Deterministic global optimization for data reconciliation and parameter estimation using error-in-variables approach

Chao-Yang Gau and Mark A. Stadtherr
Department of Chemical Engineering
University of Notre Dame
182 Fitzpatrick Hall
Notre Dame, IN 46556 USA

Running Title: Deterministic global optimization for parameter estimation

*Current address: LINDO Systems, Inc., 1415 North Dayton Street, Chicago, IL 60622, USA

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

ABSTRACT

Standard techniques for solving the optimization problem arising in parameter estimation and data reconciliation by the error-in-variables (EIV) approach offer no guarantee that the global optimum has been found. However, an interval-Newton approach can provide a powerful, deterministic global optimization methodology for the reliable solution of EIV problems in chemical process modeling, offering mathematical and computational guarantees that the global optimum has been found. This methodology is demonstrated here by applying it to the modeling of vapor-liquid equilibrium. An issue of particular interest is the effect of changing the standard deviation values appearing in the objective function.

1. INTRODUCTION

Parameter estimation is a central problem in the development of mathematical models that represent the physical phenomena underlying chemical process operations, and is thus an important issue in process systems engineering. In the classical least-squares approach to parameter estimation, it is assumed that there are a set of independent variables not subject to measurement error. The error-in-variables (EIV) approach differs in that it is assumed that there are measurement errors in *all* variables. Accounting for error in all the variables has been demonstrated (e.g., Duever et al. [1]; Patino-Leal and Reilly [2]) to lead to unbiased estimates of the parameter values, and thus to more accurate models. Furthermore, in solving the EIV problem, not only are parameter estimation results obtained, but also data reconciliation results.

Consider the problem of estimating the parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_q)^{\mathrm{T}}$ in a model of the general form $\boldsymbol{f}(\boldsymbol{\theta}, \boldsymbol{z}) = \mathbf{0}$, where $\boldsymbol{z}$ is a vector of n state variables for the system to be modeled, and $\boldsymbol{f}$ is a vector of p model functions. When the EIV approach is used, the optimization problem that must be solved has the form

$$\min_{\boldsymbol{\theta}, \tilde{\boldsymbol{z}}_i} \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{\left(\tilde{z}_{ij} - z_{ij}\right)^2}{\sigma_j^2} \tag{1}$$

subject to

$$\boldsymbol{f}(\boldsymbol{\theta}, \tilde{\boldsymbol{z}}_i) = \mathbf{0}, \quad i = 1, \ldots, m. \tag{2}$$

Here $\boldsymbol{z}_i = (z_{i1}, ..., z_{in})^{\mathrm{T}}$ represents measurements of the state variables from $i = 1, \ldots, m$ experiments, $\tilde{\boldsymbol{z}}_i = (\tilde{z}_{i1}, ..., \tilde{z}_{in})^{\mathrm{T}}$ represents the unknown "true" values associated with each measurement, and σ_j represents the standard deviation associated with the measurement of state variable j. Details concerning the formulation of this optimization problem are available elsewhere (e.g., Kim et al. [3]; Esposito and Floudas [4]), as are good introductions to the general problem of nonlinear parameter estimation (e.g, Bard [5]; Gallant [6]; Seber [7]), and to applications in chemical engineering (Englezos and Kalogerakis [8]).

The presence of the true values $\tilde{\boldsymbol{z}}_i$, $i = 1, \ldots, m$, as variables in the optimization problem has a number of practical implications. First, as already noted, it means that, in solving the EIV problem, not only are the parameter estimation results $\boldsymbol{\theta}$ obtained, but also the data reconciliation results $\tilde{\boldsymbol{z}}_i$, $i = 1, \ldots, m$. However, this comes at the expense of a substantial increase in the dimensionality of the optimization problem, which at $nm + p$ is now a function of the number of experiments. Furthermore, since the optimization is over both $\boldsymbol{\theta}$ and $\tilde{\boldsymbol{z}}_i$, this is likely to be a nonlinear optimization problem even for models that are linear in the parameters. Thus, in general, the optimization problem is nonlinear and potentially nonconvex, indicating the need to be concerned about the possible existence of multiple local minima.

Various methods have been used to solve the optimization problem defined by Eqs. (1) and (2). These include gradient-based methods such as Gauss-Newton or Gauss-Marquardt (e.g., Britt and Luecke [9]; Fabries and Renon [10]; Anderson et al. [11]; Schwetlick and Tiller [12]; Valko and Vajda [13]), generalized reduced gradient (Kim et al. [3]), and successive quadratic programming (e.g., Tjoa and Biegler [14,15]), as well as direct search methods, such as the simplex pattern search (e.g. Vamos and Hass [16]). However, these are all local methods that offer no assurance that the global minimum in the optimization problem has been found. Towards finding the global optimum, one approach is to introduce a randomized element, either in the selection of multiple initial guesses (e.g., Vamos and Hass [16]), or in the search procedure itself (e.g., Luus and Hernaez [17]). These stochastic methods still provide no guarantee that the global optimum has been found. To obtain any such guarantee requires the use of *deterministic* global optimization procedures.

A powerful deterministic approach is that suggested by Esposito and Floudas [4], who reformulate the optimization problem in terms of convex underestimating functions and then use a branch-and-bound procedure. This method provides a mathematical guarantee of global optimality. One drawback to this approach is that in general it may be necessary to perform problem reformulations and develop convex underestimators specific to each new application. Also, in principle, branch-and-bound methods implemented in floating point arithmetic may be vulnerable to rounding error problems, and thus lose their mathematical guarantees. Another deterministic approach is that suggested recently by Gau and Stadtherr [18,19], who use an interval-Newton approach. This is a general-purpose methodology that provides a mathematical guarantee of global optimality, as well as a computational guarantee, since rounding issues are dealt with through the use of interval arithmetic. In their initial study, Gau and Stadtherr [18] demonstrated the potential of the interval method-

ology by applying it to some small problems (12 to 32 variables) and finding the approach to compare favorably to the Esposito and Floudas [4] method in terms of computational efficiency. In a later feasibility study (Gau and Stadtherr [19]), they applied the interval technique to a variety of larger problems, including a heat exchanger network problem involving 264 variables. These earlier studies focused primarily on the parameter estimation aspect of the problem and did not consider the data reconcilation results. The interval methodology is demonstrated here by applying it to the modeling of vapor-liquid equilibrium, with focus on the data reconcilation results. An issue of particular interest is the effect of changing the standard deviation values appearing in the objective function.

2. METHODOLOGY

For many practical problems, the p model equations can be easily solved algebraically for p of the n state variables. Thus, by substitution into the objective function, an unconstrained formulation of the optimization problem can be obtained. The unconstrained problem can be stated

$$\min_{\boldsymbol{\theta},\tilde{\boldsymbol{v}}_i} \ \phi(\boldsymbol{\theta},\tilde{\boldsymbol{v}}_i) \tag{3}$$

where $\tilde{\boldsymbol{v}}_i$, $i = 1,\ldots,m$, refers to the $n-p$ independent state variables not eliminated using the model equations, and $\phi(\boldsymbol{\theta},\tilde{\boldsymbol{v}}_i)$ is the objective function in Eq. (1) after the p dependent state variables have been eliminated by substitution. This unconstrained formulation of the problem will be used here. However, it should be noted that the accompanying reduction in the dimensionality of the problem does not necessarily make it any easier to solve, since the objective function in the reduced space of $\boldsymbol{\theta}$ and $\tilde{\boldsymbol{v}}_i$ may be a much more complicated function than the objective function in the original space of $\boldsymbol{\theta}$ and $\tilde{\boldsymbol{z}}_i$.

For the global minimization of $\phi(\boldsymbol{\theta},\tilde{\boldsymbol{v}}_i)$, an approach based on interval analysis is used. Good introductions to interval analysis, as well as interval arithmetic and computing with intervals, include those of Neumaier [20], Hansen [21] and Kearfott [22]. Of particular interest here is the interval-Newton technique. Given a nonlinear equation system with a finite number of real roots in some initial interval, this technique provides the capability to find (or, more precisely, to enclose within a very narrow interval) all the roots of the system within the given initial interval. To apply this technique to the optimization problem of interest here, it is used to seek stationary points; that is, to solve the nonlinear equation system

$$\boldsymbol{g}(\boldsymbol{y}) = \boldsymbol{g}(\boldsymbol{\theta},\tilde{\boldsymbol{v}}_i) \equiv \nabla\phi(\boldsymbol{\theta},\tilde{\boldsymbol{v}}_i) = \boldsymbol{0}, \tag{4}$$

where for convenience the vector of independent variables has been denoted $\boldsymbol{y} = (\boldsymbol{\theta},\tilde{\boldsymbol{v}}_i)^{\mathrm{T}}$. The global minimum will be a root of this nonlinear equation system, but there may be many other roots as well, representing local minima and maxima and saddle points. To identify the global minimum, one approach is to simply find all the stationary points and then identify the point with the minimum value of the objective function. Alternatively, by including an objective range test in the solution procedure, one can effectively combine the interval-Newton approach with an interval branch-and-bound technique, so that roots of $\boldsymbol{g}(\boldsymbol{y}) = \boldsymbol{0}$ that cannot be the global minimum need not be found. If the constrained formulation of the problem is used, then instead of applying interval-Newton to solve the stationarity conditions, it is applied to the solve the Kuhn-Tucker conditions (or, more generally, the Fritz-John conditions).

The solution algorithm is applied to a se-

quence of intervals, beginning with some user-specified initial interval $\boldsymbol{Y}^{(0)}$ that provides upper and lower bounds on the independent variables. (In this section, upper case quantities are intervals, and lower case quantities real numbers.) This initial interval can be chosen to be sufficiently large to enclose all physically feasible behavior. This is in contrast to conventional local solution methods in which an initial *point* is needed, often resulting in a highly initialization-dependent procedure. It is assumed here that the global optimum will occur at an interior stationary minimum of $\phi(\boldsymbol{y})$ and not at the boundaries of $\boldsymbol{Y}^{(0)}$. Since the estimator ϕ is derived based on a product of Gaussian distribution functions corresponding to each data point, this is a very reasonable assumption for regression problems of the type considered here. Details of the basic solution algorithm used are given by Gau and Stadtherr [18] and Schnepper and Stadtherr [23]. Only the key ideas of the methodology are presented here.

For an interval $\boldsymbol{Y}^{(k)}$ in the sequence, the first step in the solution algorithm is the *function range test*. Here an *interval extension* $\boldsymbol{G}(\boldsymbol{Y}^{(k)})$ of the function $\boldsymbol{g}(\boldsymbol{y})$ is calculated. An interval extension provides upper and lower bounds on the range of values that a function may have in a given interval. It is often computed by substituting the given interval into the function and then evaluating the function using interval arithmetic. The interval extension so determined is often wider than the actual range of function values, but it always includes the actual range. If there is any component of the interval extension $\boldsymbol{G}(\boldsymbol{Y}^{(k)})$ that does not contain zero, then we may discard the current interval $\boldsymbol{Y}^{(k)}$, since the range of the function does not include zero anywhere in this interval, and thus no solution of $\boldsymbol{g}(\boldsymbol{y}) = \boldsymbol{0}$ exists in this interval. We may then proceed to consider the next interval in the sequence, since the current interval cannot contain a stationary point of $\phi(\boldsymbol{y})$. Otherwise, if $\boldsymbol{0} \in \boldsymbol{G}(\boldsymbol{Y}^{(k)})$, then testing of $\boldsymbol{Y}^{(k)}$ continues.

The next step is the *objective range test*. The interval extension $\Phi(\boldsymbol{Y}^{(k)})$, which contains the range of $\phi(\boldsymbol{y})$ over $\boldsymbol{Y}^{(k)}$, is computed. If the lower bound of $\Phi(\boldsymbol{Y}^{(k)})$ is greater than a known upper bound on the global minimum of $\phi(\boldsymbol{y})$, then $\boldsymbol{Y}^{(k)}$ cannot contain the global minimum and need not be further tested. Otherwise, testing of $\boldsymbol{Y}^{(k)}$ continues.

The next step is the *interval-Newton test*. Here the linear interval equation system

$$G'(\boldsymbol{Y}^{(k)})(\boldsymbol{N}^{(k)} - \boldsymbol{y}^{(k)}) = -\boldsymbol{g}(\boldsymbol{y}^{(k)}) \qquad (5)$$

is set up and solved for a new interval $\boldsymbol{N}^{(k)}$, where $G'(\boldsymbol{Y}^{(k)})$ is an interval extension of the Jacobian of $\boldsymbol{g}(\boldsymbol{y})$, and $\boldsymbol{y}^{(k)}$ is a point in $\boldsymbol{Y}^{(k)}$, usually taken to be the midpoint. Comparison of the current interval $\boldsymbol{Y}^{(k)}$ and the *image* $\boldsymbol{N}^{(k)}$ provides a powerful existence and uniqueness test (Kearfott [22]). There are three possible outcomes of the comparison: 1. If $\boldsymbol{N}^{(k)}$ and $\boldsymbol{Y}^{(k)}$ have a null intersection, as shown schematically in Figure 1 for a two-variable problem, then this is mathematical proof that there is no solution of $\boldsymbol{g}(\boldsymbol{y}) = \boldsymbol{0}$ in $\boldsymbol{Y}^{(k)}$. 2. If $\boldsymbol{N}^{(k)}$ is a proper subset of $\boldsymbol{Y}^{(k)}$, as shown in Figure 2, then this is mathematical proof that there is a *unique* solution of $\boldsymbol{g}(\boldsymbol{y}) = \boldsymbol{0}$ in $\boldsymbol{Y}^{(k)}$. 3. If neither of the first two conditions is true, as shown in Figure 3, then no conclusions can be made about the number of solutions in the current interval. However, it is known (Kearfott [22]) that any solutions that do exist must lie in the intersection of $\boldsymbol{N}^{(k)}$ and $\boldsymbol{Y}^{(k)}$. If this intersection is sufficiently smaller than the current interval, one can proceed by reapplying the interval-Newton test to the intersection. Otherwise, the intersection is bisected,

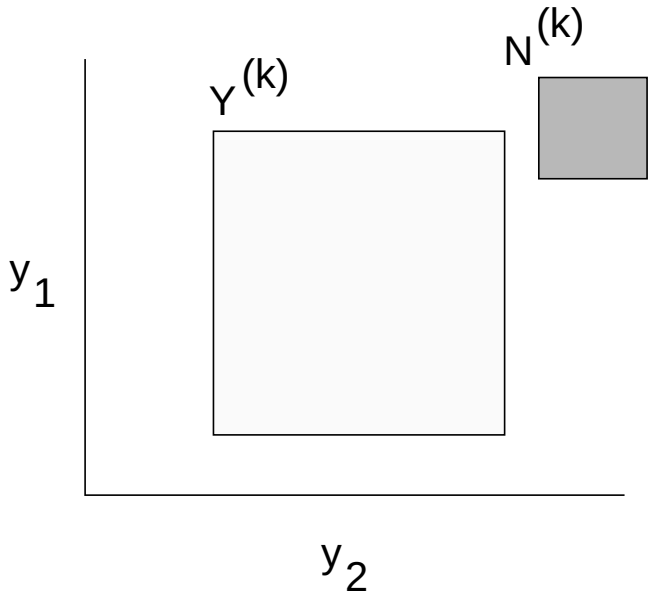

Figure 1. The computed image $\boldsymbol{N}^{(k)}$ has a null intersection with the current interval $\boldsymbol{Y}^{(k)}$. This is mathematical proof that there is no solution of the equation system in the current interval.

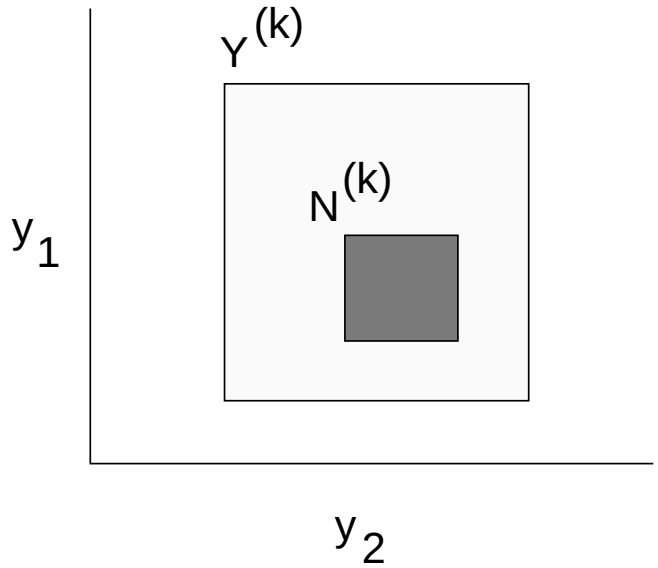

Figure 2. The computed image $\boldsymbol{N}^{(k)}$ is a subset of the current interval $\boldsymbol{Y}^{(k)}$. This is mathematical proof that there is a unique solution of the equation system in the current interval, and furthermore that this unique solution is also in the image.

and the resulting two intervals added to the sequence of intervals to be tested. These are the basic ideas of an interval-Newton/generalized-bisection (IN/GB) method. It should be noted that recent enhancements (Gau and Stadtherr [19]) to the methodology, involving the formulation and solution of the interval-Newton equation, play an important role in achieving computational efficiency on the problems considered here.

When properly implemented, this method provides a procedure that is mathematically *and* computationally guaranteed to find the global minimum of $\phi(\boldsymbol{y})$, or, if desired (by turning off the objective range test), to enclose *all* of its stationary points.

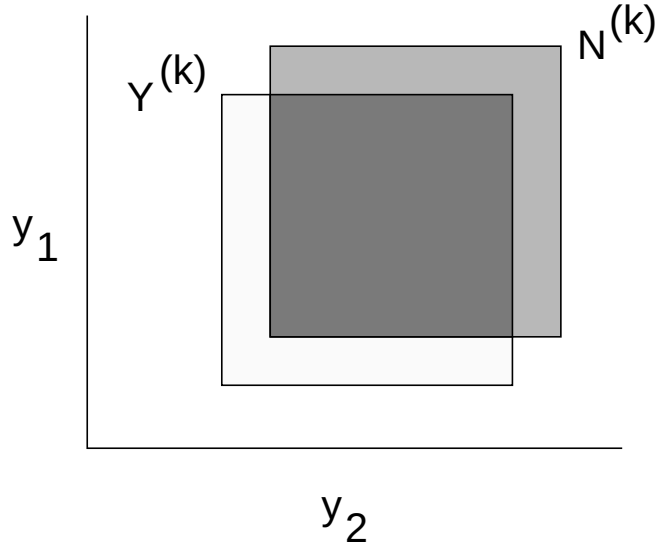

Figure 3. The computed image $\boldsymbol{N}^{(k)}$ has a nonnull intersection with the current interval $\boldsymbol{Y}^{(k)}$. Any solutions of the equation system must lie in the intersection of the image and the current interval.

3. VLE MODEL

We consider here the modeling of vapor-liquid equilibrium (VLE) using the Wilson equation for liquid-phase activity coefficient. Of interest is the binary system benzene(1)–hexafluorobenzene(2). Measurements are available (Gmehling et al. [24], part 7, p. 235) from $m = 16$ experiments for the state variable vector $\boldsymbol{z} = (x_1, y_1, P, T)^{\mathrm{T}}$, where P is the system pressure (mmHg), T is the system temperature (K), x_1 is the liquid-phase mole fraction of component 1, and y_1 is the vapor-phase mole fraction of component 1. Each experiment was performed at a measured pressure of $P = 500$ mmHg; the remaining measured values are given in Table 1. No information is

Table 1. Measurement data vector $\boldsymbol{z}_i$ for each experiment i.

i	$x_{1,i}$	$y_{1,i}$	P_i(mmHg)	T_i(°C)
1	0.0880	0.0840	500.0	67.95
2	0.1330	0.1285	500.0	68.03
3	0.1635	0.1600	500.0	68.08
4	0.1720	0.1690	500.0	68.08
5	0.2200	0.2195	500.0	68.10
6	0.2615	0.2660	500.0	68.08
7	0.3555	0.3680	500.0	67.78
8	0.3640	0.3770	500.0	67.75
9	0.4030	0.4185	500.0	67.73
10	0.5425	0.5640	500.0	67.30
11	0.6400	0.6600	500.0	67.13
12	0.7080	0.7215	500.0	67.00
13	0.8340	0.8350	500.0	66.73
14	0.8600	0.8595	500.0	66.66
15	0.9110	0.9080	500.0	66.73
16	0.9575	0.9545	500.0	66.68

available, either in Gmehling et al. [24] or in the original source of the measurement data, concerning the standard deviations associated with the measurements. A standard deviation vector of $\boldsymbol{\sigma} = (0.001, 0.01, 0.75, 0.1)^{\mathrm{T}}$ is thus assumed for now. That is, $\sigma_1 = \sigma_{x_1} = 0.001$, $\sigma_2 = \sigma_{y_1} = 0.01$, $\sigma_3 = \sigma_P = 0.75$, and $\sigma_4 = \sigma_T = 0.1$.

The model used to describe the VLE can be written as

$$P = \gamma_1 x_1 p_1^0(T) + \gamma_2(1 - x_1)p_2^0(T) \qquad (6)$$

$$y_1 = \frac{\gamma_1 x_1 p_1^0(T)}{\gamma_1 x_1 p_1^0(T) + \gamma_2(1 - x_1)p_2^0(T)}\,, \qquad (7)$$

where the pure component vapor pressures $p_1^0(T)$ and $p_2^0(T)$ are given by the Antoine relationships (with T in K and p in mmHg)

$$p_1^0(T) = \exp\left[15.8412 - \frac{2755.64}{T - 53.99}\right] \qquad (8)$$

$$p_2^0(T) = \exp\left[16.1940 - \frac{2827.54}{T - 57.66}\right] \qquad (9)$$

and the activity coefficients γ_1 and γ_2 are given by the Wilson equation

$$\ln\gamma_1 = -\ln(x_1 + \Lambda_{12}x_2) + x_2\left[\frac{\Lambda_{12}}{x_1 + \Lambda_{12}x_2} - \frac{\Lambda_{21}}{\Lambda_{21}x_1 + x_2}\right] \qquad (10)$$

$$\ln\gamma_2 = -\ln(x_2 + \Lambda_{21}x_1) - x_1\left[\frac{\Lambda_{12}}{x_1 + \Lambda_{12}x_2} - \frac{\Lambda_{21}}{\Lambda_{21}x_1 + x_2}\right]. \qquad (11)$$

Here the binary parameters Λ_{12} and Λ_{21} are given by

$$\Lambda_{12} = \frac{v_2}{v_1}\exp\left[-\frac{\theta_1}{RT}\right] \qquad (12)$$

$$\Lambda_{21} = \frac{v_1}{v_2}\exp\left[-\frac{\theta_2}{RT}\right] \qquad (13)$$

where $v_1 = 89.41$ cm^3/mol and $v_2 = 115.79$ cm^3/mol are the pure component liquid molar volumes, and θ_1 and θ_2 are the energy parameters that need to be estimated.

In order to formulate the EIV parameter estimation problem as an unconstrained optimization problem, the model, Eqs. (6)–(7), is used to eliminate P and y_1 in the objective function. Consequently, the vector of independent state variables is $\boldsymbol{v} = (x_1, T)^{\mathrm{T}}$. In the unconstrained optimization problem, the independent variables are $\boldsymbol{\theta} = (\theta_1, \theta_2)^{\mathrm{T}}$ and $\tilde{\boldsymbol{v}}_i$, $i = 1, \ldots, 16$ (16 vectors of two variables each), for a total of $2(16) + 2 = 34$ independent variables. The initial intervals on the parameters θ_1 and θ_2 were both taken as $[-10000, 200000]$ cal/mol, which covers the range of physical interest as described by Gau and Stadtherr [25]. The initial intervals on the independent state variables were chosen using plus and minus three standard deviations; that is, $\tilde{x}_{1,i} \in [x_{1,i} - 3\sigma_1, x_{1,i} + 3\sigma_1]$ and $\tilde{T}_i \in [T_i - 3\sigma_4, T_i + 3\sigma_4]$ for $i = 1, \ldots, 16$.

Table 2. Globally optimal data reconciliation results $\tilde{\boldsymbol{z}}_i$ for each experiment i when $\boldsymbol{\sigma} = (0.001, 0.01, 0.75, 0.1)^{\mathrm{T}}$.

i	$\tilde{x}_{1,i}$	$\tilde{y}_{1,i}$	$\tilde{P}_i$(mmHg)	$\tilde{T}_i$(°C)
1	0.087979	0.086223	500.063	67.9310
2	0.132970	0.131651	500.221	67.9619
3	0.163471	0.162903	500.338	67.9756
4	0.171973	0.171674	500.336	67.9762
5	0.219974	0.221641	500.406	67.9748
6	0.261499	0.265390	500.403	67.9561
7	0.355532	0.365645	499.836	67.8326
8	0.364034	0.374752	499.786	67.8180
9	0.403027	0.416519	499.897	67.7638
10	0.542534	0.564259	499.561	67.4374
11	0.639981	0.663301	499.854	67.1768
12	0.707925	0.728999	500.053	66.9859
13	0.833933	0.842084	500.142	66.6883
14	0.859959	0.864268	500.037	66.6499
15	0.911008	0.907830	500.238	66.6585
16	0.957511	0.950549	499.783	66.7455

4. RESULTS AND DISCUSSION

Using the interval methodology discussed above, with the standard deviation vector $\boldsymbol{\sigma} = (0.001, 0.01, 0.75, 0.1)^{\mathrm{T}}$, the globally optimal parameter values obtained were $\theta_1 = -443.616$ cal/mol and $\theta_2 = 1090.493$ cal/mol, with an objective function value of $\phi = 13.768$, and with the globally optimal values of the $\tilde{\boldsymbol{z}}_i$ as given in Table 2. Here the values of the independent variables $\tilde{x}_{1,i}$ and $\tilde{T}_i$ were determined during the optimization procedure, and the values of the dependent variables $\tilde{y}_{1,i}$ and $\tilde{P}_i$ were computed from the model, Eqs. (6)–(7).

The CPU time required for this 34-variable global optimization problem was 234 seconds on a Sun UltraServer2/2200 workstation (one processor). Given the size of the problem, and the fact that a rigorous global minimum was obtained, this degree of computational effort is quite small. By turning off the objective range test, thus allowing the technique to enclose *all* the stationary points, not just the global minimum, it was also ascertained that, for this problem, there was only one stationary point (the global minimum) in the specified initial interval.

We now consider the effect of the choice of the standard deviation values on the results. To do this, we focus here on the distribution of error residuals $\Delta z_{ij} = \tilde{z}_{ij} - z_{ij}$. A more rigorous quantitative statistical analysis of the results presented here is also available (Gau [26]). It is well known that a good data reconciliation possessing statistical significance should qualitatively have the same type of residual distributions for all of the measured variables. In particular, it is desired that all residuals be well spread within their respective standard deviations, with relatively few residuals exceeding $\pm\sigma_j$. This is indicative that errors in each variable have received roughly equivalent weighting in the optimization problem. To visualize this, weighted residual plots $\Delta z_{ij}/\sigma_j$ are shown in Figs. 4–7. For example, Figure 4 shows the distribution of values of $\Delta x_{1,i} = \tilde{x}_{1,i} - x_{1,i}$ relative to σ_{x_1}. Note that the standard deviation is indicated by the dashed horizontal lines at ± 1.

Examining Figs. 4–7 shows that the four error residual distributions are not similar. In particular, the errors determined for x_1 are much smaller than the assumed standard deviation for x_1. This indicates that in the optimization too much weight was put on reducing the error in x_1, relative to the other variables. Use of a larger value of σ_{x_1} will reduce the weight in the objective function of the term involving errors in x_1, and thus should lead to a better data reconcilation. Following this type of argument, and adjusting each standard deviation in turn until similarly good error distributions are seen for each variable, the standard deviation vec-

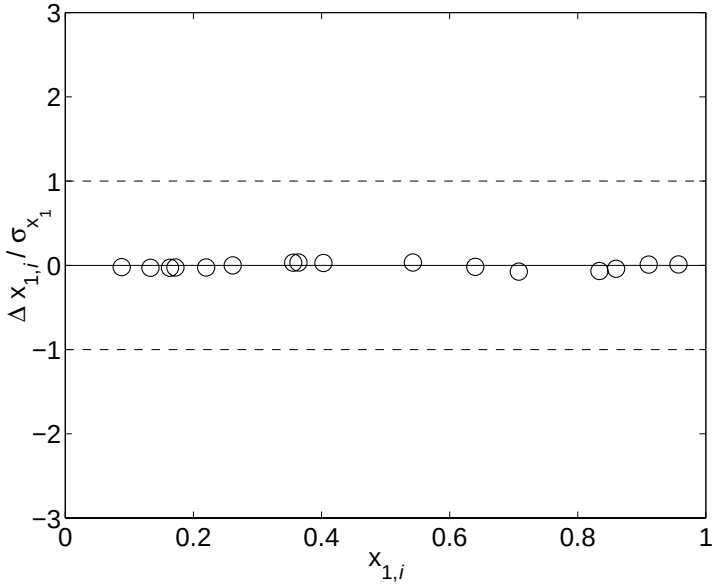

Figure 4. Globally optimal error residuals in $\Delta x_{1,i} = \tilde{x}_{1,i} - x_{1,i}$ relative to σ_{x_1} when $\boldsymbol{\sigma} = (0.001, 0.01, 0.75, 0.1)^{\mathrm{T}}$.

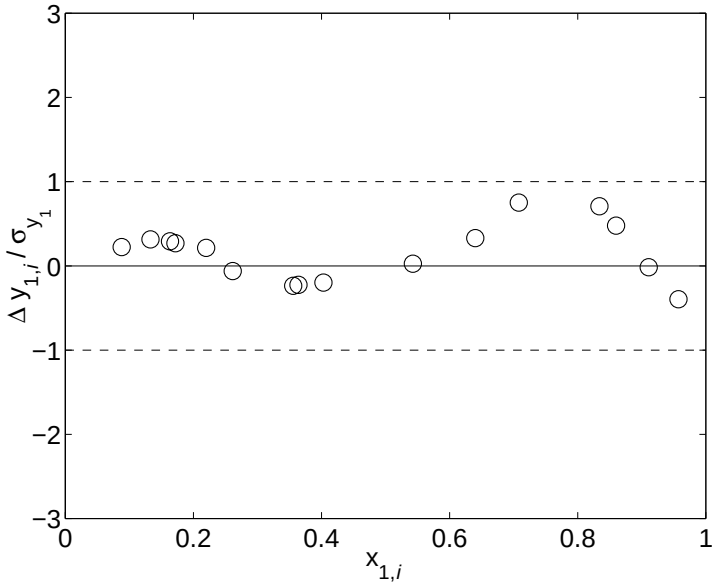

Figure 5. Globally optimal error residuals in $\Delta y_{1,i} = \tilde{y}_{1,i} - y_{1,i}$ relative to σ_{y_1} when $\boldsymbol{\sigma} = (0.001, 0.01, 0.75, 0.1)^{\mathrm{T}}$.

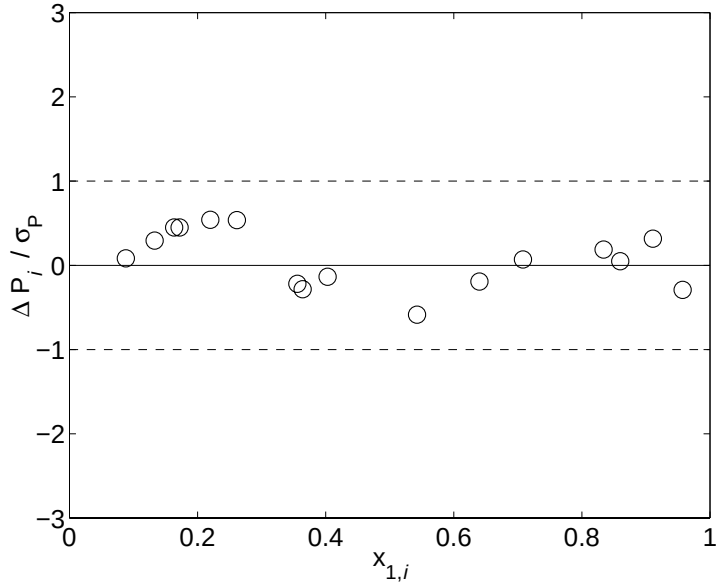

Figure 6. Globally optimal error residuals in $\Delta P_i = \tilde{P}_i - P_i$ relative to σ_P when $\boldsymbol{\sigma} = (0.001, 0.01, 0.75, 0.1)^{\mathrm{T}}$.

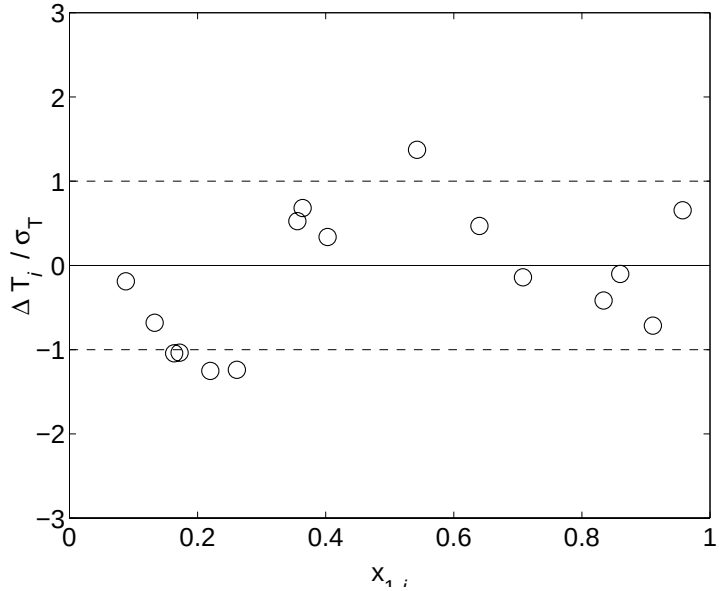

Figure 7. Globally optimal error residuals in $\Delta T_i = \tilde{T}_i - T_i$ relative to σ_T when $\boldsymbol{\sigma} = (0.001, 0.01, 0.75, 0.1)^{\mathrm{T}}$.

tor $\boldsymbol{\sigma} = (0.0025, 0.0025, 1.2, 0.07)^{\mathrm{T}}$ appears to provide reasonable results.

With the new standard deviations $\boldsymbol{\sigma} = (0.0025, 0.0025, 1.2, 0.07)^{\mathrm{T}}$, the globally optimal parameter values obtained from the interval methodology were $\theta_1 = -431.882$ cal/mol and $\theta_2 = 1038.214$ cal/mol, with an objective function value of $\phi = 30.755$, and with the globally optimal values of the $\tilde{\boldsymbol{z}}_i$ as given in Table 3.

For this 34-variable global optimization problem, the CPU time required was 1037 seconds on a Sun UltraServer2/2200 workstation (one processor). By turning off the objective range test, it was again determined that, for this problem, there was only one stationary point (the global minimum) in the specified initial interval. The increase in CPU time caused by changing the standard deviation values can be

Table 3. Globally optimal data reconciliation results $\tilde{\boldsymbol{z}}_i$ for each experiment i when $\boldsymbol{\sigma} = (0.0025, 0.0025, 1.2, 0.07)^{\mathrm{T}}$.

i	$\tilde{x}_{1,i}$	$\tilde{y}_{1,i}$	$\tilde{P}_i$(mmHg)	$\tilde{T}_i$(°C)
1	0.086729	0.085283	500.294	67.9326
2	0.131223	0.130262	500.833	67.9805
3	0.161832	0.161628	501.223	68.0072
4	0.170444	0.170510	501.222	68.0073
5	0.218706	0.220704	501.469	68.0126
6	0.261610	0.265832	501.481	67.9919
7	0.356678	0.366926	499.726	67.7963
8	0.365150	0.375970	499.569	67.7756
9	0.404081	0.417499	499.915	67.7351
10	0.542899	0.563841	498.773	67.3725
11	0.638978	0.661130	499.552	67.1571
12	0.704858	0.724819	500.030	66.9998
13	0.830847	0.838617	500.156	66.7222
14	0.857900	0.861952	499.823	66.6711
15	0.911077	0.907944	500.432	66.7051
16	0.959050	0.952662	499.017	66.7363

understood by looking at the objective function in use, namely

$$\phi(\boldsymbol{\theta}, \tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{T}}) = \sum_{i=1}^{m} \left[\frac{(\tilde{x}_{1,i} - x_{1,i})^2}{\sigma_{x_1}^2} + \frac{(\tilde{T}_i - T_i)^2}{\sigma_T^2} + \frac{(\tilde{y}_{1,i}(\boldsymbol{\theta}, \tilde{x}_{1,i}, \tilde{T}_i) - y_{1,i})^2}{\sigma_{y_1}^2} + \frac{(\tilde{P}_i(\boldsymbol{\theta}, \tilde{x}_{1,i}, \tilde{T}_i) - P_i)^2}{\sigma_P^2} \right]. \quad (14)$$

Note that the first two terms in the summation, corresponding to the errors in the independent variables, make contributions that are quadratic and convex in terms of the independent variables. However, the remaining two terms, corresponding to the errors in the dependent variables, make contributions that may be highly nonlinear and nonconvex in terms of the independent variables. Since the standard deviations are in effect weighting factors on each term in the summation, changing the standard deviations can change the character of the objective function considerably, thus affecting the difficulty of solving the problem.

The error residual distributions corresponding to the data reconciliation results in Table 3 are shown in Figs. 8–11. Note that with the adjusted values of the standard deviations the error distributions for each variable are now qualitatively similar, in that each case the residuals are well spread within their respective standard deviations, with relatively few residuals exceeding $\pm\sigma_j$. Thus, these values of the standard deviations appear to be acceptable. This can be also verified through a quantitative statistical analysis (Gau [26]). There may in fact be a range of standard deviation values that will yield an acceptable data reconcilation. To see this, we will consider one additional set of standard deviation values.

Consider now a slightly different standard deviation vector, namely $\boldsymbol{\sigma} = (0.003, 0.0029, 1.7, 0.083)^{\mathrm{T}}$. For this case, the globally optimal parameter values obtained from the interval methodology were $\theta_1 = -429.801$ cal/mol and $\theta_2 = 1029.207$ cal/mol, with an objective function value of $\phi = 19.999$, and with the globally optimal values of the $\tilde{\boldsymbol{z}}_i$ as given in Table 4. For this 34-variable global optimization problem, the CPU time required was 3535 seconds on a Sun UltraServer2/2200 workstation (one processor). By turning off the objective range test, it was determined that, in this case, there are actually *two* local minima (including the global minimum) in the specified initial interval. In addition to the local and global minimum given above, there is also a local, but not global, minimum at $\theta_1 = 278.518$ cal/mol and $\theta_2 = -279.408$ cal/mol, with an objective function value of $\phi = 161.3$, and with the values of the $\tilde{\boldsymbol{z}}_i$ as given in Table 5. This stationary point was characterized as a local

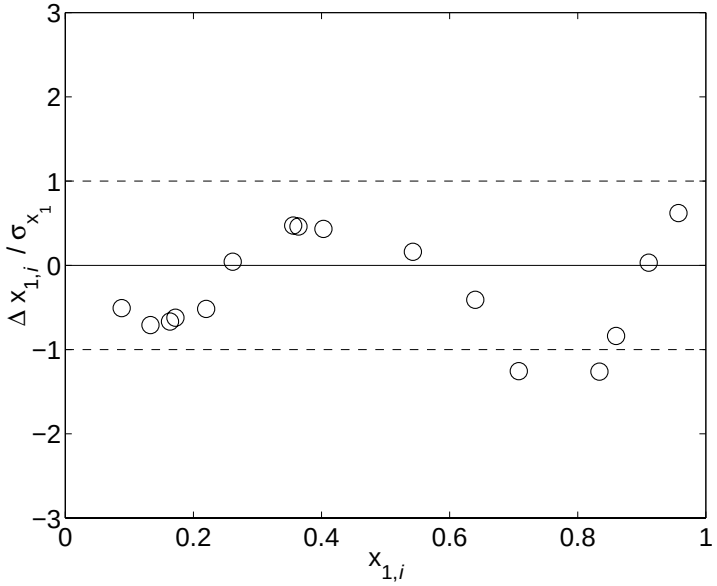

Figure 8. Globally optimal error residuals in $\Delta x_{1,i} = \tilde{x}_{1,i} - x_{1,i}$ relative to σ_{x_1} when $\boldsymbol{\sigma} = (0.0025, 0.0025, 1.2, 0.07)^{\mathrm{T}}$.

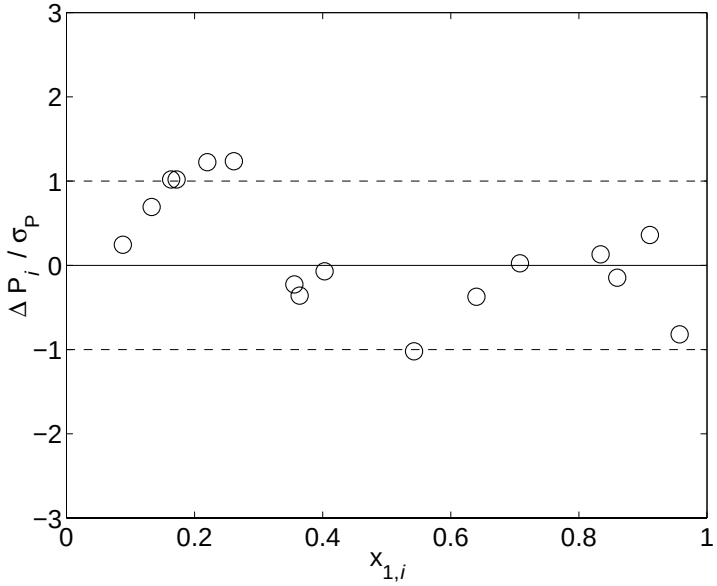

Figure 10. Globally optimal error residuals in $\Delta P_i = \tilde{P}_i - P_i$ relative to σ_P when $\boldsymbol{\sigma} = (0.0025, 0.0025, 1.2, 0.07)^{\mathrm{T}}$.

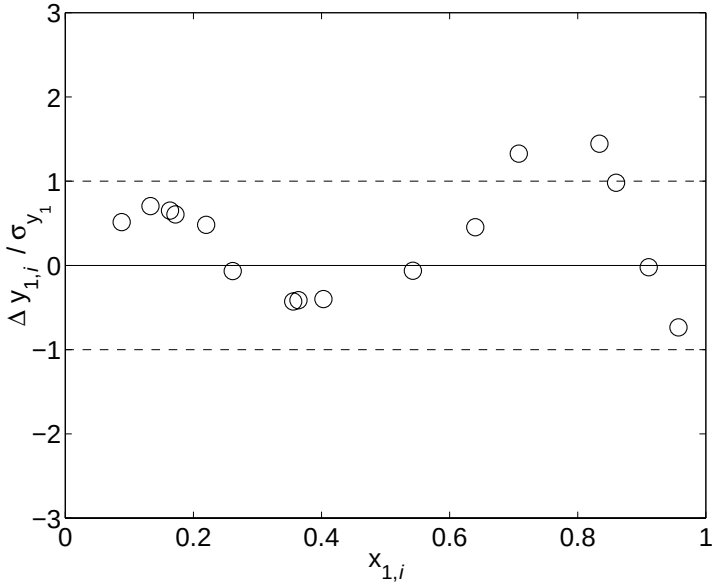

Figure 9. Globally optimal error residuals in $\Delta y_{1,i} = \tilde{y}_{1,i} - y_{1,i}$ relative to σ_{y_1} when $\boldsymbol{\sigma} = (0.0025, 0.0025, 1.2, 0.07)^{\mathrm{T}}$.

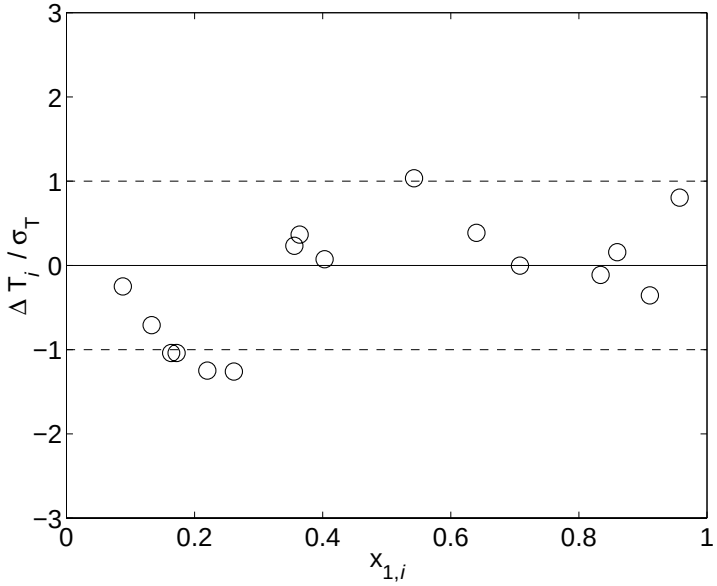

Figure 11. Globally optimal error residuals in $\Delta T_i = \tilde{T}_i - T_i$ relative to σ_T when $\boldsymbol{\sigma} = (0.0025, 0.0025, 1.2, 0.07)^{\mathrm{T}}$.

minimum by using MATLAB to compute the eigenvalues of the Hessian matrix of $\phi(\boldsymbol{\theta}, \tilde{\boldsymbol{v}}_i)$.

The presence of multiple local minima emphasizes the importance of using a deterministic global optimization approach to solve the EIV parameter estimation problem. As noted by Gau and Stadtherr [25], convergence to a local, but not global, optimum in parameter estimation problems for VLE models is not an uncommon occurrence. This can lead to the dismissal of a model as inadequate, when in fact the model may be fine, provided that the parameter estimation problem is solved correctly to a global optimum.

The error residual distributions for the global optimum are shown in Figs. 12–15. Again, the error distributions for each variable are qualitatively similar; in each case the

Table 4. Globally optimal data reconciliation results $\tilde{\boldsymbol{z}}_i$ for each experiment i when $\boldsymbol{\sigma} = (0.003, 0.0029, 1.7, 0.083)^{\mathrm{T}}$.

i	$\tilde{x}_{1,i}$	$\tilde{y}_{1,i}$	$\tilde{P}_i$(mmHg)	$\tilde{T}_i$(°C)
1	0.086657	0.085264	500.366	67.9348
2	0.131128	0.130231	501.010	67.9879
3	0.161741	0.161602	501.476	68.0184
4	0.170357	0.170489	501.476	68.0184
5	0.218635	0.220693	501.773	68.0261
6	0.261593	0.265865	501.791	68.0053
7	0.356709	0.366963	499.724	67.7915
8	0.365180	0.376001	499.539	67.7691
9	0.404128	0.417518	499.945	67.7322
10	0.542957	0.563761	498.585	67.3585
11	0.639037	0.660984	499.479	67.1520
12	0.704860	0.724601	500.021	67.0006
13	0.830794	0.838437	500.131	66.7259
14	0.857862	0.861829	499.735	66.6715
15	0.911067	0.907950	500.446	66.7120
16	0.959080	0.952793	498.795	66.7283

Table 5. Locally (but not globally) optimal data reconciliation results $\tilde{\boldsymbol{z}}_i$ for each experiment i when $\boldsymbol{\sigma} = (0.003, 0.0029, 1.7, 0.083)^{\mathrm{T}}$.

i	$\tilde{x}_{1,i}$	$\tilde{y}_{1,i}$	$\tilde{P}_i$(mmHg)	$\tilde{T}_i$(°C)
1	0.085316	0.086444	501.364	67.8931
2	0.129706	0.131479	502.370	67.9311
3	0.160478	0.162710	503.022	67.9539
4	0.169168	0.171533	503.064	67.9522
5	0.218073	0.221181	503.518	67.9533
6	0.261829	0.265597	503.554	67.9316
7	0.359288	0.364468	501.119	67.7326
8	0.367991	0.373288	500.873	67.7127
9	0.408040	0.413808	500.942	67.6895
10	0.550087	0.557028	497.678	67.3922
11	0.646782	0.653782	496.801	67.2576
12	0.711626	0.718265	496.102	67.1564
13	0.832122	0.837074	494.598	66.9469
14	0.857651	0.862058	494.162	66.8939
15	0.908019	0.911121	495.411	66.9138
16	0.955278	0.956910	495.432	66.8623

residuals are well spread within their respective standard deviations, with relatively few residuals exceeding $\pm\sigma_j$. Thus, assuming that one found the global optimum, one can conclude that these values of the standard deviations appear to be acceptable. This can be also verified through a quantitative statistical analysis (Gau [26]). Note that this global optimum is consistent with the global optimum found using the previous standard deviation vector, both in terms of the optimal parameter values and the data reconciliation results. However, the results at the local, but not global, optimum are quite different.

The error residual distributions for the local but not global optimum are shown in Figs. 16–19. These show that for *all* the variables there are a number of errors that are quite large relative to the standard deviation. Note that the error residual bounds of plus and minus three standard deviations imposed in selecting the initial search interval apply only to the independent variables x_1 and T, as larger errors can be observed in the dependent variable P. Based on this local optimum, one might well conclude that the model being used was inadequate. Of course, this conclusion would be incorrect, since as seen from the globally optimal solution, this model is capable of providing a good fit to the data. In evaluating the ability of a given model to fit a set of data, it is extremely important that the parameter estimation problem be correctly solved, to a global and not just local, optimum.

5. CONCLUDING REMARKS

We have demonstrated here that the interval-Newton approach is a powerful, deterministic global optimization methodology for the reliable solution of EIV parameter estimation problems. The approach provides both mathematical and computational *guarantees* that the *global* optimum in the parameter estimation problem has been found. Though we have concentrated here on a VLE modeling problem, the

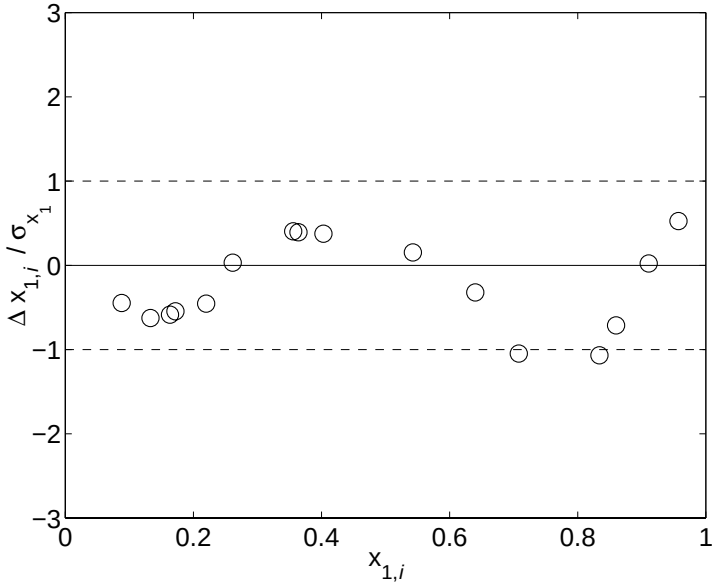

Figure 12. Globally optimal error residuals in $\Delta x_{1,i} = \tilde{x}_{1,i} - x_{1,i}$ relative to σ_{x_1} when $\boldsymbol{\sigma} = (0.003, 0.0029, 1.7, 0.083)^{\mathrm{T}}$.

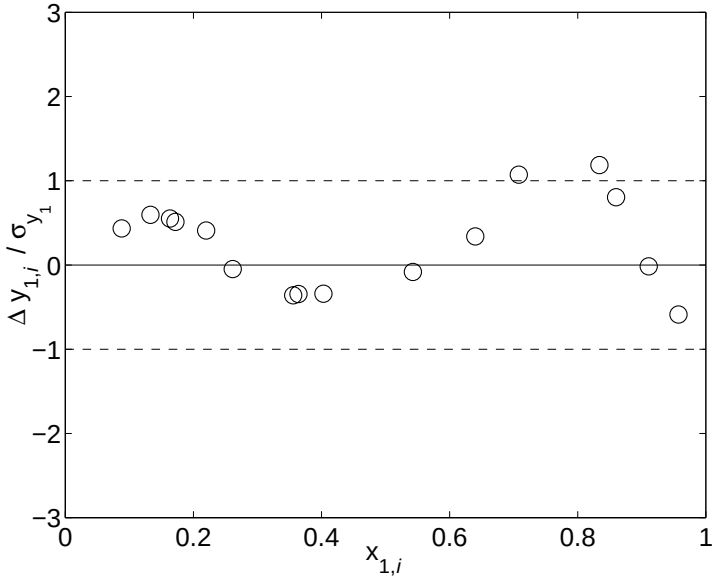

Figure 13. Globally optimal error residuals in $\Delta y_{1,i} = \tilde{y}_{1,i} - y_{1,i}$ relative to σ_{y_1} when $\boldsymbol{\sigma} = (0.003, 0.0029, 1.7, 0.083)^{\mathrm{T}}$.

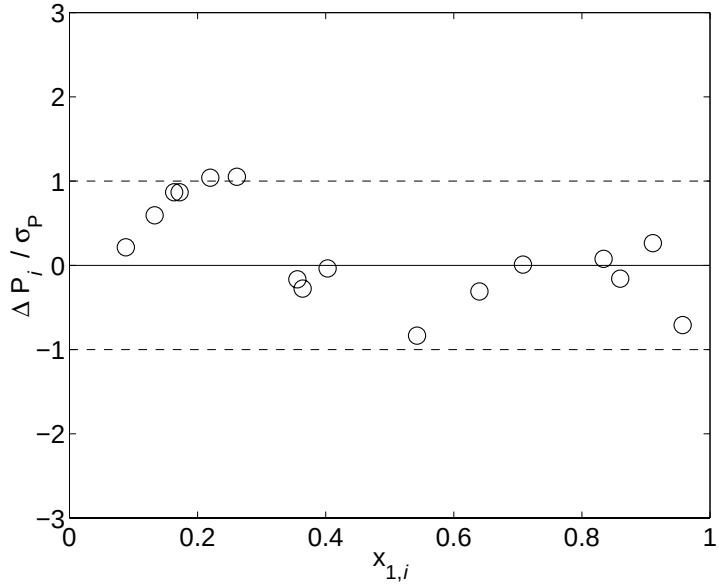

Figure 14. Globally optimal error residuals in $\Delta P_i = \tilde{P}_i - P_i$ relative to σ_P when $\boldsymbol{\sigma} = (0.003, 0.0029, 1.7, 0.083)^{\mathrm{T}}$.

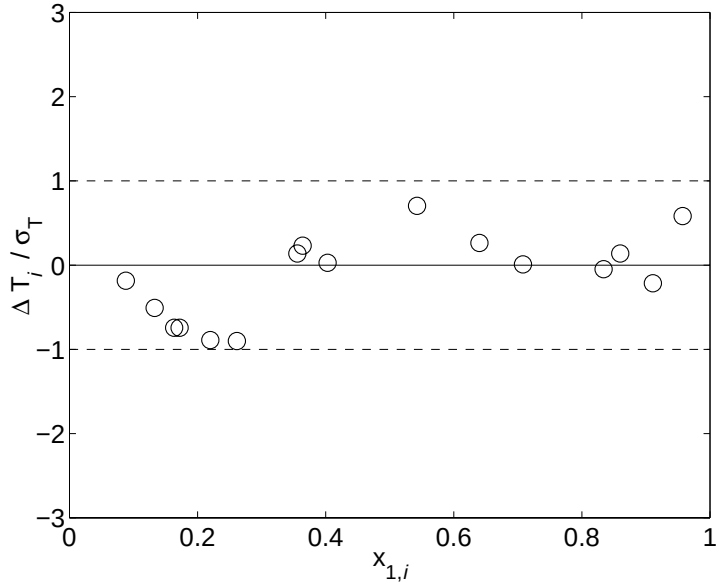

Figure 15. Globally optimal error residuals in $\Delta T_i = \tilde{T}_i - T_i$ relative to σ_T when $\boldsymbol{\sigma} = (0.003, 0.0029, 1.7, 0.083)^{\mathrm{T}}$.

interval methodology is a general-purpose approach that can be applied to a wide variety of parameter estimation and data reconciliation problems. However, the guaranteed reliability of the interval approach comes at the expense of significantly higher computation time requirements in comparison to local methods that provide no such guarantees. Thus, modelers must consider this trade-off, and ultimately decide how important it is to know for sure that the correct answer has been obtained. As seen in the case of the third standard deviation vector considered above, in evaluating the ability of a given model to fit a set of data, it is extremely important that the parameter estimation problem be correctly solved, to a global and, not just local, optimum. Failure to do so may result in the rejection of a perfectly adequate model.

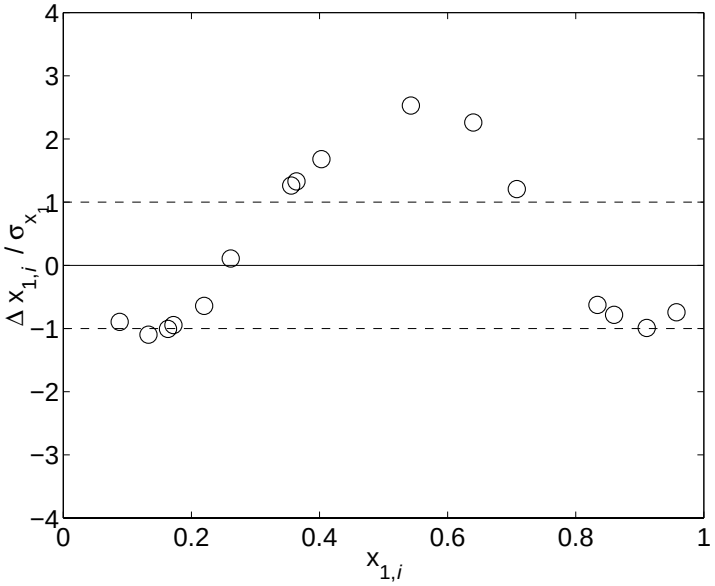

Figure 16. Locally (but not globally) optimal error residuals in $\Delta x_{1,i} = \tilde{x}_{1,i} - x_{1,i}$ relative to σ_{x_1} when $\boldsymbol{\sigma} = (0.003, 0.0029, 1.7, 0.083)^{\mathrm{T}}$.

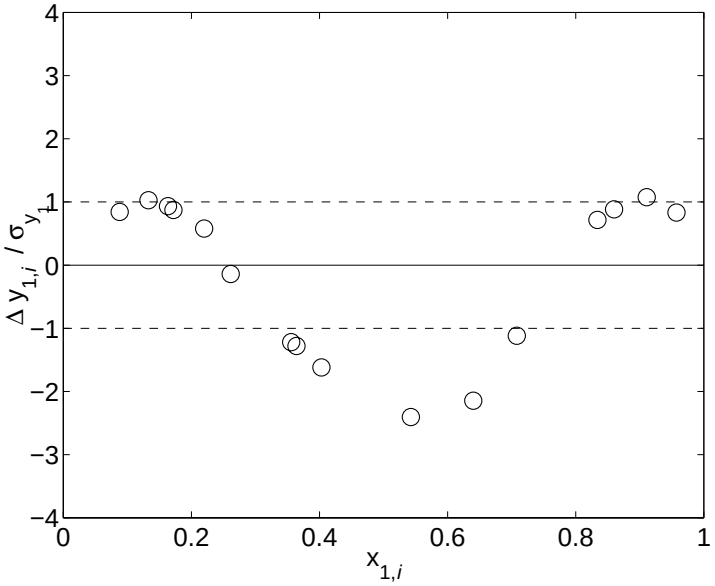

Figure 17. Locally (but not globally) optimal error residuals in $\Delta y_{1,i} = \tilde{y}_{1,i} - y_{1,i}$ relative to σ_{y_1} when $\boldsymbol{\sigma} = (0.003, 0.0029, 1.7, 0.083)^{\mathrm{T}}$.

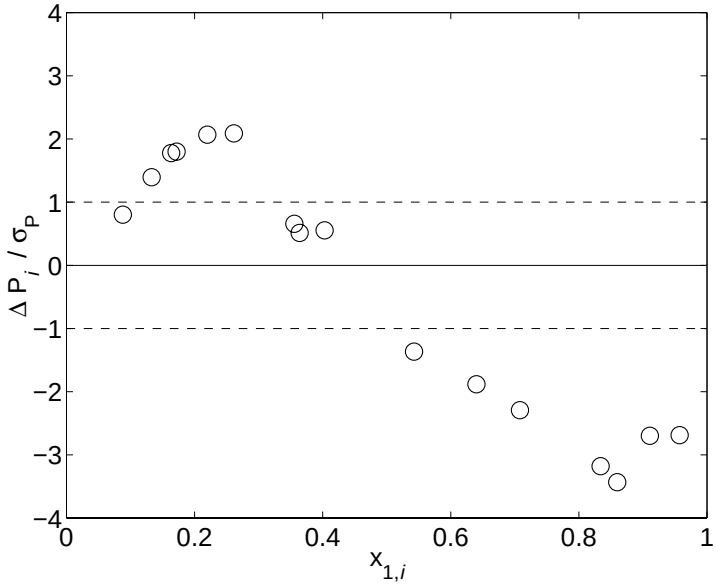

Figure 18. Locally (but not globally) optimal error residuals in $\Delta P_i = \tilde{P}_i - P_i$ relative to σ_P when $\boldsymbol{\sigma} = (0.003, 0.0029, 1.7, 0.083)^{\mathrm{T}}$.

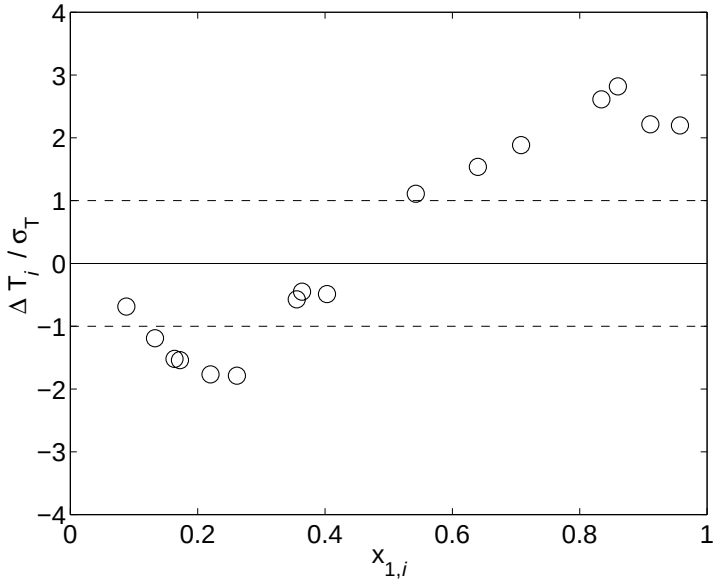

Figure 19. Locally (but not globally) optimal error residuals in $\Delta T_i = \tilde{T}_i - T_i$ relative to σ_T when $\boldsymbol{\sigma} = (0.003, 0.0029, 1.7, 0.083)^{\mathrm{T}}$.

The interval-Newton methodology described here is a powerful approach for nonlinear equation solving and global optimization. Here it has been applied to the problem of nonlinear parameter estimation and data reconcilation; however, it can also be used in a wide variety of other applications. These applications include the determination of fluid phase stability and equilibrium, using either activity coefficient models (Stadtherr et al. [27]; McKinnon et al. [28]; Tessier et al. [29]) or equation-of-state models (Hua et al. [30-33]; Xu et al. [34]), the computation of azeotropes (Maier et al. [35,36]) and critical points (Sradi et al. [37]), the determination of solid-fluid equilibrium (Xu et al. [38]), and density-functional-theory calculations of adsorption in nanoscale pores (Maier and Stadtherr [39]).

ACKNOWLEDGMENTS

This work has been supported in part by the donors of The Petroleum Research Fund, administered by the ACS, under Grant 35979-AC9, by the National Science Foundation Grant EEC97-00537-CRCD, and by the Environmental Protection Agency Grant R826-734-01-0.

REFERENCES

[1] Duever, T. A., Keeler, S. E., and Reilly, P. M. 1987, Chem. Eng. Sci., **42**, 403.

[2] Patino-Leal, H. and Reilly, P. M. 1982, AIChE J., **28**, 580.

[3] Kim, I., Liebman, M., and Edgar, T. F. 1990, AIChE J., **36**, 985.

[4] Esposito, W. R. and Floudas, C. A. 1998, Ind. Eng. Chem. Res., **37**, 1841.

[5] Bard, Y. 1974, Nonlinear Parameter Estimation, Academic Press, New York, NY, USA.

[6] Gallant, A. R. 1987, Nonlinear Statistical Models, Wiley, New York, NY, USA.

[7] Seber, G. A. F. 1989, Nonlinear Regression, Wiley, New York, NY, USA.

[8] Englezos, P. and Kalogerakis, N. 2001, Applied Parameter Estimation for Chemical Engineers, Marcel Dekkar, New York, NY, USA.

[9] Britt, H. I. and Luecke, R. H. 1973, Technometrics, **15**, 233.

[10] Fabries, J.-F. and Renon, H. 1975, AIChE J., **21**, 735.

[11] Anderson, T. F., Abrams, D. S., and Grens, E. A. 1978, AIChE J., **24**, 20.

[12] Schwetlick, H. and Tiller, V. 1985, Technometrics, **27**, 17.

[13] Valko, P. and Vajda, S. 1987, Comput. Chem. Eng., **11**, 37.

[14] Tjoa, I.-B. and Biegler, L. T. 1991, Ind. Eng. Chem. Res., **30**, 376.

[15] Tjoa, I.-B. and Biegler, L. T. 1992, Comput. Chem. Eng., **16**, 523.

[16] Vamos, R. J. and Hass, C. N. 1994, AIChE J., **40**, 556.

[17] Luus, R. and Hernaez, H. 2000, Hung. J. Ind. Chem., **28**, 201.

[18] Gau, C.-Y. and Stadtherr, M. A. 2000, Comput. Chem. Eng., **24**, 631.

[19] Gau, C.-Y. and Stadtherr, M. A. 2001, New Interval Methodologies for Reliable Chemical Process Modeling, Comput. Chem. Eng., submitted for publication.

[20] Neumaier, A. 1990, Interval Methods for Systems of Equations, Cambridge University Press, Cambridge, UK.

[21] Hansen, E. R. 1992, Global Optimization Using Interval Analysis, Marcel Dekkar, New York, NY, USA.

[22] Kearfott, R. B. 1996, Rigorous Global Search: Continuous Problems, Kluwer, Dordrecht, The Netherlands.

[23] Schnepper, C. A. and Stadtherr, M. A. 1996, Comput. Chem. Eng., **20**, 187.

[24] Gmehling, J., Onken, U., and Arlt, W. 1977–1990, Vapor-Liquid Equilibrium Data Collection, Chemistry Data Series, Vol. I, Parts 1-8, DECHEMA, Frankfurt/Main, Germany.

[25] Gau, C.-Y., Brennecke, J. F., and Stadtherr, M. A. 2000, Fluid Phase Equilib., **168**, 1.

[26] Gau, C.-Y. 2001, Reliable High Performance Computing Strategies For Chemical Process Modeling: Nonlinear Parameter Estimation, Ph.D. Thesis, University of Notre Dame, Notre Dame, IN, USA.

[27] Stadtherr, M. A., Schnepper, C. A., and

Brennecke, J. F. 1995, AIChE Symp. Ser., **91**(304), 356.

[28] McKinnon, K. I. M., Millar, C. G., and Mongeau M. 1996, in State of the Art in Global Optimization: Computational Methods and Applications, Floudas, C. A. and Pardalos, P. M. (Eds.), Kluwer, Dordrecht, The Netherlands.

[29] Tessier, S. R., Brennecke, J. F., and Stadtherr, M. A. 2000, Chem. Eng. Sci., **55**, 1785.

[30] Hua, J. Z., Brennecke, J. F., and Stadtherr, M. A. 1996, Fluid Phase Equilib., **116**, 52.

[31] Hua, J. Z., Brennecke, J. F., and Stadtherr, M. A. 1998, Comput. Chem. Eng., **22**, 1207.

[32] Hua, J. Z., Brennecke, J. F., and Stadtherr, M. A. 1998, Ind. Eng. Chem. Res., **37**, 1519.

[33] Hua, J. Z., Maier, R. W., Tessier, S. R., Brennecke, J. F., and Stadtherr, M. A. 1999, Fluid Phase Equilib., **158**, 607.

[34] Xu, G., Brennecke, J. F., and Stadtherr, M. A. 2001, Reliable Computation of Phase Stability and Equilibrium from the SAFT Equation of State, Ind. Eng. Chem. Res., in press.

[35] Maier, R. W, Brennecke, J. F., and Stadtherr, M. A. 1998, AIChE J., **44**, 1745.

[36] Maier, R. W, Brennecke, J. F., and Stadtherr, M. A. 2000, Comput. Chem. Eng., **24**, 1851.

[37] Stradi, B., Brennecke, J. F., and Stadtherr, M. A. 2001, AIChE J., **47**, 212.

[38] Xu, G., Scurto, A. M., Castier, M., Brennecke, J. F., and Stadtherr, M. A. 2000, Ind. Eng. Chem. Res., **39**, 1624.

[39] Maier, R. W, and Stadtherr, M. A. 2001, AIChE J., **47**, 1874.