Nonlinear Parameter Estimation Using Interval Analysis

Chao-Yang Gau and Mark A. Stadtherr[1]
Department of Chemical Engineering
University of Notre Dame
Notre Dame, IN 46556 USA

3rd International Conference on
Foundations of Computer Aided Process Operations,
Snowbird, UT, July 5–10, 1998

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

Summary

- The objective function in nonlinear parameter estimation problems may have multiple local optima.

- Standard methods for parameter estimation are local methods that provide no guarantee that the global optimum, and thus the best set of model parameters, has been found.

- Interval analysis provides a **mathematically and computationally guaranteed** method for reliably solving parameter estimation problems, finding the **globally optimal** parameter values.

- This is demonstrated using example problems in the modeling of vapor-liquid equilbrium.
  - Published parameter values (e.g., DECHEMA VLE Data Collection) are often locally, not globally, optimal.
  - Globally optimal parameter values are easily found using the interval approach.

Background—Interval Analysis

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

- Basic interval arithmetic for $X = [a, b]$ and $Y = [c, d]$ is X op $Y = \{x \text{ op } y \mid x \in X,\ y \in Y\}$ where 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 $f(\mathbf{x})$ for $\mathbf{x} \in \mathbf{X}$. That is, $F(\mathbf{X}) \supseteq \{f(\mathbf{x}) \mid \mathbf{x} \in \mathbf{X}\}$.

- 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 $y_{\mu i} = f_i(\mathbf{x}_\mu, \boldsymbol{\theta})$ with independent variables $\mathbf{x}_\mu = (x_{\mu 1}, \ldots, x_{\mu m})^T$ and parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$.

- Various objective functions $\phi(\boldsymbol{\theta})$ can be used to determine the parameter values that provide the "best" fit, e.g.

  - Maximum likelihood
  - Relative least squares

  Relative least squares will be used here.

- Optimization problem to determine parameters can be formulated as either a constrained or unconstrained problem. In the unconstrained case, the experimental observations are substituted directly into the objective function. The unconstrained formulation is used here.

Parameter Estimation

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

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

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

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

- Interval Newton can be combined with branch and bound so that roots of $\mathbf{g}(\boldsymbol{\theta}) = \mathbf{0}$ that cannot be the global minimum need not be found.

Interval Newton Method

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

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

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

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

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

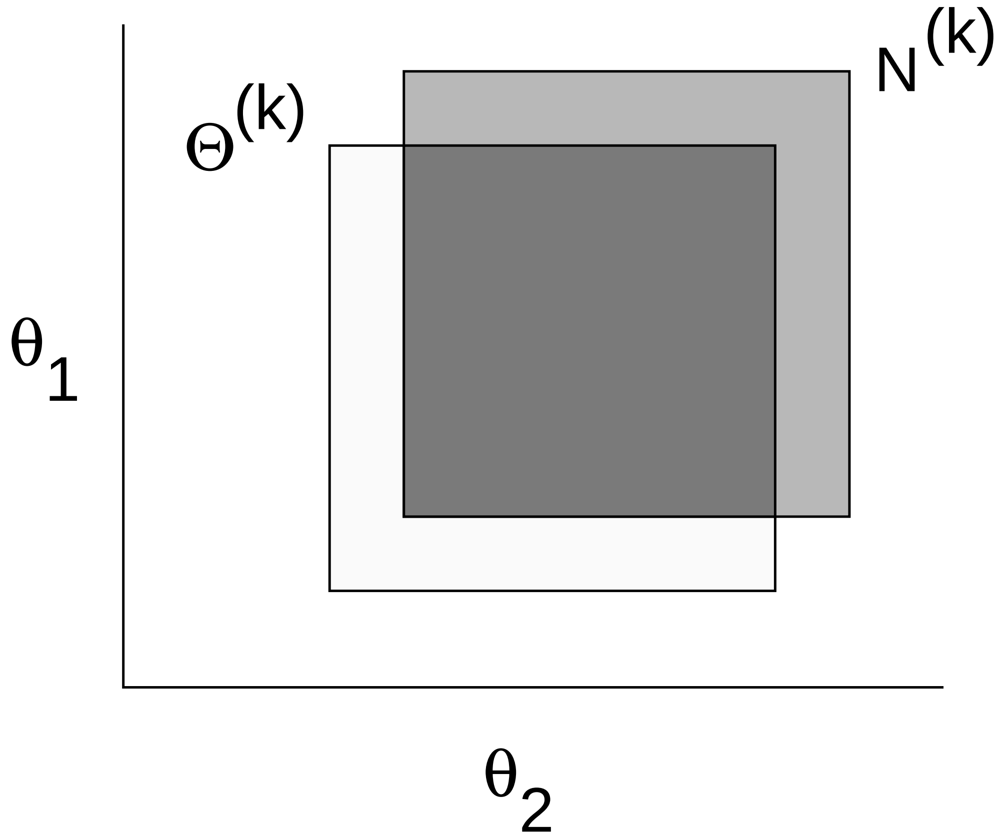
- It follows that if $\boldsymbol{\Theta}^{(k)} \cap \mathbf{N}^{(k)} = \emptyset$, then there is no root in $\boldsymbol{\Theta}^{(k)}$. This is also the conclusion if $0 \notin \mathbf{G}(\boldsymbol{\Theta}^{(k)})$.

Interval Newton Method (continued)

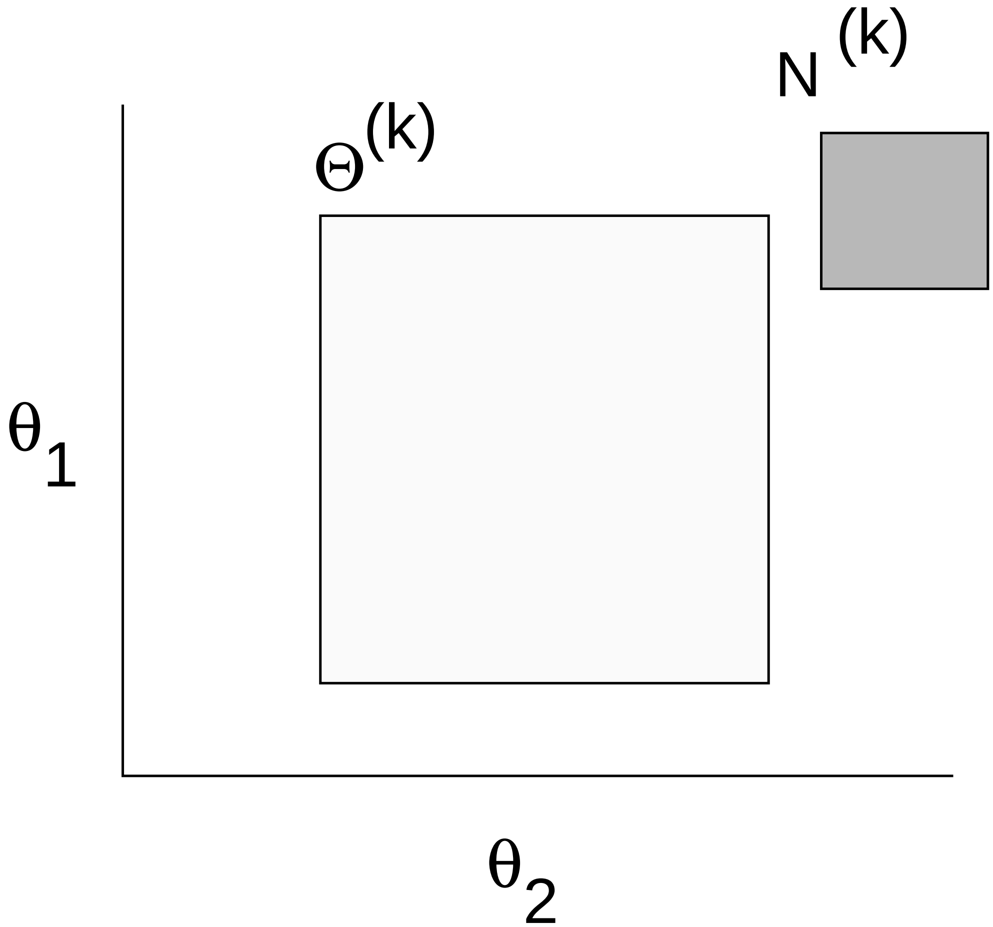
- Interval Newton provides an existence and uniqueness test: If $\mathbf{N}^{(k)} \subset \mathbf{\Theta}^{(k)}$, then:
  - There is a **unique** zero of $\mathrm{g}(\boldsymbol{\theta})$ in $\mathbf{\Theta}^{(k)}$.
  - The interval Newton iteration $\mathbf{\Theta}^{(k+1)} = \mathbf{\Theta}^{(k)} \cap \mathbf{N}^{(k)}$ 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 $\mathbf{\Theta}^{(k)}$.

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

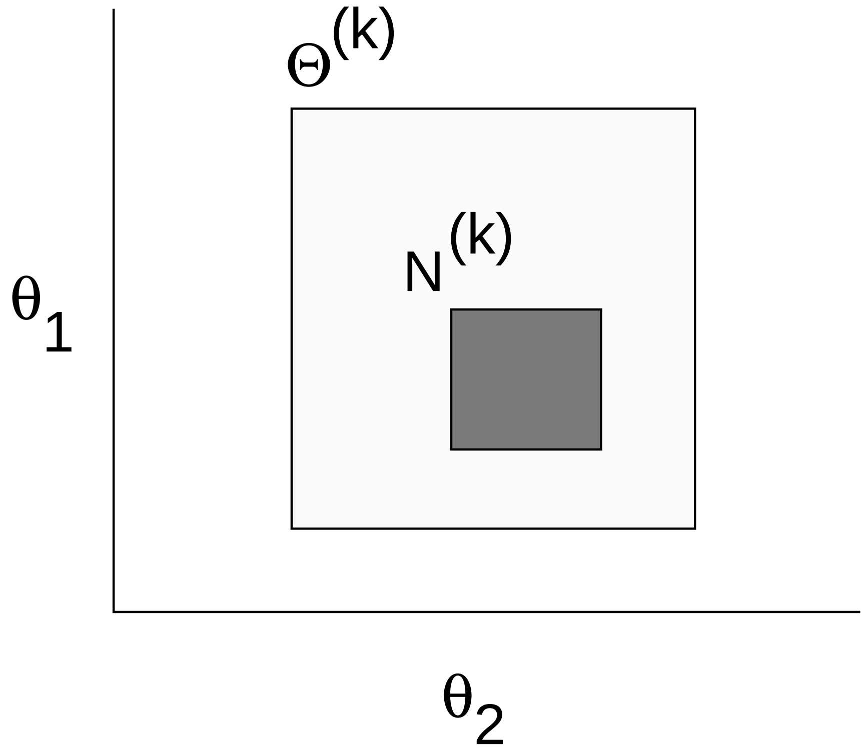
  This is the interval Newton/generalized bisection (IN/GB) approach.



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 $\mathrm{g}(\boldsymbol{\theta}) = \mathbf{0}$, 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 $\mathbf{N}^{(k)}$ (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, NRTL, UNIQUAC):

$$\gamma_{\mu i,\mathrm{calc}} = f_i(\mathbf{x}_\mu, \boldsymbol{\theta})$$

- The relative least squares objective is commonly used:

$$\phi(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{\mu=1}^{p} \left[\frac{\gamma_{\mu i,\mathrm{calc}}(\boldsymbol{\theta}) - \gamma_{\mu i,\mathrm{exp}}}{\gamma_{\mu i,\mathrm{exp}}} \right]^2 .$$

- Experimental values $\gamma_{\mu i,\mathrm{exp}}$ of the activity coefficients are obtained from VLE measurements at compositions $\mathbf{x}_\mu, \mu = 1, \ldots, p$.

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

Example Problems

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

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

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

- The energy parameters θ_1 and θ_2 must be estimated.

- Parameter estimation results for θ_1 and θ_2 are given in the DECHEMA Collection for all twelve problems.

Results

- Each problem was solved using the interval approach (INTBIS) to determine the globally optimal values of the θ_1 and θ_2 parameters.

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

- For each problem, the number of local minima in $\phi(\boldsymbol{\theta})$ was also determined.

- Table 1 presents a summary of these results and comparisons.

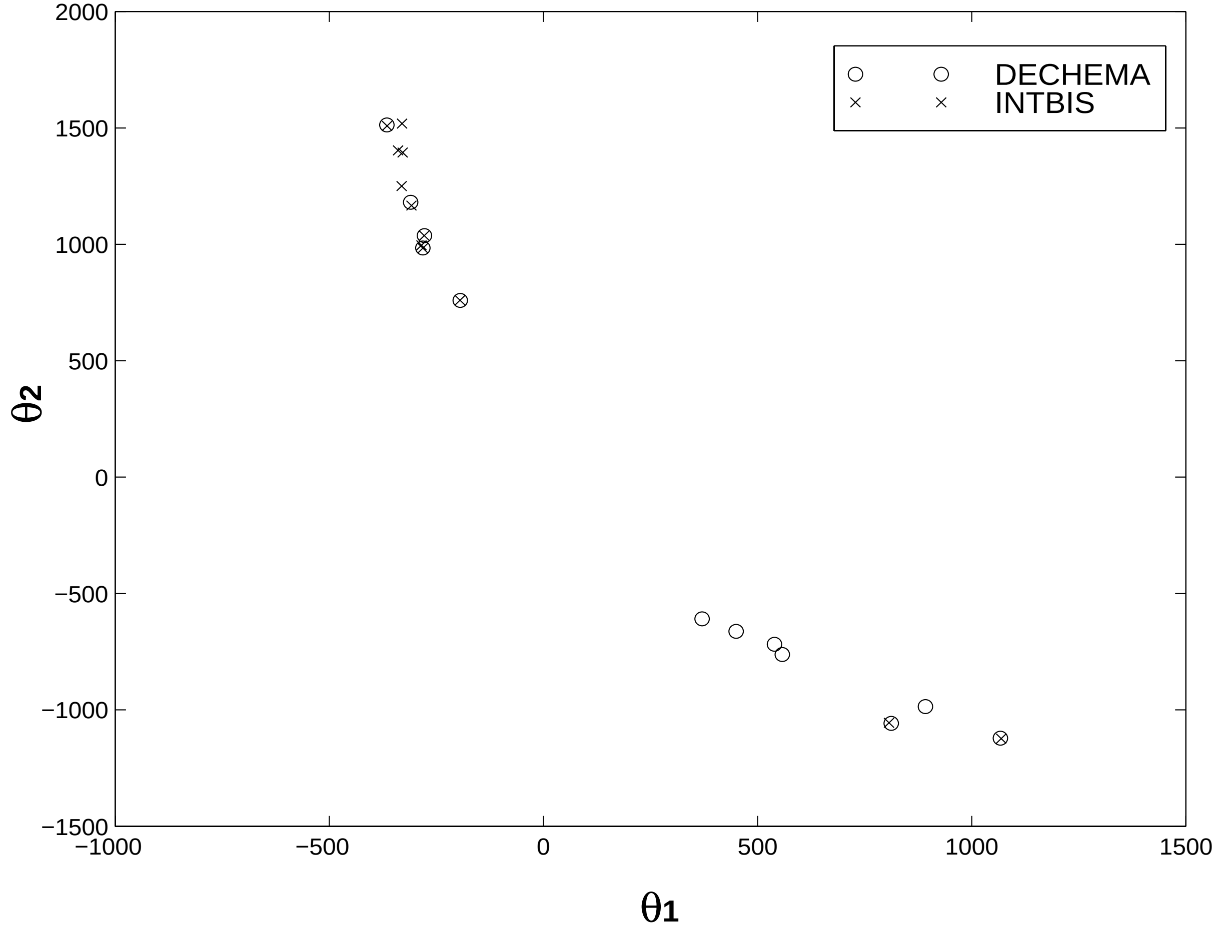
- Figure 1 shows schematically the parameter values found by INTBIS and in DECHEMA.

TABLE 1: INTBIS results vs. DECHEMA values

Data Set	P (mm Hg)	DECHEMA			INTBIS			No. of Minima
		θ_1	θ_2	$\phi(\boldsymbol{\theta})$	θ_1	θ_2	$\phi(\boldsymbol{\theta})$	
1	760	-195	759	0.0342	-168	759	0.0342	2
2	760	-278	1038	0.0106	-278	1038	0.0106	2
3	760	-310	1181	0.0151	-308	1167	0.0151	2
4	760	-282	985	0.353	-282	984	0.353	2
5	760	-366	1513	0.0257	-365	1509	0.0257	3
6	760	1067	-1122	0.0708	1065	-1120	0.0708	2
7*	200	892	-985	0.141	**-331**	**1250**	**0.0914***	2
8*	200	370	-608	0.0459	**-340**	**1404**	**0.0342***	3
9*	100	539	-718	0.165	**-285**	**996**	**0.111***	2
10*	100	450	-663	0.151	**-329**	**1394**	**0.0819***	3
11*	70	558	-762	0.0399	**-330**	**1519**	**0.0372***	3
12	25	812	-1058	0.0502	807	-1055	0.0502	2

*New globally optimal parameters found!

FIGURE 1: INTBIS results vs. DECHEMA values



Results (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 (most fall in the same cluster in Figure 1).

- 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.

- As is fairly common in this application, the stationary points lie in a steep-sided valley with a relatively flat bottom. This is shown schematically in the 3D plot of Figure 2 and the contour plot of Figure 3.

- The globally optimal parameters found using the interval approach provide a noticably better fit to the experimental data. This is shown by the relative deviation plots given in Figures 4 and 5.

TABLE 2: Stationary Points (Roots) for Data Set 10

Root	θ_1, θ_2	Eigenvalues of Hessian	$\phi(\boldsymbol{\theta})$	Status
P1	(1658, -1251)	7.55E-5, 2.58E-7	0.164	minimum
P2	(1165, -1083)	6.83E-5, -1.44E-7	0.178	saddle
P3	(452, -664)	6.97E-5, 9.42E-8	0.151	minimum
P4	(-37.8, 38.5)	9.08E-5, -3.54E-7	0.19	saddle
P5	(-329, 1394)	1.23E-4, 1.47E-7	0.0819	global minimum

FIGURE 2: Objective Function for Data Set 10 (3D Plot)

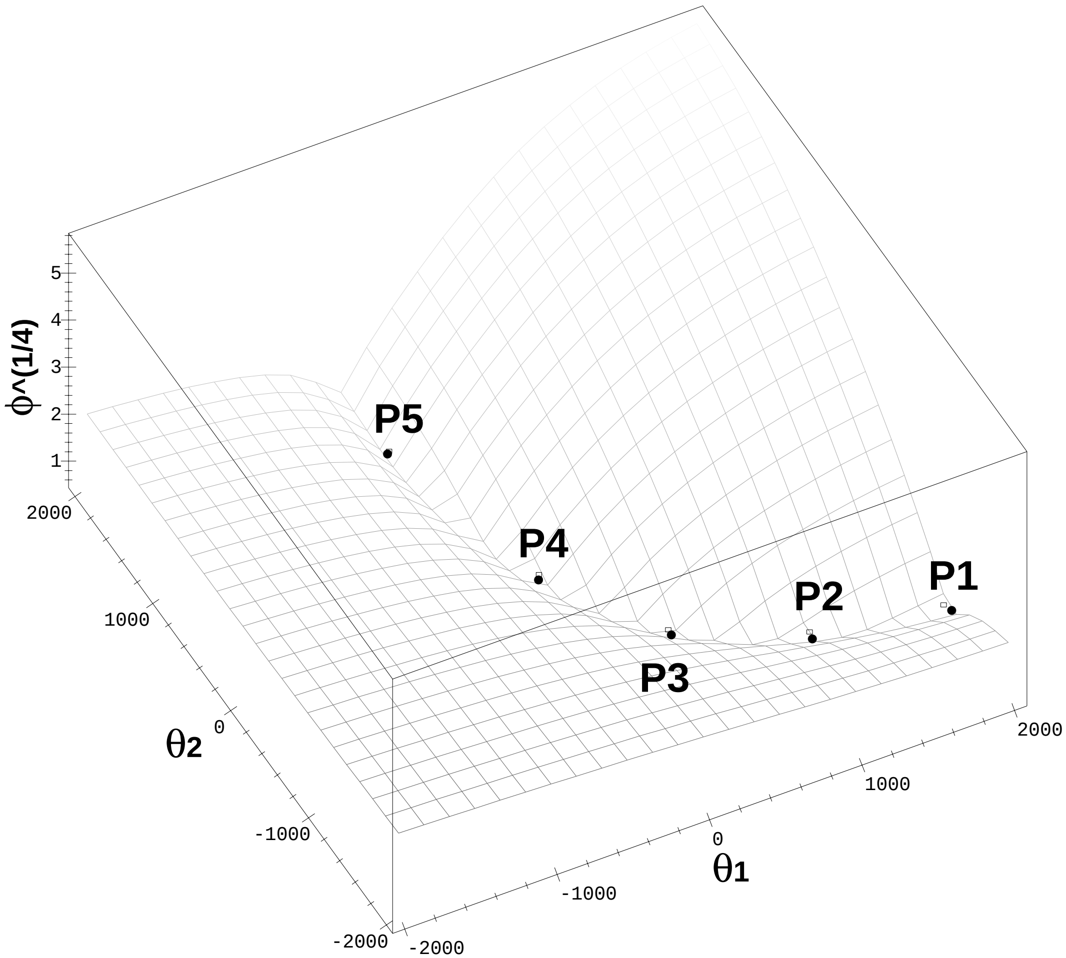
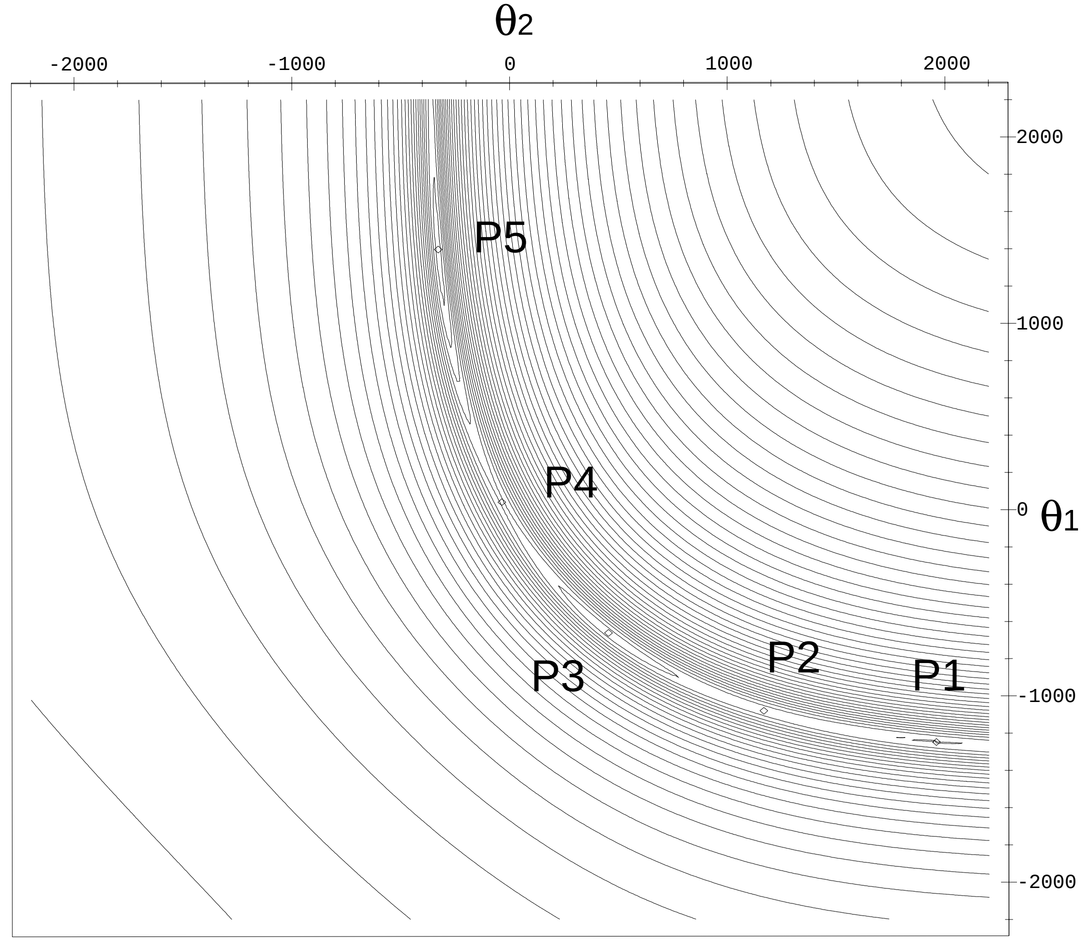


FIGURE 3: Objective Function for Data Set 10 (Contour Plot)



(contours of $\ln \phi$)

FIGURE 4: Comparison of Relative Deviation in γ_1

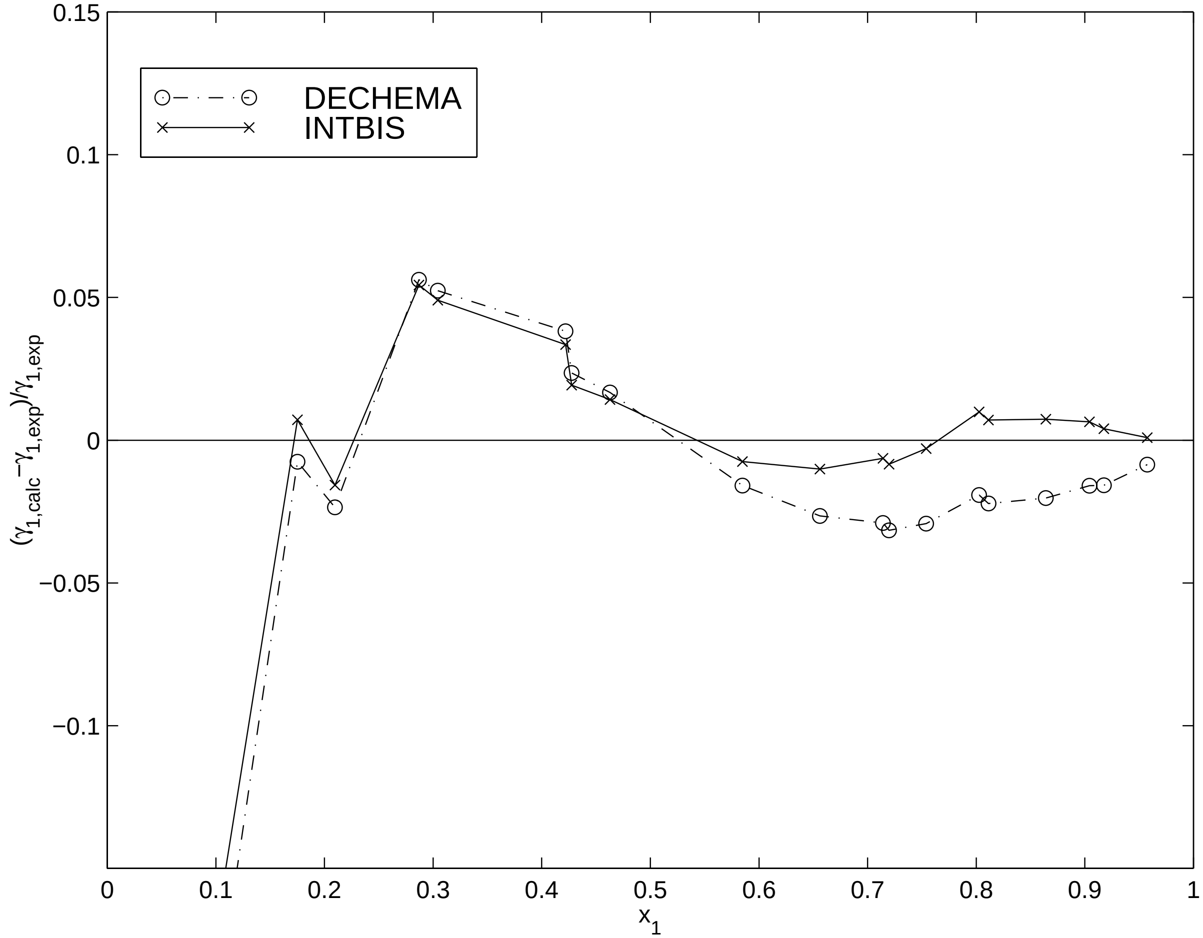
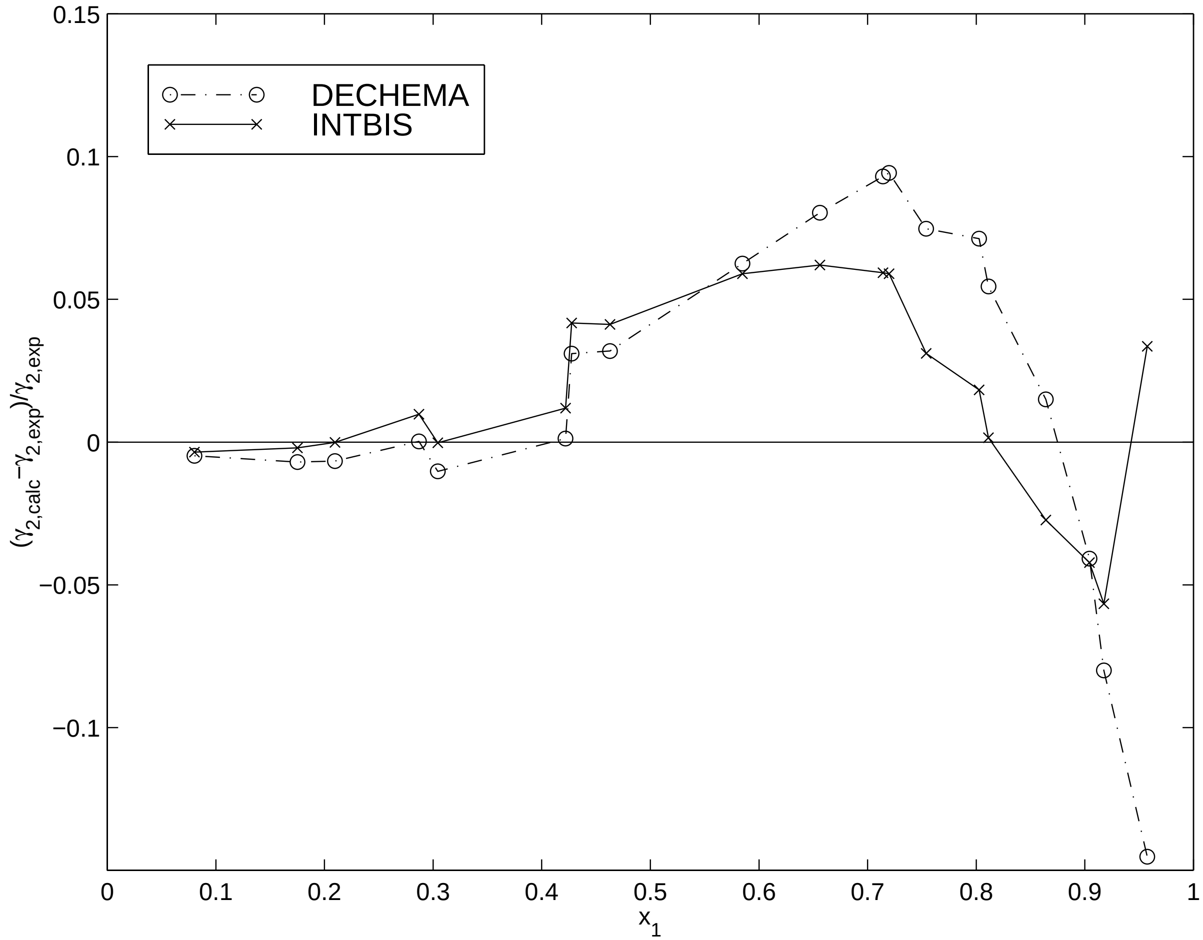


FIGURE 5: Comparison of Relative Deviation in γ_2



Computational Performance

- With initial parameter intervals of $\Theta_1^{(0)} = \Theta_2^{(0)} = [-10000, 10000]$, the computation time for the global optimization was roughly from 10 to 50 seconds on a Sun Ultra 2/1300 workstation.

- No significant efforts have been made to optimize the efficiency of the code. Tightening the evaluation of interval function extensions can potentially reduce the computation time by a order of magnitude.

- Because of the wide initial interval that can be used, as opposed to an initial point guess, the method is essentially **initialization independent**.

- The additional computation time for the interval approach, as opposed to local methods, may be well 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, 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.
  - 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
  - See also
    http://www.nd.edu/~markst