Reliable Process Modeling and Optimization Using Interval Analysis

Chao-Yang Gau, Robert W. Maier, Gang Xu and Mark A. Stadtherr

Department of Chemical Engineering
University of Notre Dame
Notre Dame, IN 46556 USA

6th SIAM Conference on Optimization
Atlanta, GA, May 10–12, 1999

1Author to whom all correspondence should be addressed. Phone: (219)631-9318; Fax: (219)631-8366; E-mail: markst@nd.edu
Outline

- Motivation
- Background
- Examples: Parameter Estimation in Modeling Vapor-Liquid Equilibrium (VLE)
- Results
- Summary and Concluding Remarks
Motivation

- In process modeling, chemical engineers frequently need to solve nonlinear equation systems in which the variables are constrained physically within upper and lower bounds; that is, to solve:

\[ f(x) = 0 \]
\[ x^L \leq x \leq x^U \]

- These problems may:
  - Have multiple solutions
  - Have no solution
  - Be difficult to converge to any solution
Motivation (continued)

• There is also frequent interest in globally minimizing a nonlinear function subject to nonlinear equality and/or inequality constraints; that is, to solve (globally):

\[
\min_x \phi(x)
\]

subject to

\[
\begin{align*}
  h(x) &= 0 \\
  g(x) &\geq 0 \\
  x^L &\leq x \leq x^U
\end{align*}
\]

• These problems may:
  
  – Have multiple local minima (in some cases, it may be desirable to find them all)
  – Have no solution (infeasible NLP)
  – Be difficult to converge to any local minima
Motivation (continued)

• One approach for dealing with these issues is *interval analysis*.

• Interval analysis can
  – Provide the engineer with tools needed to solve modeling and optimization problems with complete certainty.
  – Provide problem-solving reliability not available when using standard local methods.
  – Deal automatically with rounding error, thus providing both mathematical and computational guarantees.
Motivation (continued)

- At Notre Dame, we have successfully applied interval methods for
  - General process modeling problems (Schnepper and Stadtherr, 1996).
  - Phase stability and equilibrium problems using several different thermodynamic models (Stadtherr et al., 1994; Hua et al., 1996, 1998, 1999; Xu et al., 1998).
  - Computation of azeotropes of multicomponent mixtures (Maier et al., 1998, 1999).
  - Computation of mixture critical points (Stradi et al., 1998)

Background—Interval Analysis

- A real interval \( X = [a, b] = \{ x \in \mathbb{R} \mid a \leq x \leq b \} \) is a segment on the real number line and an interval vector \( \mathbf{X} = (X_1, X_2, \ldots, X_n)^T \) is an \( n \)-dimensional rectangle or “box”.

- Basic interval arithmetic for \( X = [a, b] \) and \( Y = [c, d] \) is \( X \ \text{op} \ Y = \{ x \ \text{op} \ y \mid x \in X, y \in Y \} \) where \( \text{op} \in \{+,-,\times,\div\} \). For example, \( X + Y = [a + c, b + d] \).

- Computed endpoints are rounded out to guarantee the enclosure.

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

- The interval extension \( F(\mathbf{X}) \) encloses all values of \( f(x) \) for \( x \in \mathbf{X} \). That is, \( F(\mathbf{X}) \supseteq \{ f(x) \mid x \in \mathbf{X} \} \).

- Interval extensions can be computed using interval arithmetic (the “natural” interval extension), or with other techniques.
Background—Parameter Estimation

- Observations $y_{\mu i}$ of $i = 1, \ldots, q$ responses from $\mu = 1, \ldots, p$ experiments are available.

- Responses are to be fit to a model $y_{\mu i} = f_i(x_{\mu}, \theta)$ with independent variables $x_{\mu} = (x_{\mu 1}, \ldots, x_{\mu m})^T$ and parameters $\theta = (\theta_1, \ldots, \theta_n)^T$. Measurement errors in $x_{\mu}$ can either be neglected or treated using the "error-in-variable" approach.

- Various objective functions $\phi(\theta)$ can be used to determine the parameter values that provide the "best" fit, e.g.
  - Maximum likelihood
  - Relative or weighted least squares

The latter will be used here.

- Optimization problem to determine parameters can usually be formulated as either a constrained or unconstrained problem. The unconstrained formulation is used here.
Parameter Estimation

- Assuming a relative least squares objective and using an unconstrained formulation, the problem is

\[
\min_{\theta} \phi(\theta) = \sum_{i=1}^{q} \sum_{\mu=1}^{p} \left[ \frac{y_{\mu i} - f_i(x_{\mu}, \theta)}{y_{\mu i}} \right]^2
\]

- A common approach for solving this problem is to use the gradient of \( \phi(\theta) \) and to seek the stationary points of \( \phi(\theta) \) by solving \( g(\theta) \equiv \nabla \phi(\theta) = 0 \). This system may have many roots, including local minima, local maxima and saddle points.

- To insure that the global minimum of \( \phi(\theta) \) is found, the capability to find all the roots of \( g(\theta) = 0 \) is needed. This is provided by the interval Newton technique.

- Interval Newton can be combined with branch and bound so that roots of \( g(\theta) = 0 \) that cannot be the global minimum need not be found.
Background–Interval Newton Method

• For the system of nonlinear equations $g(\theta) = 0$, find (enclose) all roots in a given initial interval $\Theta^{(0)}$ or determine that there are none.

• At iteration $k$, given the interval $\Theta^{(k)}$, if $0 \in G(\Theta^{(k)})$ solve the linear interval equation system

$$G'(\Theta^{(k)})(N^{(k)} - \theta^{(k)}) = -g(\theta^{(k)})$$

for the “image” $N^{(k)}$, where $G(\Theta^{(k)})$ is an interval extension of $g(\theta)$ and $G'(\Theta^{(k)})$ an interval extension of its Jacobian over the current interval $\Theta^{(k)}$, and $\theta^{(k)}$ is a point inside $\Theta^{(k)}$.

• Any root $\theta^* \in \Theta^{(k)}$ is also contained in the image $N^{(k)}$, suggesting the iteration scheme $\Theta^{(k+1)} = \Theta^{(k)} \cap N^{(k)}$ (Moore, 1966).

• It follows that if $\Theta^{(k)} \cap N^{(k)} = \emptyset$, then there is no root in $\Theta^{(k)}$. This is also the conclusion if $0 \notin G(\Theta^{(k)})$. 

10
Interval Newton Method (continued)

• Interval Newton provides an existence and uniqueness test: If $\mathcal{N}^{(k)} \subset \Theta^{(k)}$, then:
  
  – There is a **unique** zero of $g(\theta)$ in $\Theta^{(k)}$.
  – The interval Newton iteration $\Theta^{(k+1)} = \Theta^{(k)} \cap \mathcal{N}^{(k)}$ will converge quadratically to a tight enclosure of the root.
  – The point Newton method will converge quadratically to the root starting from any point in $\Theta^{(k)}$.

• If a unique root cannot be confirmed ($\mathcal{N}^{(k)} \subset \Theta^{(k)}$) or ruled out ($\Theta^{(k)} \cap \mathcal{N}^{(k)} = \emptyset$), then either:
  
  – Continue with the next iterate $\Theta^{(k+1)}$ if it is sufficiently smaller than $\mathcal{N}^{(k)}$, or
  – **Bisect** $\Theta^{(k+1)}$ and perform interval Newton on the resulting intervals.

This is the interval Newton/generalized bisection (IN/GB) approach.
Any solutions in $\Theta^{(k)}$ are also in intersection of $\Theta^{(k)}$ and $N^{(k)}$. 
There was no solution in $\Theta^{(k)}$
Unique solution in $\Theta^{(k)}$
This solution is in $N^{(k)}$
Point Newton method will converge to it
Interval Newton Method (continued)

• For $g(\theta) = 0$, this method can enclose with mathematical and computational certainty any and all solutions in a given initial interval, or can determine that there are none.

• A preconditioned interval Gauss-Seidel like technique is often used to solve for the image $N^{(k)}$ (Hansen and coworkers).

• Our implementation is based on modifications of routines taken from the packages INTBIS and INTLIB (Kearfott and coworkers).

• The interval Newton procedure can be performed on multiple intervals independently and in parallel.

• IN/GB was first implemented for process modeling problems by Schnepper and Stadtherr (1990).
Parameter Estimation in VLE Modeling

- Goal: Determine parameter values in liquid phase activity coefficient models (e.g. Wilson, van Laar, NRTL, UNIQUAC):
  \[ \gamma_{i\mu,\text{calc}} = f_i(x_{\mu}, \theta) \]

- The relative least squares objective is commonly used:
  \[ \phi(\theta) = \sum_{i=1}^{n} \sum_{\mu=1}^{p} \left[ \frac{\gamma_{i\mu,\text{calc}(\theta)} - \gamma_{i\mu,\text{exp}}}{\gamma_{i\mu,\text{exp}}} \right]^2. \]

- Experimental values \( \gamma_{i\mu,\text{exp}} \) of the activity coefficients are obtained from VLE measurements at compositions \( x_{\mu}, \mu = 1, \ldots, p \).

- Fit is usually made to binary (sometimes ternary) data. Other types of experimental data may also be used.
Example Problem 1

- The binary system water (1) and formic acid (2) was studied.

- Eleven problems, each a different data set from the DEHEMA VLE Data Collection (Gmehling et al., 1977-1990) were considered.

- The model used was the Wilson equation. This has binary interaction parameters
  \[ \Lambda_{12} = \left( \frac{v_2}{v_1} \right) \exp\left( -\frac{\theta_1}{RT} \right) \] and 
  \[ \Lambda_{21} = \left( \frac{v_1}{v_2} \right) \exp\left( -\frac{\theta_2}{RT} \right) \]
  where \( v_1 \) and \( v_2 \) are pure component molar volumes.

- The energy parameters \( \theta_1 \) and \( \theta_2 \) must be estimated.

- Parameter estimation results for \( \theta_1 \) and \( \theta_2 \) are given in the DEHEMA Collection for all eleven problems.
Results–Example 1

• Each problem was solved using the IN/GB approach to determine the globally optimal values of the $\theta_1$ and $\theta_2$ parameters.

• These results were compared to those presented in the DECHEMA Collection.

• For each problem, the number of local minima in $\phi(\theta)$ was also determined (branch and bound steps were turned off).

• Table 1 presents a summary of these results and comparisons. CPU times are on a Sun Ultra 2/1300 workstation.

• Detailed results for one data set will be shown.
<table>
<thead>
<tr>
<th>Data Set</th>
<th>Data points</th>
<th>$P$ (mmHg)</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\phi(\theta)$</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\phi(\theta)$</th>
<th>No. of Minima</th>
<th>CPU time(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15</td>
<td>760</td>
<td>-195</td>
<td>759</td>
<td>0.0342</td>
<td>-195</td>
<td>759</td>
<td>0.0342</td>
<td>2</td>
<td>14.7</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>760</td>
<td>-278</td>
<td>1038</td>
<td>0.0106</td>
<td>-278</td>
<td>1038</td>
<td>0.0106</td>
<td>2</td>
<td>12.9</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>760</td>
<td>-310</td>
<td>1181</td>
<td>0.0151</td>
<td>-308</td>
<td>1167</td>
<td>0.0151</td>
<td>2</td>
<td>17.4</td>
</tr>
<tr>
<td>4</td>
<td>28</td>
<td>760</td>
<td>-282</td>
<td>985</td>
<td>0.353</td>
<td>-282</td>
<td>984</td>
<td>0.353</td>
<td>2</td>
<td>22.0</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>760</td>
<td>-366</td>
<td>1513</td>
<td>0.0257</td>
<td>-365</td>
<td>1509</td>
<td>0.0257</td>
<td>3</td>
<td>16.0</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>760</td>
<td>1067</td>
<td>-1122</td>
<td>0.0708</td>
<td>1065</td>
<td>-1120</td>
<td>0.0708</td>
<td>2</td>
<td>21.7</td>
</tr>
<tr>
<td>7*</td>
<td>14</td>
<td>200</td>
<td>892</td>
<td>-985</td>
<td>0.1410</td>
<td>-331</td>
<td>1250</td>
<td>0.0914</td>
<td>2</td>
<td>19.5</td>
</tr>
<tr>
<td>8*</td>
<td>11</td>
<td>200</td>
<td>370</td>
<td>-608</td>
<td>0.0459</td>
<td>-340</td>
<td>1404</td>
<td>0.0342</td>
<td>3</td>
<td>17.8</td>
</tr>
<tr>
<td>9*</td>
<td>16</td>
<td>100</td>
<td>539</td>
<td>-718</td>
<td>0.1650</td>
<td>-285</td>
<td>996</td>
<td>0.1114</td>
<td>2</td>
<td>24.4</td>
</tr>
<tr>
<td>10*</td>
<td>19</td>
<td>100</td>
<td>450</td>
<td>-663</td>
<td>0.1510</td>
<td>-329</td>
<td>1394</td>
<td>0.0819</td>
<td>3</td>
<td>27.4</td>
</tr>
<tr>
<td>11*</td>
<td>9</td>
<td>70</td>
<td>558</td>
<td>-762</td>
<td>0.0399</td>
<td>-330</td>
<td>1519</td>
<td>0.0372</td>
<td>3</td>
<td>16.4</td>
</tr>
</tbody>
</table>

*: New globally optimal parameters found.
Results–Example 1 (continued)

- Each problem has multiple local minima.

- In five of the problems (data sets 7–11), the result presented in DECHEMA represents a local not global minimum.

- Using the interval approach, the global minimum was found for all problems.

- The parameter estimation results obtained from the global minimization were more consistent than those in DECHEMA.

- There are several other systems for which the results given in the DECHEMA Collection do not represent the globally best fit.
Detailed Results–Data Set 10

• This problem has five stationary points, including three minima and two saddles. Details are shown in Table 2.

• These will not all be found if the branch and bound steps are turned on, so that only the global minimum is actually sought.

• The globally optimal parameters found using the interval approach provide a noticeably better fit to the experimental data. This is shown by the relative deviation plots given in Figures 1 and 2.
### TABLE 2: Stationary Points (Roots) for Data Set 10

<table>
<thead>
<tr>
<th>Root</th>
<th>$\theta_1, \theta_2$</th>
<th>Eigenvalues of Hessian</th>
<th>$\phi(\theta)$</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>(1658, -1251)</td>
<td>7.55E-5, 2.58E-7</td>
<td>0.164</td>
<td>minimum</td>
</tr>
<tr>
<td>P2</td>
<td>(1165, -1083)</td>
<td>6.83E-5, -1.44E-7</td>
<td>0.178</td>
<td>saddle</td>
</tr>
<tr>
<td>P3</td>
<td>(452, -664)</td>
<td>6.97E-5, 9.42E-8</td>
<td>0.151</td>
<td>minimum</td>
</tr>
<tr>
<td>P4</td>
<td>(-37.8, 38.5)</td>
<td>9.08E-5, -3.54E-7</td>
<td>0.19</td>
<td>saddle</td>
</tr>
<tr>
<td>P5</td>
<td>(-329, 1394)</td>
<td>1.23E-4, 1.47E-7</td>
<td>0.0819</td>
<td>global minimum</td>
</tr>
</tbody>
</table>
FIGURE 1: Example 1—Comparison of Relative Deviation in $\gamma_1$
FIGURE 2: Example 1–Comparison of Relative Deviation in $\gamma_2$
Example Problem 2

• The binary system \textit{tert}-butanol (1) and 1-butanol (2) was studied.

• Six problems, each a different data set from the DEHEMA VLE Data Collection (Gmehling \textit{et al.}, 1977-1990) were considered.

• The model used was the Wilson equation.

• Table 3 compares parameter estimation results for $\theta_1$ and $\theta_2$ with those given in the DEHEMA Collection. New globally optimal parameter values are found in all six cases.

• The globally optimal parameters found using the interval approach provide a noticeably better fit to the experimental data. This is shown by the relative deviation plots given in Figures 3 and 4 for data set 6.
<table>
<thead>
<tr>
<th>Data Set</th>
<th>Data points</th>
<th>$P$ (mmHg)</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\phi(\theta)$</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\phi(\theta)$</th>
<th>No. of Minima</th>
<th>CPU time(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1*</td>
<td>9</td>
<td>100</td>
<td>951</td>
<td>-602</td>
<td>0.0136</td>
<td>-568</td>
<td>745</td>
<td>0.0103</td>
<td>2</td>
<td>12.8</td>
</tr>
<tr>
<td>2*</td>
<td>9</td>
<td>300</td>
<td>1068</td>
<td>-638</td>
<td>0.0158</td>
<td>-525</td>
<td>626</td>
<td>0.0130</td>
<td>2</td>
<td>11.1</td>
</tr>
<tr>
<td>3*</td>
<td>9</td>
<td>500</td>
<td>901</td>
<td>-594</td>
<td>0.0097</td>
<td>-718</td>
<td>1265</td>
<td>0.0069</td>
<td>2</td>
<td>12.4</td>
</tr>
<tr>
<td>4*</td>
<td>9</td>
<td>700</td>
<td>801</td>
<td>-561</td>
<td>0.0174</td>
<td>-734</td>
<td>1318</td>
<td>0.0137</td>
<td>2</td>
<td>12.9</td>
</tr>
<tr>
<td>5*</td>
<td>17</td>
<td>760</td>
<td>153</td>
<td>-203</td>
<td>0.1300</td>
<td>-793</td>
<td>1757</td>
<td>0.1164</td>
<td>2</td>
<td>16.2</td>
</tr>
<tr>
<td>6*</td>
<td>14</td>
<td>760</td>
<td>848</td>
<td>-606</td>
<td>0.0333</td>
<td>-865</td>
<td>2420</td>
<td>0.0111</td>
<td>2</td>
<td>11.0</td>
</tr>
</tbody>
</table>

*: New globally optimal parameters found.
FIGURE 1: Example 2—Comparison of Relative Deviation in $\gamma_1$

\[ \frac{\gamma_{1,\text{calc}} - \gamma_{1,\text{exp}}}{\gamma_{1,\text{exp}}} \]
FIGURE 2: Example 2–Comparison of Relative Deviation in $\gamma_2$
Example Problem 3

- The binary system benzene (1) and hexafluorobenzene (2) was studied.

- Ten problems, each a different data set from the DEHEMA VLE Data Collection (Gmehling et al., 1977-1990) were considered.

- The model used was the Wilson equation.

- Table 4 compares parameter estimation results for $\theta_1$ and $\theta_2$ with those given in the DEHEMA Collection. New globally optimal parameter values are found in five cases.
<table>
<thead>
<tr>
<th>Data Set</th>
<th>Data points</th>
<th>$T \ (^\circ C')$</th>
<th>DECHEMA</th>
<th>IN/GB</th>
<th>No. of Minima</th>
<th>CPU time(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1*</td>
<td>10</td>
<td>30</td>
<td>$\theta_1$</td>
<td>$\theta_2$</td>
<td>$\phi(\theta)$</td>
<td>$\theta_1$</td>
</tr>
<tr>
<td>2*</td>
<td>10</td>
<td>40</td>
<td>405</td>
<td>-405</td>
<td>0.0327</td>
<td>-459</td>
</tr>
<tr>
<td>3*</td>
<td>10</td>
<td>50</td>
<td>374</td>
<td>-374</td>
<td>0.0289</td>
<td>-449</td>
</tr>
<tr>
<td>4*</td>
<td>11</td>
<td>50</td>
<td>342</td>
<td>-342</td>
<td>0.0428</td>
<td>-424</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>60</td>
<td>-439</td>
<td>1096</td>
<td>0.0047</td>
<td>-439</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>70</td>
<td>-424</td>
<td>1035</td>
<td>0.0032</td>
<td>-425</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Data points</th>
<th>$P \ (\text{mmHg})$</th>
<th>DECHEMA</th>
<th>IN/GB</th>
<th>No. of Minima</th>
<th>CPU time(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7*</td>
<td>17</td>
<td>300</td>
<td>$\theta_1$</td>
<td>$\theta_2$</td>
<td>$\phi(\theta)$</td>
<td>$\theta_1$</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
<td>500</td>
<td>-405</td>
<td>906</td>
<td>0.0083</td>
<td>-407</td>
</tr>
<tr>
<td>9</td>
<td>17</td>
<td>760</td>
<td>-407</td>
<td>923</td>
<td>0.0057</td>
<td>-399</td>
</tr>
<tr>
<td>10</td>
<td>17</td>
<td>760</td>
<td>-333</td>
<td>702</td>
<td>0.0146</td>
<td>-335</td>
</tr>
</tbody>
</table>

*: New globally optimal parameters found.
Example Problem 3–Discussion

• Does the use of the globally optimal parameters make a significant difference when the Wilson model is used to predict vapor-liquid equilibrium (VLE)?

• A common test of the predictive power of a model for VLE is its ability to predict azeotropes.

• Using the globally optimal parameters, the Wilson equation predicts two azeotropes, which is correct.

• Using the parameters in DEHEMA that are only locally optimal, the number of azeotropes predicted is zero (three data sets) or one (two data sets).

• Using the globally optimal parameters makes the difference between predicting physical reality or not.
Computational Performance

• With initial parameter intervals of $\Theta_1^{(0)} = \Theta_2^{(0)} = [-8500, 320000]$, the computation times required ranged from roughly 10 to 25 seconds on a Sun Ultra 2/1300 workstation.

• Initial parameter intervals were chosen based on physical knowledge of infinite dilution activity coefficients.

• An inverse midpoint preconditioner was used. Significant improvements in computation time are possible using an improved preconditioner.

• Because of the wide initial interval that can be used, as opposed to an initial point guess, the method is essentially初始化independent.

• The additional computation time for the interval approach, as opposed to local methods, is compensated by the guaranteed global reliability of the results.
Concluding Remarks

- Interval analysis is a general-purpose and model-independent approach for solving parameter estimation problems in modeling VLE, providing a mathematical and computational guarantee that the global optimum is found.
  - Other VLE models could be used.
  - Other objective functions (e.g., maximum likelihood) could be used.
  - Error-in-variable approach could be used.
  - Other types of data could be used.

- Interval analysis provides powerful problem solving techniques with many other applications in the modeling of thermodynamics and phase behavior and in other process modeling problems.

- Continuing advances in computing hardware and software (e.g., compiler support for interval arithmetic) will make this approach even more attractive.
• Acknowledgments
  – ACS PRF 30421-AC9
  – NSF CTS95-22835
  – NSF DMI96-96110
  – NSF EEC97-00537-CRCD
  – EPA R824731-01-0
  – DOE DE-FG07-96ER14691
  – Sun Microsystems, Inc.

• For more information:
  – Contact Prof. Stadtherr at markst@nd.edu