for first are not as tight as those determined last. Due to this, the sequence is solved repeatedly until the bounds on all the variables do not change significantly.

Branching Method: By looking at the mathematical properties of the formulation, insight into the best branching strategy can be obtained. In the formulation, the parameter variables appear in every set of constraints, while the fitted data variables appear only in one. Therefore by branching initially on the parameters, the underestimation of the problem is improved the most. At some point though, the fitted data variables will need to be branched on to achieve final convergence. The proposed approach which accomplishes this most efficiently, looks at the sum of the deviations caused by each variable in all the nonconvex terms of the problem. These deviations are defined as follows: For a univariate concave term, $f_k(x_i)$, the deviation, $\delta_i^{UT_k}$, of the i variable is defined as:

$$\delta_i^{UT_k} \; = \; f_k(x_i^{sol}) \; - \; \left\{ f_k(x_i^l) \; + \; \frac{f_k(x_i^u) \; - \; f_k(x_i^l)}{x_i^u \; - \; x_i^l} \left(x_i^{sol} \; - \; x^l\right) \right\} \tag{20}$$

where x_i^{sol} represents the value of the ith variable at the solution to the lower bounding problem in the current region. For the kth bilinear term, $x_i x_j$, the deviation for the ith variable, $\delta_i^{BT_k}$, includes a weighting by a region reduction measure, and is defined as:

$$\delta_i^{BT_k} \; = \; \left| w_k^{sol} \; - \; x_i^{sol} \, x_j^{sol} \right| \left(\frac{x_i^u \; - \; x_i^l}{x_i^{u,orig} \; - \; x_i^{l,orig}} \right) \tag{21}$$

where w_k^{sol} is the value of the kth auxiliary variable at the solution to the lower problem, and $x_i^{l,orig}$ and $x_i^{u,orig}$ are the original lower and upper bounds on the variable x_i. For a general term, T_k, the deviation, $\delta_i^{T_k}$, of the i variable is defined as:

$$\delta_i^{T_k} \; = \; T_k(\mathbf{x}_T^{sol}) \; - \; \mathcal{L}(\mathbf{x}_T^{sol}) \tag{22}$$

where $\mathbf{x}_T^{sol}$ is the value of the vector of participating variables at the solution to the lower bounding problem, $T_k(\mathbf{x}_T^{sol})$ is the value of the original nonconvex term at this solution, and $\mathcal{L}(\mathbf{x}_T^{sol})$ is the value of the underestimating function at the solution. The total deviation, δ_i, for variable i is defined as:

$$\delta_i \; = \; \left\{ \sum_{k \, \in \, T_i} \delta_i^{T_k} \; + \; \sum_{k \, \in \, B_i} \delta_i^{BT_k} \; + \; \sum_{k \, \in \, U_i} \delta_i^{UT_k} \right\} ncon_i \tag{23}$$

where B_i, U_i, and T_i are the sets of bilinear, univariate concave, and general terms, that the ith variable participates in, and $ncon_i$ is the number of constraints in the original problem in which the ith variable appears. The branching variable index i^* is selected by:

$$i^* \; = \; \arg\max_i \; \delta_i \tag{24}$$

Other possible branching methods are presented in Adjiman et al.[22;23].

3.3 Proposed Global Optimization Algorithm

The proposed global optimization approach for the parameter estimation of nonlinear algebraic models consists of the following steps:

Step 1 Initialization

1. Formulate the problem as a constrained minimization using the guidelines discussed in section 3.2
2. Set the relative ϵ^{rel} or the absolute ϵ^{abs} convergence tolerance for αBB.
3. Set the outer iteration counter $iter^{out} = 1$
4. Set initial bounds on the parameter variables from physical significant, prior knowledge,

$$\theta^{l,orig} \leq \theta \leq \theta^{u,orig}$$

For fitted data variables which participate in nonconvex terms, set the bounds:

$$\mathbf{z}_\mu - \sigma \leq \hat{\mathbf{z}}_\mu \leq \mathbf{z}_\mu + \sigma$$

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

$$\mathbf{z}_\mu - 3\,\sigma \leq \hat{\mathbf{z}}_\mu \leq \mathbf{z}_\mu + 3\,\sigma$$

Step 2 Determine initial bounds on the data set variables, $\hat{\mathbf{z}}_\mu$. This is accomplished by solving the full non-convex 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 re-solve the problem

$$\begin{aligned} \hat{\mathbf{z}}^u_\mu &= \hat{\mathbf{z}}^u_\mu + \delta \\ \hat{\mathbf{z}}^l_\mu &= \hat{\mathbf{z}}^l_\mu - \delta \end{aligned}$$

- *Feasible, but does not have an interior solution* - Expand the active bounds by δ and re-solve.

$$\begin{aligned} \text{if } \hat{z}^{sol}_{\mu,i} &= \hat{z}^u_{\mu,i} \text{ then } \hat{z}^u_{\mu,i} = \hat{z}^u_{\mu,i} + \delta_i \\ \text{if } \hat{z}^{sol}_{\mu,i} &= \hat{z}^l_{\mu,i} \text{ then } \hat{z}^l_{\mu,i} = \hat{z}^l_{\mu,i} - \delta_i \end{aligned}$$

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

Step 3 Tighten the bounds on the parameter variables.

1. Solve (19) 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, set $\theta^{new,l} \rightarrow \theta^{old,l}$; $\theta^{new,u} \rightarrow \theta^{old,u}$ and 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.

$$obj^* \rightarrow UB \ , \quad \hat{\mathbf{z}}_\mu^* \rightarrow \hat{\mathbf{z}}_\mu^{UB} \ , \quad \theta^* \rightarrow \theta^{UB}$$

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.

$$obj^* \rightarrow obj^{1,1} \ , \quad \hat{\mathbf{z}}_\mu^* \rightarrow \hat{\mathbf{z}}_\mu^{1,1} \ , \quad \theta^* \rightarrow \theta^{1,1} \ , \quad \mathbf{w}^* \rightarrow \mathbf{w}^{1,1} \ , \quad \alpha \rightarrow \alpha^{1,1}$$

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 the following min/max problems

$$x_j^{l,*} / x_j^{u,*} = \begin{cases} \min\limits_{\theta \ \hat{\mathbf{z}}_\mu} / \max\limits_{\theta \ \hat{\mathbf{z}}_\mu} & x \\ \text{subject to} & \mathcal{L}(\theta, \hat{\mathbf{z}}_\mu) \leq 0 \\ & \hat{\mathbf{z}}_\mu^L \leq \hat{\mathbf{z}}_\mu \leq \hat{\mathbf{z}}_\mu^U \\ & \theta^L \leq \theta \leq \theta^U \end{cases} \tag{25}$$

where $\mathbf{x}$ represents the vector $[\hat{\mathbf{z}}_\mu \ \theta]$ and $j \in J$ where J is the set of variables selected for updating.

6. Select the variable to branch the domain on using the expressions developed in section 3.2.

7. Divide the region by bisecting along the selected branching variable.
8. If desired, update the bounds on selected variables in new each region by solving (25).
9. Solve the upper and lower bounding problems to local optimality in each newly created region ($r = 1, 2$). Update the upper bound if a local solution in one of the new regions is less than the current upper bound:

   if $obj^{upper} < UB$ then

$$obj^{upper} \rightarrow UB \ , \quad \hat{\mathbf{z}}_{\mu}^{*} \rightarrow \hat{\mathbf{z}}_{\mu}^{UB} \ , \quad \theta^{*} \rightarrow \theta^{UB}$$

   Store the solutions to the lower problems that are less than the current upper bound:

   if $obj^{lower} < UB$ then

$$obj^{lower} \rightarrow obj^{iter,r} \ , \quad \hat{\mathbf{z}}_{\mu}^{*} \rightarrow \hat{\mathbf{z}}_{\mu}^{iter,r} \ , \quad \theta^{*} \rightarrow \theta^{iter,r} \ , \quad \mathbf{w}^{*} \rightarrow \mathbf{w}^{iter,r} \ , \quad \alpha \rightarrow \alpha^{iter,r}$$

10. Select the region with the lowest lower bounding solution. This becomes the new lower bound:

$$LB \ = \ \min_{iter \ r} obj^{iter,r}$$

    Remove this region from the list of stored regions.
11. Increment the iteration counter $iter^{\alpha \mathrm{BB}} = iter^{\alpha \mathrm{BB}} + 1$ and return to 4.

Step 5 Check for global convergence:

if $iter^{out} = 1$ or $sol^{iter^{out}} \neq sol^{iter^{out}-1}$ then

$$\hat{\mathbf{z}}_{\mu}^{u} = \hat{\mathbf{z}}_{\mu}^{u} + \gamma \left(\hat{\mathbf{z}}_{\mu}^{u} - \hat{\mathbf{z}}_{\mu}^{l}\right) \ , \quad \hat{\mathbf{z}}_{\mu}^{l} = \hat{\mathbf{z}}_{\mu}^{l} - \gamma \left(\hat{\mathbf{z}}_{\mu}^{u} - \hat{\mathbf{z}}_{\mu}^{l}\right)$$

$$\theta^{u} = \theta^{u,orig} \ , \quad \theta^{l} = \theta^{l,orig}$$

$$iter^{out} = iter^{out} + 1 \quad \text{and return to step 2}$$

where $sol^{iter^{out}}$ is the global solution determined by αBB in the $iter^{out}$ outer iteration. Otherwise terminate.

The algorithm is presented in a flowsheet form in Figure 1. It should be noted that the convergence criterion given in step 5 guarantees convergence to the global minimum in the limit as the parameter γ is increased such that the bounds are expanded to the point in which the full statistical region is enclosed. For practical purposes a value of 0.1 was used in all the examples and has resulted in convergence to the global minimum in every case tested, even in those which the first outer iteration resulted in a local minimum. This algorithm has been implemented in a C program as an extension to the original αBB, where MINOS 5.4[31] is used to perform the needed local optimizations.

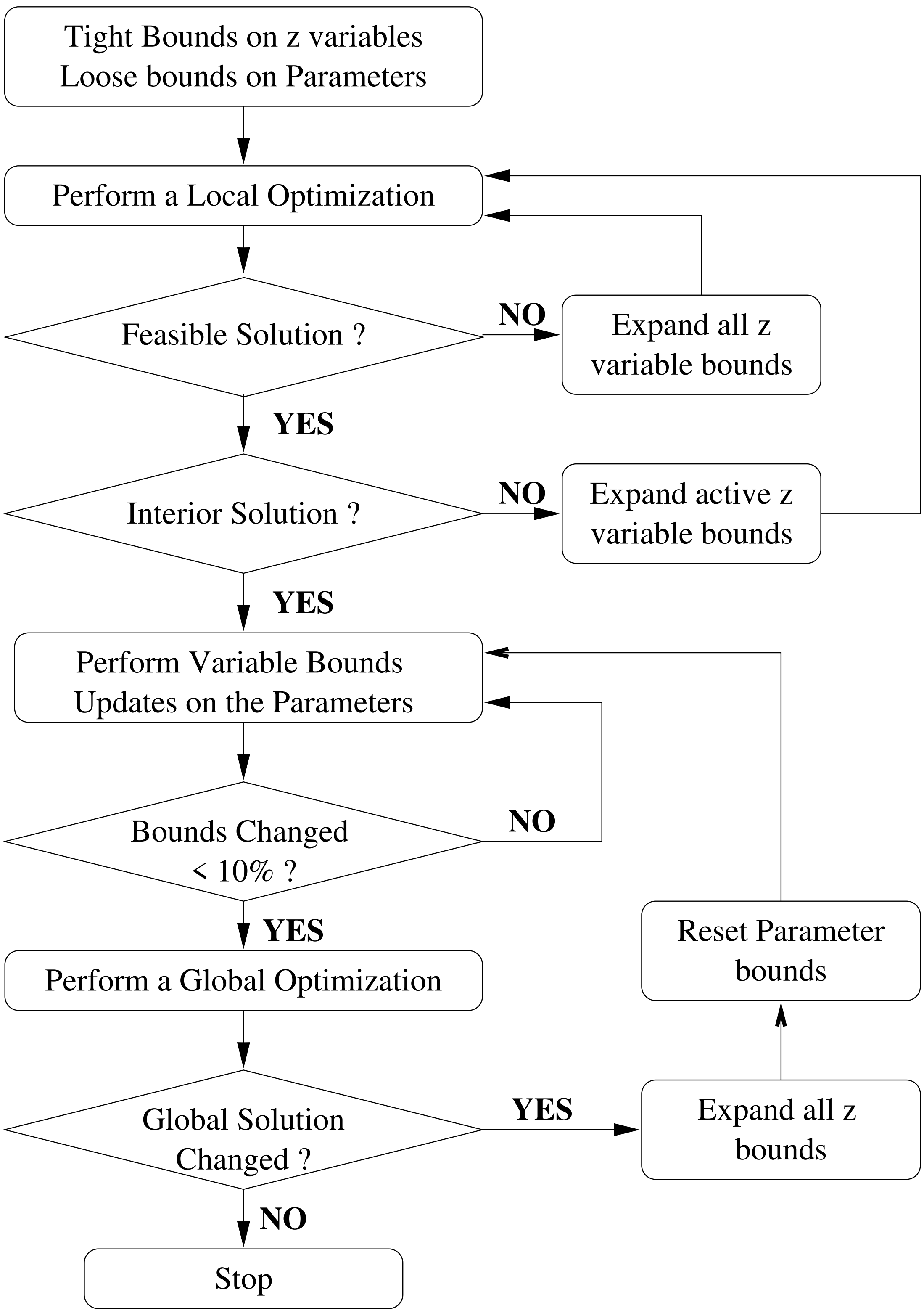

Figure 1: Flowsheet representation of proposed global optimization approach

4 Computational Studies

In order to illustrate the theoretical and computational aspects of the proposed approach, seven examples problems from various literature sources will be considered. Through these examples the important issues of (i) problem formulation; (ii) types of underestimators; (iii) unconstrained vs. constrained optimization; (iv) tightness of different underestimators; and (v) computational performance and convergence are illustrated. All problems were solved on a Hewlett Packard C160.

Example 1: Linear Fit

This example is taken from Tjoa and Biegler[14] and represents the fitting of data to a straight line of the form:

$$z_2 \;=\; \theta_1 \;+\; \theta_2\, z_1 \tag{26}$$

Since error is assumed in both z_2 and z_1, the minimization is performed over these variables in addition to the parameter set. Therefore the model is nonlinear and nonconvex due to the bilinear term, $\theta_2\, z_1$. In this example both variables have equal variance, that is, $\sigma_1^2 \;=\; \sigma_2^2$, resulting in an unweighted objective function in a constrained formulation:

$$\min_{\hat{\mathbf{z}}_\mu\, \theta} \; \sum_{\mu=1}^{10} \sum_{i=1}^{2} \left(z_{\mu,i} - \hat{z}_{\mu,i}\right)^2 \tag{27}$$

s. t.

$$-\hat{z}_{\mu,2} \;+\; \theta_1 \;+\; \theta_2\, \hat{z}_{\mu,1} = \; 0$$

The convex relaxation is generated by underestimating the bilinear term in the constraint using the added constraints given by (13). The lower bounding problem takes the form:

$$\min_{\hat{\mathbf{z}}_\mu\, \theta} \; \sum_{\mu=1}^{10} \sum_{i=1}^{2} \left(z_{\mu,i} - \hat{z}_{\mu,i}\right)^2 \tag{28}$$

s. t.

$$\begin{aligned}
-\hat{z}_{\mu,2} \;+\; \theta_1 \;+\; w_{\mu,1} \;&=\; 0\\
\hat{z}^l_{\mu,1}\, \theta_2 \;+\; \theta^l_2\, \hat{z}_{\mu,1} \;-\; \hat{z}^l_{\mu,1}\, \theta^l_2 \;-\; w_{\mu,1} \;&\leq\; 0\\
\hat{z}^u_{\mu,1}\, \theta_2 \;+\; \theta^u_2\, \hat{z}_{\mu,1} \;-\; \hat{z}^u_{\mu,1}\, \theta^u_2 \;-\; w_{\mu,1} \;&\leq\; 0\\
-\hat{z}^u_{\mu,1}\, \theta_2 \;-\; \theta^l_2\, \hat{z}_{\mu,1} \;+\; \hat{z}^u_{\mu,1}\, \theta^l_2 \;+\; w_{\mu,1} \;&\leq\; 0\\
-\hat{z}^l_{\mu,1}\, \theta_2 \;-\; \theta^u_2\, \hat{z}_{\mu,1} \;+\; \hat{z}^l_{\mu,1}\, \theta^u_2 \;+\; w_{\mu,1} \;&\leq\; 0
\end{aligned}$$

Table 1 provides the formulation statistics.

	Upper Problem	Lower Problem
Total number of variables	12	22
nonconvex variables	11	-
nonconvex terms	10	-
Total number of constraints	10	50
linear constraints	0	50

Table 1: Formulation Statistics for Example 1.

The initial bounds on the parameters were set at: θ_1: [0,10] and θ_2: [-2,2]. Two different approaches were used to generate bounds on the fitted data variables. The first involved simply fixing the bounds at $\mathbf{z}_\mu \pm 0.5$ without performing any updating. The algorithm converged in 8 iterations and 0.781 CPU sec (a relative tolerance, ϵ^{rel}, of 1% was used). The resulting solution had an objective value of 0.61857 and parameter values of $\theta_1 = 5.7084$ and $\theta_2 =$ -0.54556.

The second approach used the full algorithm presented in section 3.3. Since z_2 does not participate in any nonconvex terms, its bounds were held fixed at $z_{\mu,2} \pm 0.5$. Initial bounds on z_1 were set at $z_{\mu,1} \pm 0.08$, with a delta, δ_1, of 0.01. The algorithm converged in a total of 4 αBB iterations (in 2 outer solutions), with a CPU time of 0.641 sec. The solution point was the same as in the first approach.

In this case, due to the relatively small size of the problem and few nonconvex terms, the fitted data variable bounds updating method does not offer a significant computational savings over using fixed values for the bounds. The additional computational time required to perform two separate global solutions nearly offsets the decrease in the total number of iterations.

Example 2: Polynomial Fit

This example is also taken from Tjoa and Biegler[14] and uses the same data set as the previous example. The model is now a third order polynomial:

$$z_2 = \theta_1 + \theta_2 z_1 + \theta_3 z_1^2 + \theta_4 z_1^3 \tag{29}$$

This model results in a number of general nonconvex terms. In order to simplify these terms the following substitutions are introduced:

$$b_1 = z_1^2 \tag{30}$$

$$b_2 = z_1^3 \tag{31}$$

Bounds on these substitution variables are generated in each region based on these relationships. As in the first example, the variance associated with each variable is assumed equal. The resulting problem formulation is:

$$\min_{\hat{\mathbf{z}}_\mu\,\theta} \sum_{\mu=1}^{10} \sum_{i=1}^{2} (z_{\mu,i} - \hat{z}_{\mu,i})^2 \tag{32}$$

s. t.

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

This formulation results in two very close local solutions presented in Table 2. These solutions were generated by repeatedly solving the upper problem to local optimality from randomly chosen starting points.

Obj	θ_1	θ_2	θ_3	θ_4
0.485152	6.0153	-0.9998	0.15247	-0.01324
0.488658	5.9634	-0.9521	0.14044	-0.01235

Table 2: Local solutions for Example 2.

The generation of the convex relation is more involved in this case than the first. The bilinear terms are underestimated by the added constraints given in (13). In addition, the term $-\hat{z}^2_{\mu,1}$ is treated as a univariate concave term, and the terms $\pm\hat{z}^3_{\mu,1}$ are treated as general nonconvex terms. The resulting lower problem formulation is:

$$\min_{\hat{\mathbf{z}}_\mu\,\theta} \sum_{\mu=1}^{10} \sum_{i=1}^{2} (z_{\mu,i} - \hat{z}_{\mu,i})^2 \tag{33}$$

s. t.

$$\begin{aligned} -\hat{z}_{\mu,2} + \theta_1 + w_{\mu,1} + w_{\mu,2} + w_{\mu,3} &= 0 \\ -b_{\mu,1} + \hat{z}^2_{\mu,1} &\le 0 \\ b_{\mu,1} - \left\{ \left(\hat{z}^l_{\mu,1}\right)^2 + \frac{\left(\hat{z}^u_{\mu,1}\right)^2 - \left(\hat{z}^l_{\mu,1}\right)^2}{\hat{z}^u_{\mu,1} - \hat{z}^l_{\mu,1}} \left(\hat{z}_{\mu,1} - \hat{z}^l_{\mu,1}\right) \right\} &\le 0 \\ -b_{\mu,2} + \hat{z}^3_{\mu,1} + \alpha_{\mu,1}\left(\hat{z}^u_{\mu,1} - \hat{z}_{\mu,1}\right)\left(\hat{z}^l_{\mu,1} - \hat{z}_{\mu,1}\right) &\le 0 \\ b_{\mu,2} - \hat{z}^3_{\mu,1} + \alpha_{\mu,2}\left(\hat{z}^u_{\mu,1} - \hat{z}_{\mu,1}\right)\left(\hat{z}^l_{\mu,1} - \hat{z}_{\mu,1}\right) &\le 0 \\ w_{\mu,1} &\longrightarrow \theta_1\,\hat{z}_{\mu,1} \\ w_{\mu,2} &\longrightarrow \theta_2\,b_{\mu,1} \\ w_{\mu,3} &\longrightarrow \theta_3\,b_{\mu,2} \end{aligned}$$

The added constraints for the bilinear terms are included in the formulation, but not shown above for simplicity. For the general nonconvex terms, analytical values of α were calculated and can be found in Appendix B. The formulation statistics appear in Table 3.

	Upper Problem	Lower Problem
Total number of variables	44	74
nonconvex variables	13	-
nonconvex terms	60	-
Total number of constraints	50	170
linear constraints	0	140

Table 3: Formulation Statistics for Example 2.

Two different approaches for 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 $\theta_2,\ \theta_3,\ \theta_4$, and bounds updating at each iteration before branching is performed on two variables randomly chosen from the set $\{\theta_2,\ \theta_3,\ \theta_4, \hat{z}_{\mu,1}\}$.

The first approach uses fixed bounds on the variables $\hat{z}_{\mu,1}$. Table 4 shows that even for relatively tight bounds, the algorithm does not converge in 1000 iterations. In the second, the full algorithm is 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. This example involves a much greater number of nonconvex terms, all of which involve the fitted data variable $\hat{z}_{\mu,1}$ either directly or indirectly through the substitution variables. As a result, the bounds on this set of variables have a pronounced effect on the convergence characteristics of the algorithm, that is, the computational effort and the iterations required to solve the problem.

Bounds on $\hat{z}_{\mu,1}$	Iterations	Rel. Convergence	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

Table 4: Convergence Rate for Example 2 using various bounds on $z_{\mu,1}$. (†) initial bounds of $z_{\mu,1}\ \pm\ 0.08$ with $\delta_1 = 0.01$. (‡) Total αBB iterations for two outer iterations (218 iterations, 158.2 cpu sec and 386 iterations, 307.43 cpu sec).

Figure 2 shows how the parameter bounds change at each iteration of the initial updating (step 3 in the proposed method given in section 3.3). These results were obtained from the first outer iteration of the solution presented above. 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 computation time).

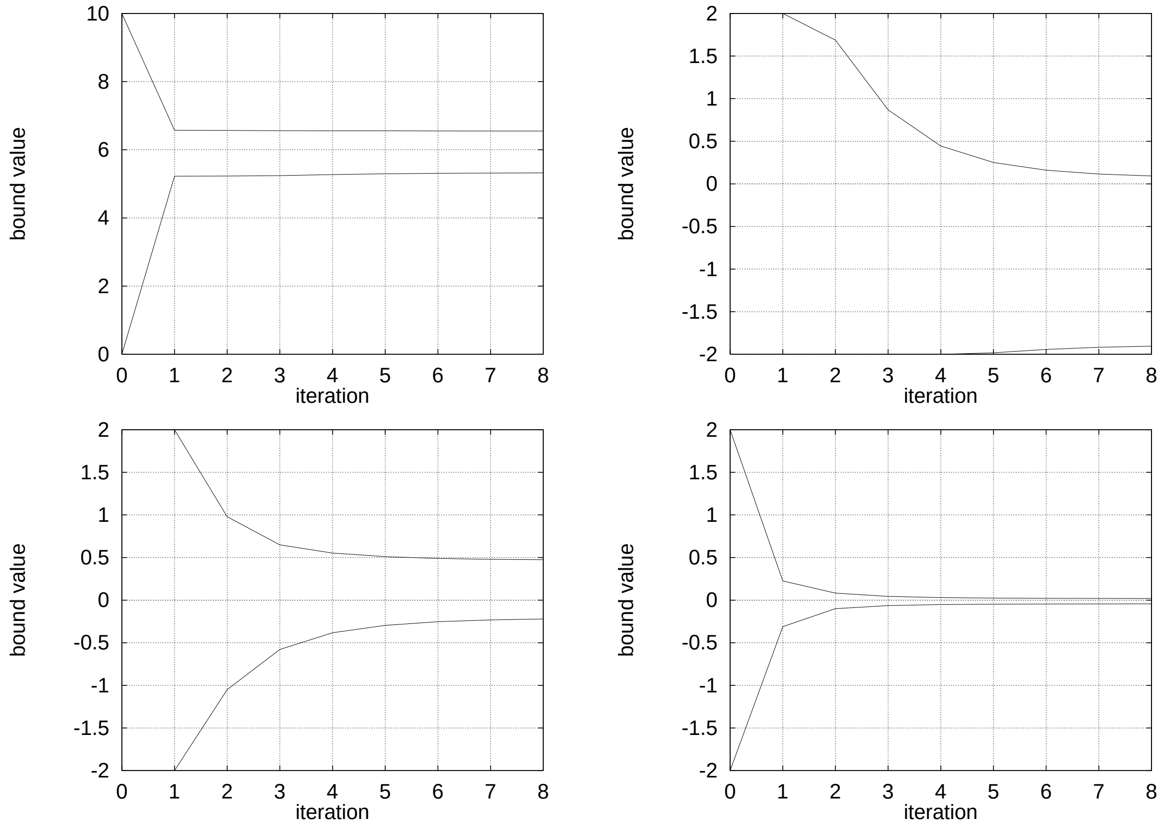

Figure 2: Parameter bounds progression for initial updating for Example 2. (θ_1 to θ_4, top left to bottom right)

Example 3: Non-Linear Fit

This example appears in Tjoa and Biegler[14] and Rod and Hancil[12]. The model equation takes the form of:

$$z_2 = \theta_1 + \frac{1}{z_1 - \theta_2} \tag{34}$$

Data sets of 25, 50, 75, and 100 points were generated for this problem using values of $\theta_1 = 2.00$ and $\theta_2 = 6.00$. The values for z_1 ranged from 0 to 3.0. Random errors from a uniform distribution [-0.025, 0.025] were added to both z_1 and z_2. An example of a data set of 100 points along with the model equation used to generate the set is shown in Figure 3.

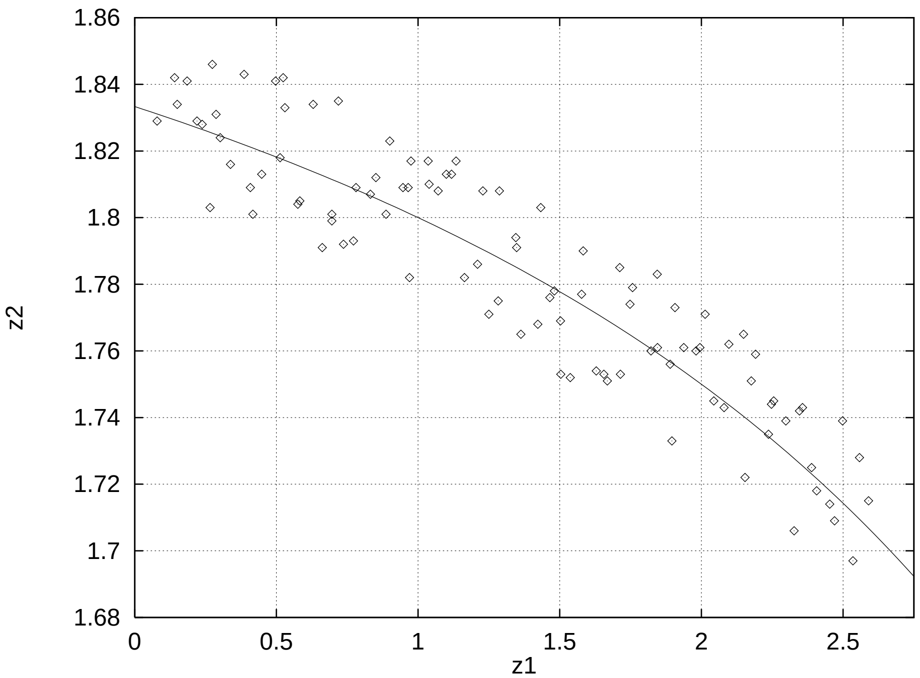

Figure 3: Sample 100 point data set for Example 3 and the true model equation (–)

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:

$$\min_{\hat{\mathbf{z}}_\mu\, \theta} \sum_{\mu=1}^{m} \sum_{i=1}^{2} \left(\hat{z}_{\mu,i} - z_{\mu,i}\right)^2 \tag{35}$$

s. t.

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

If the relationship $\hat{z}^u_{\mu,1} < \theta^l_2$ 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. The resulting convex relaxation is:

$$\min_{\hat{\mathbf{z}}_\mu\, \theta} \sum_{\mu=1}^{m} \sum_{i=1}^{2} \left(\hat{z}_{\mu,i} - z_{\mu,i}\right)^2 \tag{36}$$

s. t.

$$
\begin{aligned}
-\hat{z}_{\mu,2} \;+\; \theta_1 \;+\; \tfrac{1}{\hat{z}_{\mu,1} \;-\; \theta_2} \;+\; \alpha_\mu \Big\{ \left(\hat{z}^u_{\mu,1} \;-\; \hat{z}_{\mu,1}\right) \left(\hat{z}^l_{\mu,1} \;-\; \hat{z}_{\mu,1}\right) & \\
+\; \left(\theta^u_2 \;-\; \theta_2\right) \left(\theta^l_2 \;-\; \theta_2\right) \Big\} & \;\leq\; 0 \\
\hat{z}_{\mu,2} \;-\; \theta_1 \;-\; \tfrac{1}{\hat{z}_{\mu,1} \;-\; \theta_2} & \;\leq\; 0
\end{aligned}
$$

The expression for the value of α_μ can be found in Appendix B. A second constrained formulation is possible in which all the nonconvexities take the form of bilinear terms.

$$
\min_{\hat{\mathbf{z}}_\mu \, \theta} \; \sum_{\mu=1}^{m} \; \sum_{i=1}^{2} \; \left(\hat{z}_{\mu,i} - z_{\mu,i}\right)^2 \tag{37}
$$

s. t.

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

The resulting convex relaxation is:

$$
\min_{\hat{\mathbf{z}}_\mu \, \theta} \; \sum_{\mu=1}^{m} \; \sum_{i=1}^{2} \; \left(\hat{z}_{\mu,i} - z_{\mu,i}\right)^2 \tag{38}
$$

s. t.

$$
\begin{aligned}
& -w_{\mu,1} \;+\; w_{\mu,2} \;+\; w_{\mu,3} \;-\; w_4 \;=\; 1 \\
& w_{\mu,1} \;\rightarrow\; \hat{z}_{\mu,1}\, \hat{z}_{\mu,2} \\
& w_{\mu,2} \;\rightarrow\; \hat{z}_{\mu,2}\, \theta_2 \\
& w_{\mu,3} \;\rightarrow\; \hat{z}_{\mu,1}\, \theta_1 \\
& w_4 \;\rightarrow\; \theta_1\, \theta_2
\end{aligned}
$$

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

$$
\min_{\hat{\mathbf{z}}_\mu \, \theta} \; \sum_{\mu=1}^{m} \; \left(z_{\mu,1} \;-\; \hat{z}_{\mu,1}\right)^2 \;+\; \left(\theta_1 \;+\; \frac{1}{\hat{z}_{\mu,1} \;-\; \theta_2} \;-\; z_{\mu,2}\right)^2 \tag{39}
$$

Expanding (39) and grouping the terms results in:

$$
\begin{aligned}
\min_{\hat{\mathbf{z}}_\mu \, \theta} \;\; & -\sum_{\mu=1}^{m} \; 2\, z_{\mu,2}\, \theta_1 \;+\; 2\, z_{\mu,1}\, \hat{z}_{\mu,1} \\
& +\; \sum_{\mu=1}^{m} \; z^2_{\mu,2} \;+\; z^2_{\mu,1} \;+\; \hat{z}^2_{\mu,1} \;+\; \theta^2_1 \;+\; \tfrac{1}{\left(\hat{z}_{\mu,1} \;-\; \theta_2\right)^2} \\
& +\; \sum_{\mu=1}^{m} \; \tfrac{2\, \theta_1 \;-\; 2\, z_{\mu,2}}{\hat{z}_{\mu,1} \;-\; \theta_2}
\end{aligned} \tag{40}
$$

In formulation (40) only the last term is nonconvex and treated as a term of general structure. The resulting convex relaxation is:

$$\begin{aligned} \min_{\hat{\mathbf{z}}_\mu\, \theta} \quad & -\sum_{\mu=1}^{m} 2\, z_{\mu,2}\, \theta_1 \;+\; 2\, z_{\mu,1}\, \hat{z}_{\mu,1} \\ & + \sum_{\mu=1}^{m} z_{\mu,2}^2 \;+\; z_{\mu,1}^2 \;+\; \hat{z}_{\mu,1}^2 \;+\; \theta_1^2 \;+\; \frac{1}{(\hat{z}_{\mu,1} - \theta_2)^2} \\ & + \sum_{\mu=1}^{m} \frac{2\,\theta_1 - 2\, z_{\mu,2}}{\hat{z}_{\mu,1} - \theta_2} \;+\; \alpha_\mu \Big\{ \left(\hat{z}_{\mu,1}^u - \hat{z}_{\mu,1}\right)\left(\hat{z}_{\mu,1}^l - \hat{z}_{\mu,1}\right) \\ & \qquad + \left(\theta_1^u - \theta_1\right)\left(\theta_1^l - \theta_1\right) + \left(\theta_2^u - \theta_2\right)\left(\theta_2^l - \theta_2\right) \Big\} \end{aligned} \tag{41}$$

Due to the complexity of the α term (see Appendix B), it is not possible to determine where the maximum value occurs apriori. Therefore, within the αBB, a valid bound on α is determined using interval arithmetic methods[30;22;23]. Table 5 contains the statistics for each formulation.

	Alpha		Bilinear		Unconstrained	
	Upper	Lower	Upper	Lower	Upper	Lower
Total Variables	2m+2	2m+2	2m+2	5m+3	2m+2	2m+2
Nonconvex Variables	m+1	-	2m+2	-	m+2	-
Nonconvex Terms	m	-	3m+1	-	m	-
Total Constraints	2m	2m	m	13m+4	-	-
Linear Constraints	0	0	0	13m+4	-	-

Table 5: Formulation Statistics for Example 3. In each case, m refers to the number of data points used in the estimation.

In each attempted solution, bounds on the variables were set at: θ_1: [1:10], θ_2: [1-3.5:10] (depending on the size of the data set due to the discontinuity at $\theta_2 = \hat{z}_{\mu,1}$), 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 section 3.3, but bounds on the data set variables were held constant. Table 6 shows the results using the different formulations for one data set of 25 points. The results show that the α formulation converged quickly, while both the bilinear and unconstrained formulations did not achieve convergence in 1000 iterations. This result is quite counter-intuitive from looking at the problem. First, an unconstrained formulation is much easier to solve from a local optimization standpoint, as is evident from the time per iteration. But in the framework of the αBB the problem is much more difficult due to the increased complexity of the nonconvexities. Analytical expressions for α were determined, but their form still required the use of approximations in calculating the numerical values. This resulted in poor underestimation which is evident in the poor convergence rate. Secondly, the bilinear formulation possess an underestimator which represents the convex hull of each term in the constraint. This should lead to better convergence, but in generating this formulation, the number of variables in nonconvex terms was doubled and the number of constraints was greatly increased. Therefore twice as many variables may need to be branched on in order to achieve convergence. In generating the best possible formulation, not only should the forms of the nonconvex terms be considered, but also the number of terms and size of the variable set which participates in those terms.

Formulation	Iterations	Rel. Convergence	CPU sec	sec/iter
Alpha	21	1%	11.2	0.54
Bilinear	1000	4.65%	540.2	0.54
Unconstrained	1000	>100%	282.8	0.28

Table 6: Convergence of different formulations of Example 3 using a 25 point data set.

The size of the data set has also a distinct effect on the convergence of the algorithm. Table 7 provides the obtained values for the parameters and objective function for each data set size. As the number of data points increases, the number of iterations required to solve the problem decreases, while the total solution time increases. The rate of increase in solution time is a result of the way the local solver, MINOS, scales with the number of variables. Figure 4 shows this result graphically. Also noteworthy is how the resulting fitted parameters approach the true values as the number of data sets is increased. This illustrates well the unbiased estimates produced using maximum likelihood.

Data Points	θ_1	θ_2	Obj. Value
25	2.022 ± 0.048	5.70 ± 1.38	$4.68 \pm 0.24 \times 10^{-3}$
50	2.010 ± 0.013	5.79 ± 0.27	$9.74 \pm 0.16 \times 10^{-3}$
75	2.009 ± 0.007	5.80 ± 0.15	$1.47 \pm 0.13 \times 10^{-2}$
100	1.999 ± 0.003	6.05 ± 0.07	$1.98 \pm 0.11 \times 10^{-2}$

Table 7: Parameter estimates for Example 3 with different sizes of data sets. Results presented are means and standard deviations using five separate, randomly generated data sets of each size.

Example 4: Respiratory Mechanical Model

This model appears in Csendes and Ratz[19]. The model equation takes the form of:

$$z_\mu = \left(\theta_1 + \frac{\theta_2}{\omega_\mu^{\theta_3}}\right) + \jmath\left(\omega_\mu\,\theta_4 - \frac{\theta_5}{\omega_\mu^{\theta_3}}\right) \tag{42}$$

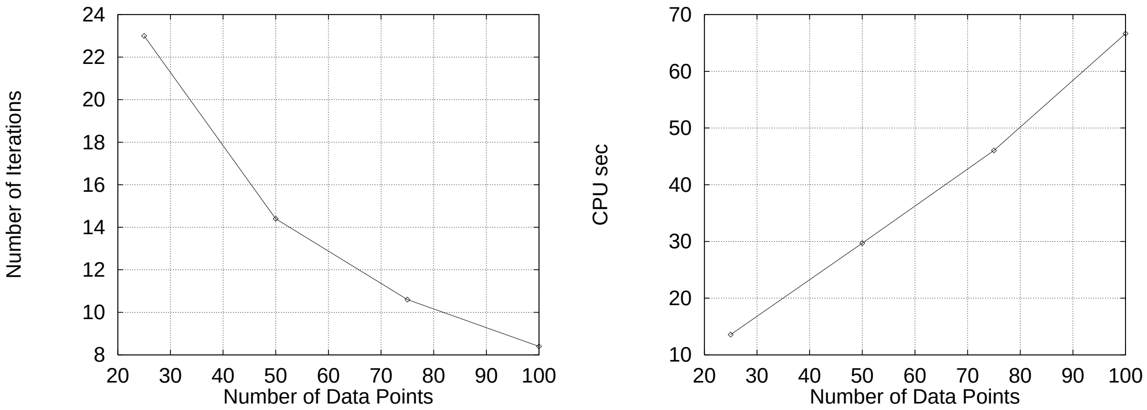

Figure 4: Number of iterations and computational time for Example 3 with different data set sizes. Results are the average using five separate, randomly generated data sets of each size.

where $\jmath$ refers to $\sqrt{-1}$, and $\omega_\mu = \frac{\mu\,\pi}{20}$. In formulating this problem, the real and imaginary parts of $\hat{z}_\mu$ will be handled separately. Let $\hat{z}^r_\mu$ represent the real part and $\hat{z}^i_\mu$ the imaginary. For this problem two separate formulations are possible. First an unconstrained formulation can be written:

$$\min_{\theta} \sum_{\mu=1}^{6} \left[\left(z^r_\mu - \theta_1 - \theta_2\, \omega_\mu^{-\theta_3} \right)^2 + \left(z^i_\mu - \omega_\mu\, \theta_4 + \theta_5\, \omega_\mu^{-\theta_3} \right)^2 \right] \tag{43}$$

Each of the above square terms are treated as general nonconvex terms and an α underestimator is used. The resulting convex relaxation is :

$$\begin{aligned} \min_{\theta} \sum_{\mu=1}^{6} & \left[\left(z^r_\mu - \theta_1 - \theta_2\, \omega_\mu^{-\theta_3} \right)^2 + \alpha_{\mu,1} \sum_{j=1}^{3} \left(\theta^u_j - \theta_j \right) \left(\theta^l_j - \theta_j \right) + \right. \\ & \left. \left(z^i_\mu - \omega_\mu\, \theta_4 + \theta_5\, \omega_\mu^{-\theta_3} \right)^2 + \alpha_{\mu,2} \sum_{j=3}^{5} \left(\theta^u_j - \theta_j \right) \left(\theta^l_j - \theta_j \right) \right] \end{aligned} \tag{44}$$

Due to the nature of the nonconvexities in (44), analytical values of α were not easily identified. Therefore either numerical methods or fixed values need to be used in the solution. A second, constrained formulation can be generated with the use of the substitution $b_\mu = \omega_\mu^{-\theta_3}$. The resulting formulation is:

$$\min_{\hat{\mathbf{z}}_\mu\, \theta} \sum_{\mu=1}^{6} \left[\left(\hat{z}^r_\mu - z^r_\mu \right)^2 + \left(\hat{z}^i_\mu - z^i_\mu \right)^2 \right] \tag{45}$$

s. t.

$$\begin{array}{rcrcrcl} \hat{z}^i_\mu & - & \omega_\mu\,\theta_4 & + & \theta_5\,b_\mu & = & 0 \\ & & \hat{z}^r_\mu & - & \theta_1 & - & \theta_2\,b_\mu = 0 \\ & & & & b_\mu & - & \omega_\mu^{-\theta_3} \leq 0 \\ & & & & -b_\mu & + & \omega_\mu^{-\theta_3} \leq 0 \end{array}$$

The nonconvexities are either in the form of bilinearities, or univariate concave terms ($-\omega_\mu^{-\theta_3}$ is concave, therefore the corresponding positive term is convex). The resulting formulation of the relaxed problem is:

$$\min_{\hat{\mathbf{z}}_\mu\,\theta} \sum_{\mu=1}^{6} \left[\left(\hat{z}^r_\mu - z^r_\mu \right)^2 + \left(\hat{z}^i_\mu - z^i_\mu \right)^2 \right] \tag{46}$$

s. t.

$$\begin{aligned} \hat{z}^i_\mu - \omega_\mu\,\theta_4 + w_{\mu,1} &= 0 \\ \hat{z}^r_\mu - \theta_1 - w_{\mu,2} &= 0 \\ b_\mu - \omega_\mu^{-\theta_3} &\leq 0 \\ -b_\mu + \left\{ \omega_\mu^{-\theta_3^l} + \frac{\omega_\mu^{-\theta_3^u} - \omega_\mu^{-\theta_3^l}}{\theta_3^u - \theta_3^l} \left(\theta_3 - \theta_3^l \right) \right\} &\leq 0 \\ w_{\mu,1} &\rightarrow \theta_5\,b_\mu \\ w_{\mu,2} &\rightarrow \theta_2\,b_\mu \end{aligned}$$

Table 8 provides the problem statistics for each of the two different formulations.

	Constrained		Unconstrained	
	Upper	Lower	Upper	Lower
Total Variables	23	35	5	5
Nonconvex Variables	9	-	5	-
Nonconvex Terms	18	-	12	-
Total Constraints	24	72	-	-
Linear Constraints	0	48	-	-

Table 8: Formulation statistics for Example 4.

The bounds used were [0.0, 1.0] for $\theta_{1,2,4,5}$ and [1.1, 1.3] for θ_3. Since the fitted data variables either do not appear in the problem (unconstrained formulation), or appear only in linear and convex terms (constrained formulation) their bounds do not have any effect on the underestimation. The bounds are therefore set at $z_{\mu,i} \pm (1 + \jmath)$ and no updating is performed.

Table 9 shows the results using both formulations. Using numerically calculated values for α, for the unconstrained formulation, convergence was not achieved in 10,000 iterations. The use of constant α values results in convergence to the known global solution, but theoretical guarantees do not hold since the convexity of the underestimator is not ensured. On the other hand, the constrained formulation converges quickly with theoretical guarantees. The global solution found (using the constrained formulation) was θ = [0.606298, 0.556761, 1.13181, 0.750199, 0.621899], with an objective value of 0.212460. This global solution was obtained in 1.38 CPU sec.

Formulation	α values	Iterations	CPU sec
Unconstrained	Numerical	10000†	687.6
	1	908	51.1
	0.1	65	3.89
	0.01	9	0.661
Constrained	Analytical	11	1.38

Table 9: Convergence Rates for Example 4. An absolute tolerance of 0.01 is used for convergence. (†) Did not converge, absolute difference of 0.0640.

Csendes and Ratz[19] obtained a solution with an objective value enclosed by [0.20788313, 0.21756222] using interval methods to solve the unconstrained formulation. Their solution required about 8 hours on a SUN SPARCSTATION with 222,275 objective function evaluations, and 618,406 partial derivative evaluations.

Example 5: Kowalik Problem

This problem appeared in Moore et al.[18]. The model equation is:

$$z_1 \;=\; \theta_1 \, \frac{z_2^2 \;+\; z_2\,\theta_2}{z_2^2 \;+\; z_2\,\theta_3 \;+\; \theta_4} \tag{47}$$

In this example, only z_1 contains error and therefore z_2 is treated as a constant. Two different formulations are possible. An unconstrained formulation becomes:

$$\min_{\theta} \sum_{\mu=1}^{11} \left[z_{\mu,1} \;-\; \frac{\theta_1\, z_{\mu,2}^2 \;+\; \theta_1\,\theta_2\, z_{\mu,2}}{z_{\mu,2}^2 \;+ z_{\mu,2}\,\theta_3 \;+\; \theta_4} \right]^2 \tag{48}$$

Each term in (48) will be treated as a term of general structure. The resulting formulation of the convex relaxation is:

$$\min_{\theta} \sum_{\mu=1}^{11} \quad \left[z_{\mu,1} \;-\; \frac{\theta_1\, z_{\mu,2}^2 \;+\; \theta_1\,\theta_2\, z_{\mu,2}}{z_{\mu,2}^2 \;+ z_{\mu,2}\,\theta_3 \;+\; \theta_4} \right]^2 \tag{49}$$

$$+ \alpha_\mu \sum_{j=1}^{4} \left(\theta_j^u - \theta_j\right)\left(\theta_j^l - \theta_j\right)$$

The form of the nonconvex terms in (49) does not allow for the calculation of analytical values of α. Therefore in the solution of the lower bounding problem, numerical methods will be used to determine their value. A simpler constrained formulation can be generated by rewriting the model equation:

$$\min_{\hat{z}_{\mu,1}\,\theta} \sum_{\mu=1}^{11} \left(\hat{z}_{\mu,1} - z_{\mu,1}\right)^2 \tag{50}$$

s. t.

$$\hat{z}_{\mu,1}\, z_{\mu,2} - \theta_1\, z_{\mu,2}^2 + z_{\mu,2}\,\hat{z}_{\mu,1}\,\theta_3 + \hat{z}_{\mu,1}\,\theta_4 - z_{\mu,2}\,\theta_2\,\theta_1 = 0$$

In (50) all the nonconvexities are in the form of bilinear terms, making the determination of the convex underestimator much easier. The resulting lower bounding problem is:

$$\min_{\hat{z}_{\mu,1}\,\theta} \sum_{\mu=1}^{11} \left(\hat{z}_{\mu,1} - z_{\mu,1}\right)^2 \tag{51}$$

s. t.

$$\begin{aligned}
&\hat{z}_{\mu,1}\, z_{\mu,2} - \theta_1\, z_{\mu,2}^2 + z_{\mu,2}\, w_{\mu,1} + w_{\mu,2} - z_{\mu,2}\, w_3 = 0\\
&w_{\mu,1} \rightarrow \hat{z}_{\mu,1}\,\theta_3\\
&w_{\mu,2} \rightarrow \hat{z}_{\mu,1}\,\theta_4\\
&w_3 \rightarrow \theta_2\,\theta_1
\end{aligned}$$

Table 10 contains the statistics for both of the formulations.

	Constrained		Unconstrained	
	Upper	Lower	Upper	Lower
Total Variables	15	38	4	4
Nonconvex Variables	15	-	4	-
Nonconvex Terms	23	-	11	-
Total Constraints	11	103	-	-
Linear Constraints	0	103	-	-

Table 10: Formulation Statistics for Example 5.

The bounds on all the parameters are set at: [-0.2892,0.2893] and on the data set variable, $\hat{z}_{\mu,2}$ of $z_{\mu,2} \pm 0.02$. Tables 11 and 12 show the local solutions found for the unconstrained and the constrained formulations, respectively. The unconstrained formulation, as in the previous example, did not converge to global optimality in 10,000 iterations. The constrained problem was solved using initial updating on both the parameters and the data set variables (initial bounds of $z_{\mu,2} \pm 0.01$ and $\delta_2 = 0.005$), as well as bounds updating at each iteration before branching on two randomly chosen candidates from the full set of variables. Convergence (absolute tolerance of 1e-8) was achieved in a total of 1905 αBB iterations in 319.5 CPU sec (two outer iterations).

Objective	θ_1	θ_2	θ_3	θ_4
0.03075×10^{-2}	0.19283	0.19088	0.12314	0.13578
0.12250×10^{-2}	0.22347	$-0.28920^{\dagger}$	0.03645	-0.10512
0.16262×10^{-2}	0.23050	-0.13221	0.25584	-0.05299
7.54451×10^{-2}	0.11559	$-0.28920^{\dagger}$	$-0.28920^{\dagger}$	$-0.28920^{\dagger}$

Table 11: Local solutions for Example 5, unconstrained formulation. (†) Variable is at its lower bound for the problem.

Objective	θ_1	θ_2	θ_3	θ_4
0.03075×10^{-2}	0.19283	0.19088	0.12314	0.13578
0.12250×10^{-2}	0.22347	$-0.28920^{\dagger}$	0.03645	-0.10512
0.20697×10^{-2} ‡	0.23329	-0.18019	0.15986	-0.06347
0.21918×10^{-2} ‡	0.23266	-0.13022	0.20372	-0.04461
0.22631×10^{-2} ‡	0.23353	-0.10108	0.23365	-0.03411
0.23198×10^{-2} ‡	0.23372	-0.07067	0.26351	-0.02338
0.24601×10^{-2} ‡	0.23375	-0.08323	0.25156	-0.02784

Table 12: Local solutions for Example 5, constrained formulation. (†) Variable is at its lower bound for the problem, (‡) one or more data set variables were at their bounds.

Moore et al.[18] employed an interval analysis approach and obtained a solution with an objective value enclosed by [0.00030748585, 0.00030748613] using an unconstrained formulation. Their solution required 4057.1 CPU sec. on a SPARC station SLC using 545,490 function evaluations, 492,923 Jacobian evaluations, and 200,627 Hessian evaluations.

Example 6: CSTR Model

This example features a slightly more complicated, multiple equation model. Consider a model which represents a steady state adiabatic CSTR with an irreversible first order reaction

($A \xrightarrow{k_1} B$) presented by Kim et al.[17].

Five different quantities were measured (using simulated data with noise added): inlet concentration of A, outlet concentrations of A and B, and inlet and outlet temperatures. The model consists of simple mass and energy balances around the reactor:

$$\frac{1}{\tau}(A_o - A) - k_1 A = 0 \tag{52}$$

$$\frac{1}{\tau}(B_o - B) + k_1 A = 0 \tag{53}$$

$$\frac{1}{\tau}(T_o - T) + \frac{-\Delta H_r}{\rho C_p}(k_1 A) = 0 \tag{54}$$

where A_o and B_o are the inlet concentrations of components A and B respectively; A and B are the outlet concentrations of the two components; T_o and T are the inlet and outlet temperatures respectively; τ is the residence time of the reactor (100 s); ΔH_r is the heat of reaction (-4180 J/mol A); ρ is the density of the reaction mixture (1.0 g/l); and C_p is the heat capacity of the reaction mixture (4.18 J/g K). The rate expression for k_1 takes the form:

$$k_1 = c_1 \exp \frac{-Q_1}{RT} \tag{55}$$

where c_1 and Q_1 are the Arrhenius constants (parameters to be determined). Parameter transformations were used in the estimation, which results in the following rate expressions:

$$k_1 = \theta_1 \exp\left[-\theta_2\left(\frac{T_r}{T} - 1\right)\right] \tag{56}$$

with:

$$\theta_1 = c_1 \exp \frac{-Q_1}{RT_r} \tag{57}$$

$$\theta_2 = \frac{Q_1}{RT_r} \tag{58}$$

where T_r is a reference temperature (800 K). The vector of measured variables, z, is defined as $z \equiv [A_o \; A \; B \; T_o \; T]$. Also it is assumed that the feed is pure A, therefore $B_o = 0$. The full nonconvex formulation for the problem is:

$$\min_{\hat{\mathbf{z}}_\mu \, \theta} \sum_{\mu=1}^{10} \sum_{i=1}^{5} \frac{(\hat{z}_{\mu,i} - z_{\mu,i})^2}{\sigma_i^2} \tag{59}$$

s.t.

$$\begin{aligned}
\tfrac{1}{\tau}\,\hat{z}_{\mu,1} \;-\; \tfrac{1}{\tau}\,\hat{z}_{\mu,2} \;-\; k_\mu\,\hat{z}_{\mu,2} \;&=\; 0\\
-\tfrac{1}{\tau}\,\hat{z}_{\mu,3} \;+\; k_\mu\,\hat{z}_{\mu,2} \;&=\; 0\\
\tfrac{1}{\tau}\,\hat{z}_{\mu,4} \;-\; \tfrac{1}{\tau}\,\hat{z}_{\mu,5} \;-\; \tfrac{\Delta H_r}{\rho\,C_p}\,(k_\mu\,\hat{z}_{\mu,2}) \;&=\; 0\\
k_\mu \;-\; \exp\left[\tfrac{-\theta_2\,T_r}{\hat{z}_{\mu,5}} \;+\; \theta_2\right] \;&=\; 0
\end{aligned}$$

In order to aid in the generation of a convex relaxation, various transformations and simplifications can be made. Three different formulations for the convex relaxation can be generated. One common simplification in all three formulations involves the use of a substitution variable to assist in the convexification of the rate expression constraints

$$b_\mu \;=\; \frac{-\theta_2\,T_r}{\hat{z}_{\mu,5}} \;+\; \theta_2 \tag{60}$$

The rate expression constraints are then simplified into:

$$k_\mu \;=\; \theta_1\,\exp(b_\mu) \tag{61}$$

In the first formulation, the term which appears in (61) is treated as a general nonconvex term with analytically determined α values (found in Appendix B), and the substitution constraints are multiplied through by $\hat{z}_{\mu,5}$ to produce terms with are bilinear in nature.

A second formulation can be generated by taking the natural logarithm of the rate expression constraints. This transforms these constraints into:

$$\ln k_\mu \;-\; \ln\theta_1 \;-\; b_\mu \;\le\; 0 \tag{62}$$

$$-\ln k_\mu \;+\; \ln\theta_1 \;+\; b_\mu \;\le\; 0 \tag{63}$$

The nonconvex terms in these constraints are in the form of univariate concave terms ($\ln x$ is concave). These terms are underestimated using a linearization.

A third formulation involves the logarithmic transforms (62) and (63) as well as a fraction term underestimator. In this formulation the substitution constraint is kept in its original form given by (60). The statistics of each of these three formulations is found in Table 13.

The experimental data sets were generated by added normally distributed random noise to a set of 10 simulated data points ($\theta_1 = 0.01717s^{-1}$ and $\theta_2 = 12.58$) found in the reference. The noise added had a standard deviation, $\sigma = [0.001\ 0.001\ 0.001\ 1.0\ 1.0]$. Bounds on the parameters were taken as θ_1: [0.0001:0.1] and θ_2: [5:15]. Initial bounds were set on the fitted data variables $\hat{z}_{\mu,2}$ and $\hat{z}_{\mu,5}$ of $\mathbf{z}_\mu - \sigma \le \hat{\mathbf{z}}_\mu \le \mathbf{z}_\mu + \sigma$ with $\delta_i = \frac{1}{4}\sigma_i$. The other fitted data variable bounds were set to $\pm 3\,\sigma$ since they do not participate in any nonconvex terms. Initial bounds updating was performed on the parameters, θ, and the substitution variables,

	Formulation 1		Formulation 2		Formulation 3	
	Upper	Lower	Upper	Lower	Upper	Lower
Total Variables	72	102	72	102	72	92
Nonconvex Variables	42	-	42	-	32	-
Nonconvex Terms	70	-	70	-	60	-
Total Constraints	60	180	60	180	60	140
Linear Constraints	0	160	0	160	0	100

Table 13: Formulation Statistics for Example 6.

$\mathbf{b}_\mu$, until convergence. The bounds on the substitution variables $\mathbf{b}_\mu$ were also updated in each new region after branching. Branching was performed on both the parameters and the fitted data variables $\hat{z}_{\mu,2}$ and $\hat{z}_{\mu,5}$.

Table 14 shows the results for each of the three different formulations. In each case the same data set was used. The optimization resulted in parameter values of $\theta_1 = 0.0168$ and $\theta_2 = 12.4332$ with an objective function value of 29.04731.

Formulation	Iterations†	Rel. Convergence	CPU sec
1	1000	18.5 %	1253.2
2	174	1%	309.0
3	43	1%	282.2

Table 14: Convergence of different formulations for Example 6. (†) Total αBB iterations for all required outer iterations to reach convergence.

Formulation 1 does not even converge in the first outer iteration. It was found that the α values varied little with the size of the region, thus leading to loose convex relaxations even in small regions. Formulations 2 and 3 both converged in two outer iterations with similar CPU times, but greatly different iteration counts. Formulation 2 required three times the iterations, but with only a 10% increase in CPU time. The increase in iterations can be explained from the difference in the number of nonconvex variables and terms (formulation 3 having 10 less in each category). On the other hand formulation 3 has twice as many nonlinear constraints in the relaxed problem, thus the time required to solve each sub-problem is higher, nearly enough to offset the decrease in the number of iterations.

Example 7: Liquid-Vapor Equilibrium Model

This example appeared in Kim et al.[17]. A two parameter Van Laar equation was used to model binary Vapor-Liquid Equilibrium data which consists of four measured variables (P, T, x, y). The parameters were estimated for the system methanol (1) and 1,2-dichloroethane (2) using five experimental data points found in Reid et al.[32]. First the equilibrium conditions are written:

$$\gamma_1 \, x_1 \, p_1^o(T) \; - \; y \, P \;=\; 0 \tag{64}$$

$$\gamma_2 \, (1 \; - \; x_1) \, p_2^o(T) \; - \; (1 \; - \; y_1) \, P \;=\; 0 \tag{65}$$

where $p_1^o(T)$ and $p_2^o(T)$ are the pure component vapor pressures at the system temperature T; P is the system pressure; x_1 and y_1 are the liquid and vapor fractions respectively of methanol; and γ_1 and γ_2 are the activity coefficients of the components. The pure component vapor pressures are defined using an Antoine equation:

$$p_i^o(T) \;=\; \exp\left[C_{i,1} \; - \; \frac{C_{i,2}}{T \; - \; C_{i,3}}\right] \tag{66}$$

where p_i^o is in mmHg and T is in K. The constants used for each component can be found in Table 15. The activity coefficients are defined using a two parameter Van Laar equation:

$$\gamma_1 \;=\; \exp\left[\frac{A}{RT}\left(1 \; + \; \frac{A}{B}\,\frac{x_1}{1 \; - \; x_1}\right)^{-2}\right] \tag{67}$$

$$\gamma_2 \;=\; \exp\left[\frac{B}{RT}\left(1 \; + \; \frac{B}{A}\,\frac{1 \; - \; x_1}{x_1}\right)^{-2}\right] \tag{68}$$

	$C_{i,1}$	$C_{i,2}$	$C_{i,3}$
Methanol ($i = 1$)	18.5875	3626.55	34.29
1,2-dichloroethane ($i = 2$)	16.1764	2927.17	50.22

Table 15: Antoine Coefficients used in Example 7.

The parameters as well as the system temperature are scaled by a reference temperature, T_r, equal to 323.15 K. The parameter set is then defined as:

$$\theta \;=\; \left[\frac{A}{RT_r} \quad \frac{B}{RT_r}\right]$$

and the data set as:

$$\mathbf{z} = \begin{bmatrix} x_1 & y_1 & \frac{T}{T_r} & P \end{bmatrix}$$

Using these definitions and taking the natural logarithm of the equilibrium expressions, the model results in the following 6 equations

$$\ln(\gamma_1) + \ln(z_1) + \ln(p_1^o) - \ln(z_2) - \ln(P) = 0 \tag{69}$$

$$\ln(\gamma_2) + \ln(1 - z_1) + \ln(p_2^o) - \ln(1 - z_2) - \ln(P) = 0 \tag{70}$$

$$\ln(p_i^o) - C_{i,1} + \frac{C_{i,2}}{T_r\, z_3 - C_{i,3}} = 0 \tag{71}$$

$$\ln(\gamma_1) - \frac{\theta_1}{z_3}\left(1 + \frac{\theta_1}{\theta_2}\frac{z_1}{1 - z_1}\right)^{-2} = 0 \tag{72}$$

$$\ln(\gamma_2) - \frac{\theta_2}{z_3}\left(1 + \frac{\theta_2}{\theta_1}\frac{1 - z_1}{z_1}\right)^{-2} = 0 \tag{73}$$

Substitution of equations (71)-(73) into (69) and (70) results in the following two equation model:

$$\frac{\theta_1}{z_3}\left(1 + \frac{\theta_1}{\theta_2}\frac{z_1}{1 - z_1}\right)^{-2} + \ln(z_1) + C_{1,1} - \frac{C_{1,2}}{T_r z_3 - C_{1,3}} - \ln(z_2) - \ln(P) = 0 \tag{74}$$

$$\frac{\theta_2}{z_3}\left(1 + \frac{\theta_2}{\theta_1}\frac{1 - z_1}{z_1}\right)^{-2} + \ln(1 - z_1) + C_{2,1} - \frac{C_{2,2}}{T_r z_3 - C_{2,3}} - \ln(1 - z_2) - \ln(P) = 0 \tag{75}$$

The standard deviations were chosen similar to the reference[17]:

$$\sigma = \begin{bmatrix} 0.005 & 0.015 & 3.09 \times 10^{-4} & 0.75 \end{bmatrix}$$

In order to aid in the underestimation of various terms in the constraint set, the following substitutions are made:

$$b_1 = \left(1 + \frac{\theta_1}{\theta_2} b_3\right)^{-2} \tag{76}$$

$$b_2 = \left(1 + \frac{\theta_2}{\theta_1} b_4\right)^{-2} \tag{77}$$

$$b_3 = \frac{z_1}{1 - z_1} \tag{78}$$

$$b_4 = \frac{1 - z_1}{z_1} \tag{79}$$

The resulting optimization problem becomes:

$$\min_{\hat{\mathbf{z}}_\mu\,\theta} \; \sum_{\mu=1}^{5}\sum_{i=1}^{4} \frac{\left(\hat{z}_{\mu,i} \;-\; z_{\mu,i}\right)^2}{\sigma_i^2} \tag{80}$$

s. t.

$$\begin{aligned}
\frac{\theta_1\, b_{\mu,1}}{\hat{z}_{\mu,3}} + \ln(\hat{z}_{\mu,1}) + C_{1,1} - \frac{C_{1,2}}{T_r\,\hat{z}_{\mu,3} - C_{1,3}} - \ln(\hat{z}_{\mu,2}) - \ln(\hat{z}_{\mu,4}) &\leq 0\\
-\frac{\theta_1\, b_{\mu,1}}{\hat{z}_{\mu,3}} - \ln(\hat{z}_{\mu,1}) - C_{1,1} + \frac{C_{1,2}}{T_r\,\hat{z}_{\mu,3} - C_{1,3}} + \ln(\hat{z}_{\mu,2}) + \ln(\hat{z}_{\mu,4}) &\leq 0\\
\frac{\theta_2\, b_{\mu,2}}{\hat{z}_{\mu,3}} + \ln(1 - \hat{z}_{\mu,1}) + C_{2,1} - \frac{C_{2,2}}{T_r\,\hat{z}_{\mu,3} - C_{2,3}} - \ln(1 - \hat{z}_{\mu,2}) - \ln(\hat{z}_{\mu,4}) &\leq 0\\
-\frac{\theta_2\, b_{\mu,2}}{\hat{z}_{\mu,3}} - \ln(1 - \hat{z}_{\mu,1}) - C_{2,1} + \frac{C_{2,2}}{T_r\,\hat{z}_{\mu,3} + C_{2,3}} + \ln(1 - \hat{z}_{\mu,2}) + \ln(\hat{z}_{\mu,4}) &\leq 0\\
-\frac{1}{\sqrt{b_{\mu,1}}} + \frac{\theta_1\, b_{\mu,3}}{\theta_2} &\leq -1\\
\frac{1}{\sqrt{b_{\mu,1}}} - \frac{\theta_1\, b_{\mu,3}}{\theta_2} &\leq 1\\
-\frac{1}{\sqrt{b_{\mu,2}}} + \frac{\theta_2\, b_{\mu,4}}{\theta_1} &\leq -1\\
\frac{1}{\sqrt{b_{\mu,2}}} - \frac{\theta_2\, b_{\mu,4}}{\theta_1} &\leq 1\\
b_{\mu,3} - \frac{\hat{z}_{\mu,1}}{1 - \hat{z}_{\mu,1}} &\leq 0\\
-b_{\mu,3} + \frac{\hat{z}_{\mu,1}}{1 - \hat{z}_{\mu,1}} &\leq 0\\
b_{\mu,4} - \frac{1 - \hat{z}_{\mu,1}}{\hat{z}_{\mu,1}} &\leq 0\\
-b_{\mu,4} + \frac{1 - \hat{z}_{\mu,1}}{\hat{z}_{\mu,1}} &\leq 0
\end{aligned}$$

This formulation contains quite a few univariate concave terms. All of the positive natural logarithms as well as the following terms fall under this class and will be underestimated using a linearization:

$$\frac{-C_{1,2}}{T_r\hat{z}_{\mu,3} \;-\; C_{1,3}} \quad \frac{-C_{2,2}}{T_r\hat{z}_{\mu,3} \;-\; C_{2,3}} \quad \frac{-1}{\sqrt{b_{\mu,1}}} \quad \frac{-1}{\sqrt{b_{\mu,2}}} \quad \frac{-\hat{z}_{\mu,1}}{1 \;-\; \hat{z}_{\mu,1}} \quad \frac{-(1 \;-\; \hat{z}_{\mu,1})}{\hat{z}_{\mu,1}}$$

Also the terms with opposite signs of the ones listed above are convex and therefore do not need to be underestimated. The terms of the form $\frac{xy}{z}$ were underestimated using an α underestimator. The expressions used can be found in Appendix B.

Another formulation is possible by underestimating the fractional terms using linear cuts derived in a similar fashion as for bilinear terms. This is accomplished by first substituting an auxiliary variable for the term:

$$w \;=\; \frac{x\,y}{z} \tag{81}$$

This constraint can then be rewritten by multiplied through by z:

$$z\,w \;=\; x\,y \tag{82}$$

Now a single bilinear variable is introduced for these two terms in order to enforce (82). The four additional linear cuts are added with bounds on w taken as:

$$w^U = \frac{x^U\, y^U}{z^l} \qquad w^L = \frac{x^L\, y^L}{z^U} \tag{83}$$

For this formulation the added variables and the terms they relate to are:

$$\begin{aligned}
w_{\mu,5} &\longrightarrow w_{\mu,1}\,\hat{z}_{\mu,3} = \theta_1\, b_{\mu,1}\\
w_{\mu,6} &\longrightarrow w_{\mu,2}\,\hat{z}_{\mu,3} = \theta_2\, b_{\mu,2}\\
w_{\mu,7} &\longrightarrow w_{\mu,3}\,\theta_2 = \theta_1\, b_{\mu,3}\\
w_{\mu,8} &\longrightarrow w_{\mu,4}\,\theta_1 = \theta_2\, b_{\mu,4}
\end{aligned}$$

The statistics for each formulation can be found in Table 16.

	α Formulation		Bilinear Formulation	
	Upper	Lower	Upper	Lower
Total Variables	42	42	42	82
Nonconvex Variables	42	-	42	-
Nonconvex Terms	95	-	95	-
Total Constraints	60	60	100	92
Linear Constraints	0	10	0	42

Table 16: Formulation Statistics for Example 7.

Bounds on the parameters were set at [1:2], while initial bounds on the data variables were set at $\mathbf{z}_\mu \pm \frac{1}{2}\sigma$, with $\delta_i = 0.1\sigma_i$. Initial bounds updating was performed on the parameters, θ, and the substitutions variables $\mathbf{b}_\mu$. Branching was performed on the variable set consisting of the parameters, θ, and the fitted data variables, $\hat{\mathbf{z}}_\mu$. Bounds updating at each region was performed on the substitution variables, $\mathbf{b}_\mu$, as well as 6 randomly chosen variables from $\hat{z}_{\mu,1-3}$. A relative convergence, ϵ^{rel}, of 5% was used. Table 17 compares the performance of each formulation. Both formulations converged in two outer iterations to an objective function of 3.32185, with parameter values, $\theta_1 = 1.9117$ and $\theta_2 = 1.6082$.

The results show that the two different formulations converge in almost exactly the same number of iterations and CPU time. The similar number of iterations is expected since both formulations have the same number of nonconvex variables and terms. The similar CPU times are also expected. Even though the bilinear formulation does have more variables, these additional variables appear only in linear constraints, and both formulations contain the same number of nonlinear constraints.

Formulation	Iterations†	CPU sec
α Formulation	278	1625.47
Bilinear Formulation	260	1693.26

Table 17: Convergence of different formulations for Example 7. (†) Total αBB iterations for all required outer iterations to reach convergence.

5 Conclusion

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 extensive studied and their effect on the performance of the proposed algorithm was illustrated through a series of computation studies on problems taken from the literature.

The problem formulation has a great impact on the performance of the algorithm. It was found that even though local solution methods perform well on an unconstrained formulation, the branch and bound global optimization approach does not converge in a reasonable amount of time. This is attributed to the increased complexity of the nonconvex terms present through the substitution of the nonlinear models into a quadratic objective function. This results in the inability to develop analytical convex underestimators, and even in the case that this is possible the underestimation is not tight. Constrained formulations, on the other hand, offered the ability, in general, to produce tight analytical underestimators and reasonable convergence to the global minimum. Even between similar constrained formulations, differences in the convergence rate have been shown to occur as a result of differences in the generation of the underestimating functions.

Both the branching methods and the variable bounds calculations are influenced by the by the inherent structure of the problem. Two distinct classes of variables are present, the parameters and the fitted data variables. As a result, each class is treated differently in the respective methods. In the generation of tight initial variable bounds, the fitted data variable bounds are found using a series of local optimizations, variable bound expansions, and global optimizations. The parameter bounds, on the other hand, are determined by solving a series of feasibility problems. In each case these iterative approaches have shown to offer great improvements over single pass methods. In the determination of a branching candidate at each iteration in the global optimization, the effect the variables have on the problem is taken into account. Since each parameter variable appears in every set of constraints, they are branched on first. This allows for a rapid initial improvement in the lower bound. At some point, though, the fitted data variables are branched on to allow for final convergence.

The computational results obtained on a variety of the test problems demonstrate the effectiveness, while comparisons with interval analysis methods indicate the favorable performance of the proposed deterministic global optimization approach.

Acknowledgments

Financial support from the National Science Foundation is gratefully acknowledged.

References

(1) Bard, Y. *Nonlinear Parameter Estimation*; Academic Press: New York, 1974.

(2) Bard, Y.; Lapidus, L. Kinetics Analysis by Ditigal Parameter Estimation. *Catalysis Reviews* **1968**, *2* (1), 67.

(3) Southwell, W. H. Fitting Expermental Data. *J. of Comp. Physics* **1969**, *4*, 465.

(4) Schwetlick, H.; Tiller, V. Numerical Methods for Estimating Parameters in Nonlinear Models with Errors in the Variables. *Technometrics* **1985**, *27* (1), 17.

(5) Marquardt, D. W. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *J. Soc. Ind. Appl. Math.* **1963**, *11* (2), 431.

(6) Britt, H. I.; Luecke, R. H. The Estimation of Parameters in Nonlinear Implict Models. *Technometrics* **1973**, *15* (2), 233.

(7) Fariss, R. H.; Law, V. H. An Efficient Computational Technique for the Generalized Application of Maximum Likelihood to Improve Correlation of Experimental Data. *Comp. Chem. Engng.* **1979**, *3*, 95.

(8) Fabries, J.; Renon, H. Method of Evalutation and Reduction of Vapor-Liquid Equilbrium Data of Binary-Mixtures. *AIChE J.* **1975**, *21* (4), 735.

(9) Anderson, T. F.; Abrams, D. S.; Grens II, E. A. Evaluation of Parameters for Nonlinear Thermodynamic Models. *AIChE J.* **1978**, *24* (1), 20.

(10) Patino-Leal, H.; Reilly, P. M. Statistical Estimation of Parameters in Vapor-Liquid Equilibrium. *AIChE J.* **1982**, *28* (4), 580.

(11) Valko, P.; Vajda, S. An Extended Marquardt-type Procedure for Fitting Error-in-Variables Models. *Comp. Chem. Engng.* **1987**, *11* (1), 37.

(12) Rod, V.; Hancil, V. Iterative Estimation of Model Parameters when Measurements of All Variables are Subject to Error. *Comp. Chem. Engng.* **1980**, *4*, 33.

(13) Dovi, V. G.; Paladino, O. Fitting of Experimental Data to Implicit Models Using a Constrained Variation Method. *Comp. Chem. Engng.* **1989**, *13* (6), 731.

(14) Tjoa, T. B.; Biegler, L. T. Reduced Successive Quadratic Programming Strategy for Errors-in-Variables Estimation. *Comp. Chem. Engng.* **1992**, *16* (6), 523.

(15) Ricker, N. L. Comparison of Methods for Nonlinear Parameter Estimation. *Ind. Eng. Process Des. Dev.* **1984**, *23*, 283.

(16) Stewart, W. E.; Caracotsios, M.; Sorensen, J. P. Parameter Estimation from Multiresponse Data. *AIChE J.* **1992**, *38* (5), 641.

(17) Kim, I.; Liebman, M. J.; Edgar, T. F. Robust Error-in-Variables Estimation Using Nonlinear Programming Techniques. *AIChE J.* **1990**, *36* (7), 985.

(18) Moore, R.; Hansen, E.; Leclerc, A. Rigorous Methods for Global Optimization. In *Recent Advances in Global Optimization*; Floudas, C. A.; Pardalos, P. M., Eds.; Princeton Series in Computer Science Princeton University Press, 1992, p 321.

(19) Csendes, T.; Ratz, D. Subdivision Direction Selection in Interval Methods for Global Optimization. *Journal of Global Optimization* **1995**, *7* (2), 183.

(20) Androulakis, I.; Maranas, C. D.; Floudas, C. A. αBB : A Global Optimization Method for General Constrained Nonconvex Problems. *Journal of Global Optimization* **1995**, *7*, 337.

(21) Adjiman, C. S.; Androulakis, I. P.; Maranas, C. D.; Floudas, C. A. A Global Optimization Method, αBB, for Process Design. *Computers chem. Engng, Suppl.* **1996**, *20*, S419.

(22) Adjiman, C. S.; Androulakis, I. P.; Floudas, C. A.; Neumaier, A. A Global Optimization Method, αBB, for General Twice–Differentiable NLPs – I. Theoretical Advances *Computers chem. Engng.* **1998**, in press.

(23) Adjiman, C. S.; Androulakis, I. P.; Floudas, C. A. A Global Optimization Method, αBB, for General Twice–Differentiable NLPs – II. Implementation and Computational Results *Computers chem. Engng.* **1998**, in press.

(24) Al-Khayyal, F. A.; Falk, J. E. Jointly Constrained Biconvex Programming. *Maths Ops Res.* **1983**, *8*, 273.

(25) Al-Khayyal, F. A. Jointly Constrained Bilinear Programs and Related Problems : An Overview. *Computers Math. Applic.* **1990**, *19* (11), 53.

(26) McCormick, G. P. Computationaly of Global Solutions to Factorable Nonconvex Programs: Part I - Convex Underestimations Problems. *Mathematical Programming* **1976**, *10*, 147.

(27) Maranas, C. D.; Floudas, C. A. Finding All Solutions of Nonlinearly Constrained Systems of Equations. *Journal of Global Optimization* **1995**, *7* (2), 143.

(28) Maranas, C. D.; Floudas, C. Global Minimum Potential Energy Conformations of Small Molecules. *Journal of Global Optimization* **1994b**, *4*, 135.

(29) Harding, S. T.; Floudas, C. A. Global Optimization in Multiproduct and Multipurpose Batch Design under Uncertainty. *I&EC Res.* **1997**, *36* (5), 1644.

(30) Adjiman, C. S.; Floudas, C. A. Rigorous Convex Underestimators for General Twice–Differentiable Problems. *Journal of Global Optimization* **1996**, *9*, 23.

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

(32) Reid, R. C.; Prausnitz, J. M.; Sherwood, T. K. *The Properties of Gases and Liquids*; McGraw-Hill Book Company, 3rd ed. 1977.

Appendix A: Data and Fitted Variable Values

Data		Example 1		Example 2	
$z_{\mu,1}$	$z_{\mu,2}$	$\hat{z}_{\mu,1}$	$\hat{z}_{\mu,2}$	$\hat{z}_{\mu,1}$	$\hat{z}_{\mu,2}$
0.0	5.9	-0.049	5.81	0.057	5.95
0.9	5.4	0.86	5.32	0.82	5.29
1.8	4.4	1.97	4.71	1.90	4.58
2.6	4.6	2.50	4.42	2.45	4.29
3.3	3.5	3.50	3.87	3.44	3.84
4.4	3.7	4.27	3.46	4.31	3.48
5.2	2.8	5.26	2.91	5.31	3.02
6.1	2.8	5.96	2.53	6.00	2.64
6.5	2.4	6.43	2.27	6.46	2.35
7.4	1.5	7.50	1.69	7.45	1.55

Table 18: Data and fitted values for Examples 1 and 2.

Data	Fitted Values
$5 - 5\jmath$	$5.13 - 4.93\jmath$
$3 - 2\jmath$	$2.67 - 2.07\jmath$
$2 - \jmath$	$1.91 - 1.10\jmath$
$1.5 - 0.5\jmath$	$1.54 - 0.58\jmath$
$1.2 - 0.2\jmath$	$1.34 - 0.23\jmath$
$1.1 - 0.1\jmath$	$1.20 + 0.04\jmath$

Table 19: Data and fitted values for Example 4.

Data		Fitted Values
$z_{\mu,1}$	$1/z_{\mu,2}$	$\hat{z}_{\mu,1}$
0.1957	0.25	0.1944
0.1947	0.5	0.1928
0.1735	1	0.1824
0.1600	2	0.1489
0.0844	4	0.0928
0.0627	6	0.0624
0.0456	8	0.0456
0.0342	10	0.0355
0.0323	12	0.0288
0.0235	14	0.0241
0.0246	16	0.0207

Table 20: Data and fitted values for Example 5.

z_1		z_2		z_3		z_4		z_5	
data	fitted	data	fitted	data	fitted	data	fitted	data	fitted
0.9871	0.9985	0.8906	0.8826	0.1157	0.1159	547.47	547.84	663.48	663.78
1.0003	0.9878	0.8350	0.8437	0.1380	0.1441	531.77	531.49	676.04	675.63
1.0039	1.0011	0.8255	0.8282	0.1850	0.1729	512.21	512.06	684.81	684.95
0.9760	0.9920	0.8020	0.7874	0.2005	0.2046	490.59	490.84	695.47	695.44
1.0129	1.0058	0.7520	0.7660	0.2420	0.2398	464.67	464.80	703.69	704.62
1.0083	1.0005	0.7193	0.7242	0.2739	0.2763	438.47	438.34	714.90	714.62
1.0075	0.9995	0.6861	0.6822	0.3215	0.3173	408.04	407.66	726.09	724.94
0.9994	0.9997	0.6388	0.6377	0.3741	0.3600	375.56	375.35	735.44	735.32
1.0007	1.0013	0.5970	0.5962	0.3926	0.4052	340.26	340.39	745.70	745.55
0.9973	1.0071	0.5580	0.5587	0.4703	0.4484	306.55	306.56	753.94	754.96

Table 21: Data and fitted values for Example 6.

z_1		z_2		z_3		z_4	
data	fitted	data	fitted	data	fitted	data	fitted
0.30	0.299	0.591	0.596	1.00	1.00	483.80	483.99
0.40	0.400	0.602	0.612	1.00	1.00	493.20	493.28
0.50	0.500	0.612	0.624	1.00	1.00	499.90	499.67
0.70	0.699	0.657	0.667	1.00	1.00	501.40	501.29
0.90	0.901	0.814	0.810	1.00	1.00	469.70	469.71

Table 22: Data and fitted values for Example 7.

Appendix B: α Calculations for the Examples

Example 2

$$\begin{array}{rcl} \hat{z}_{\mu,1}^3 & \longrightarrow & \alpha_{\mu,1} = \max\left(0\,,\,-3\,\hat{z}_{\mu,1}^l\right) \\ -\hat{z}_{\mu,1}^3 & \longrightarrow & \alpha_{\mu,2} = \max\left(0\,,\,3\,\hat{z}_{\mu,1}^u\right) \end{array}$$

Example 3

<u>α formulation:</u>

$$\frac{1}{\hat{z}_{\mu,1} - \theta_2} \longrightarrow \alpha_\mu = \frac{2}{(\theta_2^l - \hat{z}_{\mu,1}^u)^3}$$

<u>unconstrained formulation:</u>

$$\frac{2\,\theta_1 - 2\,z_{\mu,2}}{\hat{z}_{\mu,1} - \theta_2} \longrightarrow \alpha_\mu = 2 \max_{\hat{z}_{\mu,1}\,\theta_1\,\theta_2} \frac{z_{\mu,2} - \theta_1 - \frac{\sqrt{2}}{2}\sqrt{\left(\hat{z}_{\mu,1} - \theta_2\right)^2 + 2\left(\theta_1 - z_{\mu,2}\right)^2}}{\left(\hat{z}_{\mu,1} - \theta_2\right)^3}$$

Example 6

$$\frac{\theta_1\,b_{\mu,1}}{\hat{z}_{\mu,3}} \longrightarrow \alpha = \frac{-\theta_1^U\,b_{\mu,1}^L + \sqrt{(\theta_1^U)^2(b_{\mu,1}^L)^2 + (\hat{z}_{\mu,3}^L)^2(\theta_1^U)^2 + (\hat{z}_{\mu,3}^L)^4 + (\hat{z}_{\mu,3}^L)^2(b_{\mu,1}^L)^2}}{2(\hat{z}_{\mu,3}^L)^3}$$

$$-\frac{\theta_1\,b_{\mu,1}}{\hat{z}_{\mu,3}} \longrightarrow \alpha = \frac{\theta_1^U\,b_{\mu,1}^U + \sqrt{(\theta_1^U)^2(b_{\mu,1}^U)^2 + (\hat{z}_{\mu,3}^L)^2(\theta_1^U)^2 + (\hat{z}_{\mu,3}^L)^4 + (\hat{z}_{\mu,3}^L)^2(b_{\mu,1}^U)^2}}{2(\hat{z}_{\mu,3}^L)^3}$$

$$\frac{\theta_2\,b_{\mu,2}}{\hat{z}_{\mu,3}} \longrightarrow \alpha = \frac{-\theta_2^U\,b_{\mu,2}^L + \sqrt{(\theta_2^U)^2(b_{\mu,2}^L)^2 + (\hat{z}_{\mu,3}^L)^2(\theta_2^U)^2 + (\hat{z}_{\mu,3}^L)^4 + (\hat{z}_{\mu,3}^L)^2(b_{\mu,2}^L)^2}}{2(\hat{z}_{\mu,3}^L)^3}$$

$$-\frac{\theta_2\,b_{\mu,2}}{\hat{z}_{\mu,3}} \longrightarrow \alpha = \frac{\theta_2^U\,b_{\mu,2}^U + \sqrt{(\theta_2^U)^2(b_{\mu,2}^U)^2 + (\hat{z}_{\mu,3}^L)^2(\theta_2^U)^2 + (\hat{z}_{\mu,3}^L)^4 + (\hat{z}_{\mu,3}^L)^2(b_{\mu,2}^U)^2}}{2(\hat{z}_{\mu,3}^L)^3}$$

$$\frac{\theta_1\,b_{\mu,3}}{\theta_2} \longrightarrow \alpha = \max\left[\frac{-\theta_1^U\,b_{\mu,3}^L + \sqrt{(\theta_1^U)^2(b_{\mu,3}^L)^2 + (\theta_2^L)^2(\theta_1^U)^2 + (\theta_2^L)^4 + (\theta_2^L)^2(b_{\mu,3}^L)^2}}{2(\theta_2^L)^3},\right.$$
$$\left.\frac{-\theta_1^U\,b_{\mu,3}^U + \sqrt{(\theta_1^U)^2(b_{\mu,3}^U)^2 + (\theta_2^L)^2(\theta_1^U)^2 + (\theta_2^L)^4 + (\theta_2^L)^2(b_{\mu,3}^U)^2}}{2(\theta_2^L)^3}\right]$$

$$-\frac{\theta_1\,b_{\mu,3}}{\theta_2} \longrightarrow \alpha = \frac{\theta_1^U\,b_{\mu,3}^U + \sqrt{(\theta_1^U)^2(b_{\mu,3}^U)^2 + (\theta_2^L)^2(\theta_1^U)^2 + (\theta_2^L)^4 + (\theta_2^L)^2(b_{\mu,3}^U)^2}}{2(\theta_2^L)^3}$$

$$\frac{\theta_1\, b_{\mu,4}}{\theta_1} \quad \longrightarrow \quad \alpha = \max\left[\frac{-\theta_2^U\, b_{\mu,4}^L + \sqrt{(\theta_2^U)^2(b_{\mu,4}^L)^2 \;+\; (\theta_1^L)^2(\theta_2^U)^2 \;+\; (\theta_1^L)^4 \;+\; (\theta_1^L)^2(b_{\mu,4}^L)^2}}{2(\theta_1^L)^3}, \right.$$

$$\left. \frac{-\theta_1^U\, b_{\mu,3}^U \;+\; \sqrt{(\theta_2^U)^2(b_{\mu,4}^U)^2 \;+\; (\theta_1^L)^2(\theta_2^U)^2 \;+\; (\theta_1^L)^4 \;+\; (\theta_1^L)^2(b_{\mu,4}^U)^2}}{2(\theta_1^L)^3}\right]$$

$$-\frac{\theta_2\, b_{\mu,4}}{\theta_1} \quad \longrightarrow \quad \alpha \;=\; \frac{\theta_2^U\, b_{\mu,4}^U \;+\; \sqrt{(\theta_2^U)^2(b_{\mu,4}^U)^2 \;+\; (\theta_1^L)^2(\theta_2^U)^2 \;+\; (\theta_1^L)^4 \;+\; (\theta_1^L)^2(b_{\mu,4}^U)^2}}{2(\theta_1^L)^3}$$