Interval Robustness in Nonsmooth Nonlinear Parameter Estimation

Humberto Muñoz and R. B. Kearfott
Department of Mathematics,
University of Louisiana at Lafayette
Box 4-1010, Lafayette, LA 70504-1010 USA
hnm2294@louisiana.edu
rbk@louisiana.edu

Abstract

The reliable solution of nonlinear parameter estimation problems is an important computational problem in chemical engineering. Classical solution methods for these problems are local methods, and may not be reliable to find the global optimum. Interval arithmetic can be used to compute completely reliably the global optimum for the nonlinear parameter estimation problem. Experimental results are given for parameter estimation in vapor-liquid equilibrium (VLE) models originally set forth by *Stadtherr–Gau.* To complement *Stadtherr's* work, we consider not only least squares, but also the l_1 and l_∞ nonsmooth estimators.

1 Introduction

Parameter estimation is a common problem in many areas of science and engineering, including such applications as real time optimization. Its goal is to estimate accurate model parameters that provide the best fit to measured data, despite small-scale noise in the data or occasional large-scale measurement errors (outliers). In general, the estimation techniques are based on some kind of least squares or maximum likelihood criterion, and these require the solution of a nonlinear and nonconvex optimization problem.

The standard methods (gradient-based approaches: Gauss-Newton methods, Gauss-Marquardt methods, and successive quadratic programming methods, or non-gradient methods, such as the simplex pattern search) used to solve these problems are local methods that provide no guarantee the global optimum, and the best model parameters have been found.

Interval arithmetic can be used to compute completely reliably the global optimum for the nonlinear parameter estimation problem. *Stadtherr–Gau* [1], used interval analysis in an application that deals with nonlinear parameter estimation in vapor-liquid equilibrium (VLE) models. The reliable solution of nonlinear parameter estimation problems is an important computational problem in chemical engineering. Classical solution methods for these problems are local methods and may not be reliable to find the global optimum. *Stadtherr–Gau* considered the least squares estimator l_2 in their work.

The goal of this work is show the results of the nonsmooth estimators l_1 and l_∞, in the solution of parameter estimation in VLE models using nonsmooth optimization techniques in interval arithmetic, and also compare them to *Stadtherr*'s work.

The second section introduces parameter estimation, and the three different objective estimators l_1, l_2 and l_∞ used in this work. The third section presents basic concepts in interval arithmetic and the algorithm solution. The fourth section is devoted to the application in VLE modeling. Finally, the fifth section shows numerical results and conclusions.

2 Parameter Estimation

Suppose that n observations of m response variables, $y_{ji}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n$ are available, and that the responses are to be fit to a model of the form $y_{ji} = f_i(x_j, \theta)$, with independent variables $x_j = (x_{j1}, x_{j2}, \ldots, x_{jp})^T$ and parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_q)^T$. Various objective functions (or estimators) $\phi(\theta)$ can be used to obtain the parameter values that provide the best fit. In many circumstances, a maximum likelihood estimate is most appropriate. However, assuming a normal likelihood in the errors, this can be simplified to the widely relative least squares criterion or the l_2 norm of the relative errors, and to obtain the objective function

$$\phi(\theta) = \sum_{i=1}^{n} \sum_{j=1}^{m} \left[\frac{y_{ji} - f_i(x_j, \theta)}{y_{ji}} \right]^2. \tag{1}$$

Similarly, using the l_1 and l_∞ norms we get, respectively, the objective functions:

$$\phi(\theta) = \sum_{i=1}^{m}\sum_{j=1}^{n}\left|\frac{y_{ji} - f_i(x_j,\theta)}{y_{ji}}\right|, \tag{2}$$

or

$$\phi(\theta) = \max_{1\le i\le m, 1\le j\le n}\left|\frac{y_{ji} - f_i(x_j,\theta)}{y_{ji}}\right|. \tag{3}$$

These can be treated either as a constrained or, if the experimental observations are substituted directly into the objective function, unconstrained formulation of the problem here. In general, there are a variety of standard techniques to minimize ϕ that provide local minimum, but no assurance that a global minima has been found. Using interval arithmetic can provide such a technique.

3 Interval Arithmetic: Basic Concepts

Real interval arithmetic is based on closed intervals of real numbers, i.e. $\boldsymbol{x} = [\underline{x}, \overline{x}]$. A real interval vector is $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$, where $\boldsymbol{x}_i = [\underline{x}_i, \overline{x}_i]$ can be interpreted geometrically as an n-dimensional box. There are good introductions to interval analysis in *Kearfott* [6], *Hansen* [2], *Neumaier* [8].

The interval extensions and interval Newton methods have been developed for global solution of nonlinear systems of equations and for global optimization. These techniques provide the capability to narrowly enclose all roots of the systems within the given initial interval. It is well known that when the functions are given by smooth expressions, without conditional branches, this technique is quadratically convergent. For instance, the unconstrained minimization of the relative least squares function, $\phi(\theta)$, a common approach is to use the gradient of $\phi(\theta)$ and seek a solution of $g(\theta) = \nabla\phi(\theta) = 0$. The global minimum will be a root of this nonlinear equation system, but there may be many other roots as well, representing local extremes and saddle points. Thus, for this approach to be reliable, it is necessary to find all the roots of $g(\theta) = 0$, and this is provided by the interval Newton techniques. Additional details concerning the implementation of interval Newton methods, and the theory underlying them, is available from *Kearfott* [6], *Hansen* [2], *Neumaier* [8]. Special algorithms developed in GlobSol handle non-smooth problems such as l_1 and l_∞ optimizations with the same techniques as smooth problems, and under certain conditions the interval Newton method converges linearly (See *Kearfott* [4], [5], *Ratz* [10]).

In practice, the interval Newton procedure can also be combined with an interval branch and bound technique, so that roots of $g(\theta) = 0$ that cannot be the global minimum need not be found. The solution algorithm is applied to a sequence of intervals, beginning with some initial interval vector $\boldsymbol{\theta}^{(0)}$ given by the user. The initial interval can be chosen to be sufficiently large to enclose all physically feasible behavior. It is assumed that the global optimum will occur at an interior stationary minimum of $\phi(\theta)$ and not on the boundary of $\boldsymbol{\theta}^{(0)}$. Since the estimator $\phi(\theta)$ is derived from a product of normal distribution or double exponential functions corresponding to each data point, only a stationary global minimum is reasonable for statistical regression problems such as considered here.

4 Solution Algorithm by *Stadtherr–Gau*

Proposed in [1], The following algorithm is similar to generic algorithms that have appeared, essentially in *Moore* [7],*Ratschek–Rohn* [9],*Hansen* [2], and *Kearfott* [6].

For an interval $\boldsymbol{\theta}^{(k)}$ in the sequence, follow the steps:

1. *Function range test.* An interval extension $G(\boldsymbol{\theta}^{(k)})$ of the function $g(\theta) = \nabla\phi(\theta)$ is computed.

    a. If there is any component of the interval extension $G(\boldsymbol{\theta}^{(k)})$ that does not contain zero, the current interval $\boldsymbol{\theta}^{(k)}$ is discarded, thus no solution of $g(\theta) = 0$ exists in this interval.

    b. Otherwise, if $0 \in G(\boldsymbol{\theta}^{(k)})$, then testing of $\boldsymbol{\theta}^{(k)}$ continues.

2. *Objective range test.* An interval extension $\Phi(\boldsymbol{\theta})$ of the function $\phi(\theta)$ is computed.

    a. If the lower bound of $\Phi(\boldsymbol{\theta})$ is greater than a known upper bound on the global minimum of $\phi(\theta)$, then $\boldsymbol{\theta}^{(k)}$ cannot contain the global minimum and it is discarded.

    b. Otherwise, testing of $\boldsymbol{\theta}^{(k)}$ continues. The upper bound of $\phi(\theta)$ can be determined and updated in different ways. Here we use point evaluations of $\phi(\theta)$ done at the midpoint of previous tested $\boldsymbol{\theta}$ intervals that may contain stationary points.

3. *Interval Newton test.* Here the linear interval equation system

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

is set up and solved for a new interval $\boldsymbol{N}^{(k)}$, where $G'(\boldsymbol{\theta}^{(k)})$ is an interval extension of the Jacobian of $g(\theta)$, i.e. the Hessian of $\phi(\theta)$, over the current interval $\boldsymbol{\theta}^{(k)}$, and $\theta^{(k)} \in \boldsymbol{\theta}^{(k)}$, usually taken to be the midpoint of $\boldsymbol{\theta}^{(k)}$. In *Kearfott 1996* [6], *Hansen* [2], *Neumaier* [8] it is shown that any root $\theta^* \in \boldsymbol{\theta}^{(k)}$ is also contained in $\boldsymbol{N}^{(k)}$, implying that if $\boldsymbol{\theta}^{(k)} \cap \boldsymbol{N}^{(k)} = \emptyset$, then there is no root of $g(\theta) = 0$ in $\boldsymbol{\theta}^{(k)}$, and suggesting the iteration scheme $\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} \cap \boldsymbol{N}^{(k)}$. The foregoing suggests a series of tests to determine whether a stationary point (root of $g(\theta) = 0$) that might be the global minimum of $\phi(\theta)$ can be contained in $\boldsymbol{\theta}^{(k)}$:

   a. If $\boldsymbol{\theta}^{(k)} \cap \boldsymbol{N}^{(k)} = \emptyset$, then $\boldsymbol{\theta}^{(k)}$ is discarded.

   b. Evaluate $\phi(\theta^{(k)})$ to determine and update an upper bound on the global minimum for use in step 2.

   c. If $\boldsymbol{N}^{(k)} \subset \boldsymbol{\theta}^{(k)}$, then there is exactly one root of $g(\theta) = 0$ in $\boldsymbol{\theta}^{(k)}$, which may correspond to the global minimum.

   d. If neither of the above is true, then no further conclusion can be drawn.

In the last case, one could then repeat the root inclusion test on the next interval Newton iterate $\boldsymbol{\theta}^{(k+1)}$, assuming it is sufficiently smaller that $\boldsymbol{\theta}^{(k)}$, or one could bisect $\boldsymbol{\theta}^{(k+1)}$ and repeat the root inclusion test on the resulting intervals. This is the basic idea of interval Newton/generalized bisection (IN/GB). The worst-case computational complexity of the (IN/GB) algorithm is exponential in the number of variables. However, process modeling problems involving over a hundred variables have been successfully solved using this approach (*Schnepper and Stadtherr,* [11]).

5 Application to VLE Modeling

Because of its importance in the design of separation systems, much attention has been given to modeling the thermodynamics of phase equilibrium in fluid mixtures. As an example, we consider here the estimation from binary (VLE) data of the energy parameters in the Wilson equation

$$\frac{g^E}{RT} = -x_1 \ln(x_1 + \Lambda_{12} x_2) - x_2 \ln(x_2 + \Lambda_{21} x_1) = x_1 \ln \gamma_1 + x_2 \ln \gamma_2, \qquad (4)$$

where g^E is the molar excess Gibbs energy for a binary system, x_1 and x_2 the liquid-phase mole fractions. From (4) expressions for the activity coefficients γ_1 and γ_2 are

$$\ln \gamma_1 = -\ln(x_1 + \Lambda_{12} x_2) + x_2 \left[\frac{\Lambda_{12}}{x_1 + \Lambda_{12} x_2} - \frac{\Lambda_{21}}{x_2 + \Lambda_{21} x_1} \right], \qquad (5)$$

$$\ln \gamma_2 = -\ln(x_2 + \Lambda_{21} x_1) - x_1 \left[\frac{\Lambda_{12}}{x_1 + \Lambda_{12} x_2} - \frac{\Lambda_{21}}{x_2 + \Lambda_{21} x_1} \right], \qquad (6)$$

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

$$\Lambda_{12} = \frac{v_2}{v_1} exp \left[\frac{-\theta_1}{RT} \right], \qquad (7)$$

$$\Lambda_{21} = \frac{v_1}{v_2} exp \left[\frac{-\theta_2}{RT} \right], \qquad (8)$$

v_1 and v_2 are the pure component liquid molar volumes, T is the system temperature and θ_1 and θ_2 are the energy parameters that need to be estimated.

Given VLE measurements and assuming an ideal vapor phase, experimental values $\gamma_{1,exp}$ and $\gamma_{2,exp}$ of the activity coefficients can be obtained from the relation

$$\gamma_{i,exp} = \frac{y_{i,exp} P_{exp}}{x_{i,exp} P_i^0}, \qquad i = 1, 2, \qquad (9)$$

where P_i^0 is the vapor pressure of pure component i at the system temperature T. For the example problems here we follow *Gmehling et at.* [3] and consider the three different objective functions in the introduction, with $y_{ji} = \gamma_{ji,exp}$ and $f_i(x_j, \theta) = \gamma_{ji,calc(\theta)}$, i.e. the relative errors are

$$\frac{y_{ji} - f_i(x_j, \theta)}{y_{ji}} = \frac{\gamma_{ji,exp} - \gamma_{ji,calc(\theta)}}{\gamma_{ji,exp}}, \qquad j = 1, \ldots, n, \quad i = 1, 2, \qquad (10)$$

where $\gamma_{ji,calc(\theta)}$ are calculated from the Wilson equation at the same conditions (temperature, pressure and composition) used to measure $\gamma_{ji,exp}$.

Using the solution algorithm on section 1.2, with the different objective functions $\phi(\theta)$, defined in (1), (2), (3) we can estimate the energy parameters, θ_1 and θ_2. The next section presents results and discussion of these three estimations.

6 Results and Conclusions

In the following tables $\theta^{(2)}$, $\theta^{(1)}$, $\theta^{(\infty)}$, denote the solutions for the objective functions defined by (1), (2) and (3) respectively, and $\theta^{(D)}$ is the solution published in *Gmehling et at.* [3]. Each column $l_\infty(\theta)$, $l_1(\theta)$ and $l_2(\theta)$ considers the evaluation of these objective functions in all solutions.

(a) Without the diagonal numbers in Table 1, a robust comparison among the three different estimators can be considered. It can be established by column ranking, assigning the scores 3, 2, and 1 from the minimum value to the largest value in each column. The most robust solution would be the solution with largest accumulative score. In Table 1 the l_2 solution is the most robust with score 7, followed by the l_1 solution with score 6, and the worst solution is the l_∞ solution with score 5.

(b) For Tables 2, 3, 4 the scoring for the three estimators is the same, 6. Considering this simple robust method, is impossible to choose what estimator is the best. In this case it is necessary to use other robust procedure to compare these estimators, for instance, observe the sensitivity of these estimators with respect to different kinds of outliers.

(c) In general, this interval technique of parameter estimation is model independent. The approach presented here is general purpose and can be used in connection with other objective functions, such as maximum likelihood, or other type of data.

(d) In all tables the evaluations of the three objective functions in $\theta^{(2)}$, $\theta^{(1)}$, $\theta^{(\infty)}$ are lower than their corresponding evaluations in $\theta^{(D)}$. This shows that using the nonsmooth parameter estimators we have obtained similar results than those in (*Stadtherr–Gau* [1]), using the smooth least squares estimator, for the same four data set. In other words, the solutions $\theta^{(D)}$ published in *Gmehling et at.* [3] correspond to local optima only whereas interval approaches find globally optimal parameter values.

(e) Since this is a very wide interval based on physical considerations, we believe that it is extremely likely that it will contain the globally optimal parameter values. However, it should be emphasized again that the solution algorithm is, of course, only guaranteed to converge to a global solution, that is a stationary point within this chosen initial parameter interval. It should also be noted that other approaches, including the use of system-specific information, could be used to establish reasonable initial bounds.

Table 1: Results for the Data Set one

T(∘ C)	Solution	$l_\infty(\theta)$	$l_1(\theta)$	$l_2(\theta)$
30	$\theta^{(\infty)} = (-455.5, 1135)$	0.0620	0.4451	0.0140
30	$\theta^{(1)} = (-454.1, 1255)$	0.0848	0.3639	0.0130
30	$\theta^{(2)} = (-468.5, 1320)$	0.0713	0.3758	0.0118
30	$\theta^{(D)} = (437, -437)$	0.1280	0.7139	0.0383

Table 2: Results for the Data Set two

T(∘ C)	Solution	$l_\infty(\theta)$	$l_1(\theta)$	$l_2(\theta)$
40	$\theta^{(\infty)} = (-440.8, 1058.1)$	0.0490	0.3705	0.0101
40	$\theta^{(1)} = (-454.1, 1255)$	0.0848	0.3639	0.0080
40	$\theta^{(2)} = (-468.5, 1320)$	0.0713	0.3758	0.0078
40	$\theta^{(D)} = (405, -405)$	0.1163	0.6624	0.0329

Table 3: Results for the Data Set three

T(∘ C)	Solution	$l_\infty(\theta)$	$l_1(\theta)$	$l_2(\theta)$
50	$\theta^{(\infty)} = (-425.9, 986.4)$	0.0472	0.2500	0.0081
50	$\theta^{(1)} = (-447.0, 1158.7)$	0.0507	0.2499	0.0057
50	$\theta^{(2)} = (-449.7, 1162.6)$	0.0489	0.2567	0.0057
50	$\theta^{(D)} = (374, -374)$	0.1088	0.6248	0.0289

Table 4: Results for the Data Set four

T(∘ C)	Solution	$l_\infty(\theta)$	$l_1(\theta)$	$l_2(\theta)$
50	$\theta^{(\infty)} = (-423.3, 976.9)$	0.0413	0.3246	0.0084
50	$\theta^{(1)} = (-388.2, 861.4)$	0.0619	0.2923	0.0093
50	$\theta^{(2)} = (-417.9, 969.3)$	0.0475	0.3130	0.0081
50	$\theta^{(D)} = (342, -342)$	0.1343	0.6838	0.0426

References

[1] C. Y. Gau and M. A. Stadtherr. Nonlinear parameter estimation using interval analysis, 1999. Preprint, University of Notre Dame, July 26.

[2] E. R. Hansen. *Global Optimization Using Interval Analysis*. Marcel Dekker, Inc., New York, 1992.

[3] U. Onken J. Gmehling and W. Arlt. Vapor-liquid equilibrium data collection. In *Chemistry Data Series*, volume 1. DECHEMA, Frankfurt/Main, Germany, 1977–1990.

[4] R. B. Kearfott. Interval extensions of non-smooth functions for global optimization and nonlinear systems solvers. *Computing*, 57:149–162, 1996.

[5] R. B. Kearfott. Treating non-smooth functions as smooth functions in global optimization and nonlinear systems solvers. In G. Alefeld, A. Frommer, and B. Lang, editors, *Scientific Computing and Validated Numerics*, Mathematical Research, volume 90, pages 160–172, Berlin, 1996. Akademie Verlag.

[6] R. Baker Kearfott. *Rigorous Global Search: Continuous Problems.* Kluwer Academic Publishers, Netherlands, 1996.

[7] R. E. Moore. *Methods and Applications of Interval Analysis.* SIAM, Philadelphia, 1979.

[8] A. Neumaier. *Interval Methods for Systems of Equations.* Cambridge University Press, Cambridge, England, 1990.

[9] H. Ratschek and J. Rokne. *New Computer Methods for Global Optimization.* Wiley, New York, 1988.

[10] D. Ratz. *Automatic Slope Computation and its Application in Nonsmooth Global Optimization.* Shaker Verlag, Aachen, 1998.

[11] C. A. Schnepper and M. A. Stadtherr. Robust process simulation using interval methods. *Comput. Chem.Eng.*, 20:187–199, 1996.