Improved Accuracy of Nonlinear Parameter Estimation with LAV and Interval Arithmetic Methods

Humberto MUÑOZ
Department of Mathematics, Southern University and A&M College
Baton Rouge, LA 70813, USA

and

Nigel GWEE
Department of Computer Science, Southern University and A&M College
Baton Rouge, LA 70813, USA

ABSTRACT

The reliable solution of nonlinear parameter estimation problems is an important computational 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 non-convex optimization problem. Classical solution methods for these problems are local methods, and may not be reliable for finding the global optimum, with no guarantee the best model parameters have been found. Interval arithmetic can be used to compute completely and reliably the global optimum for the nonlinear parameter estimation problem. Finally, experimental results will compare the least squares, l_2, and the least absolute value, l_1, estimates using interval arithmetic in a chemical engineering application.

Keywords: Global optimization, interval computations, non-smooth optimization, least absolute value, least squares estimators.

1. INTRODUCTION

In general, parameter estimation techniques are based on some kind of least squares or maximum likelihood criterion, and these require the solution of a nonlinear and non-convex 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 that both the global optimum and the best model parameters have been found. Interval arithmetic can be used to compute reliably the global optimum for a nonlinear parameter estimation problem.

The second section introduces the topic of parameter estimation. The third section presents basic concepts in interval arithmetic and non-smooth interval optimization techniques. The fourth section shows numerical results in a chemical engineering application and finally, the fifth section presents conclusions and future work.

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 fitted 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 used relative least squares (LS) criterion or the l_2 norm of the relative errors, and to obtain the objective function

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

Similarly, using the l_1 norm for the least absolute value (LAV) criterion the objective function is

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

LAV has been used successfully in multiple regression [2]. In general, there are a variety of standard techniques to minimize ϕ that provide a local minimum, but no assurance that a global minimum has been found. This paper shows that the LAV criterion combined with interval arithmetic techniques is a reliable tool in parameter estimation.

3. INTERVAL ARITHMETIC

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. Comprehensive introductions to interval analysis can be found in [6], [5] and [10].

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 enclose narrowly 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, in 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.

Non-smooth phenomena in mathematics and optimization occur naturally and frequently, and there is a need to be able to deal with them. Recent literature presents associated techniques which are well suited to this purpose. Many practical optimization problems, in particular those containing expressions such as $|E(x)|$ and $\max\{E(x), F(x)\}$, $E, F : \mathbb{R}^n \to \mathbb{R}$ or functions defined by If-then-else branches, result in discontinuous functions or functions whose derivatives have jump discontinuities. For such problems, gradient type methods cannot be applied. There are different approaches to overcome this difficulty.

Consider a function $f : \mathbb{R}^n \to \mathbb{R}$, and a point $x \in \mathbb{R}^n$. The classical gradient of f at x is defined only when f is differentiable at x. For non-differentiable functions f, various generalizations of the gradient are summarized in [1] which are termed Clarke subgradients. A common feature of these subgradients is that they only capture information about the epigraph of f near $(x, f(x))$. In [8] *Muñoz and Kearfott* showed that it is possible to find an interval enclosure for all these techniques that incorporates information of f in a complete neighborhood of x.

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 those considered here.

Elementary Operations

Real interval arithmetic was introduced in its modern form in [7], and is based on arithmetic within the set of closed intervals of real numbers. Fundamental definitions of interval vectors, interval arithmetic operations, and interval slope sets now follow.

Definition 1 A real bounded and closed *interval* is defined by

$$\boldsymbol{x} = [\underline{x}, \bar{x}] := [\inf \boldsymbol{x}, \sup \boldsymbol{x}] \in \mathbb{IR},$$

$\check{x}$ is the center or midpoint for the interval $\boldsymbol{x}$ An n-dimensional *interval vector*, (also called *box*) is defined by

$$\boldsymbol{X} := (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)^T \in \mathbb{IR}^n,$$

where $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ are real intervals and $\mathbb{IR}^n$ is the set of real interval vectors. An *interval matrix* $\mathbf{A} = (\mathbf{a_{ij}})$ is a matrix all of whose entries $\mathbf{a_{ij}}$ are intervals.

Definition 2 For the intervals $\boldsymbol{x} = [\underline{x}, \bar{x}]$ and $\boldsymbol{y} = [\underline{y}, \bar{y}]$, the basic interval operations are defined as follows.

$$\begin{aligned}
\boldsymbol{x} + \boldsymbol{y} &= [\underline{x} + \underline{y}, \bar{x} + \bar{y}] \\
\boldsymbol{x} - \boldsymbol{y} &= [\underline{x} - \bar{y}, \bar{x} - \underline{y}] \\
\boldsymbol{x} \times \boldsymbol{y} &= [m, M] \quad \text{where} \\
m = &\ \min\{\underline{x}\underline{y}, \underline{x}\bar{y}, \bar{x}\underline{y}, \bar{x}\bar{y}\},
\end{aligned}$$

$$
\begin{array}{rrl}
 & M = & \max\{\underline{x}\underline{y}, \underline{x}\bar{y}, \bar{x}\underline{y}, \bar{x}\bar{y}\} \\
\text{if} \quad 0 \notin \boldsymbol{x} & \frac{1}{\boldsymbol{x}} & = \left[\frac{1}{\bar{x}}, \frac{1}{\underline{x}}\right] \\
\boldsymbol{x} \div \boldsymbol{y} & = & \boldsymbol{x} \times \dfrac{1}{\boldsymbol{y}}
\end{array} \tag{3}
$$

The following example illustrates interval arithmetic evaluations by using the formulas in Eq. (3).

Example 1 Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the function $f(x, y) := (y - x^2)^2 + 2x$. The interval evaluation of f on the interval vector $\boldsymbol{X} = (\boldsymbol{x}, \boldsymbol{y})^T = ([0.7, 1], [0.0, 0.3])^T$ is obtained as follows. For any $x \in \boldsymbol{x}$ and $y \in \boldsymbol{y}$,

$$
\begin{array}{rcl}
f(x,y) & = & (y - x^2)^2 + 2x \\
 & \in & (\boldsymbol{y} - \boldsymbol{x}^2)^2 + 2\boldsymbol{x} \\
 & = & ([0.0, 0.3] - [0.7, 1]^2)^2 + 2[0.7, 1] \\
 & = & ([0.0, 0.3] - [0.49, 1])^2 + [1.4, 2] \\
 & = & ([-1.0, -0.19]^2 + [1.4, 2] \\
 & = & [0.036, 1] + [1.4, 2] \\
 & = & [1.436, 3].
\end{array}
$$

We will limit ourselves to the following definition of interval slopes, taken from [6], p. 27.

Definition 3 Let $f : \mathbb{R}^n \to \mathbb{R}$. The vector $\mathbf{S}$ is said to be a **slope set** for f over $\boldsymbol{x}$ and centered on the interval vector $\check{\boldsymbol{x}}$ if, for every $x \in \boldsymbol{x}$ and $\check{x} \in \check{\boldsymbol{x}}$, $\boldsymbol{x} \neq \check{\boldsymbol{x}}$, there is a $S \in \mathbf{S}$ such that

$$
f(x) - f(\check{x}) \quad = \quad S \cdot (x - \check{x}). \tag{4}
$$

Any smallest such set of vectors satisfying this condition will be denoted by $\mathbf{S}^\sharp(f, \boldsymbol{x}, \check{\boldsymbol{x}})$. The smallest interval vector (or interval hull) that contains $\mathbf{S}^\sharp(f, \boldsymbol{x}, \check{\boldsymbol{x}})$ will be denoted by $\square\mathbf{S}^\sharp(f, \boldsymbol{x}, \check{\boldsymbol{x}})$, and is called the *interval slope* of f over $\boldsymbol{x}$, where it is not necessary that f is differentiable.

Usually, $\check{\boldsymbol{x}}$ is a point $\check{x}$ or a very small box. In the one-dimensional case, the slope function of f with center in $\check{x}$, $\mathbf{s}(f, x, \check{x})$, can be defined by

$$
\mathbf{s}(f, x, \check{x}) = \begin{cases} \frac{f(x) - f(\check{x})}{x - \check{x}}, & x \neq \check{x}, \\ \hat{s}, & x = \check{x}, \end{cases} \tag{5}
$$

where $\hat{s} \in \mathbb{R}$ may be arbitrarily chosen. Assuming f to be continuously differentiable, we can define $\hat{s} = f'(\check{x})$, and thus make $\mathbf{s}(f, x, \check{x})$ continuous in x and $\check{x}$.

Example 2 Let $f : \mathbb{R} \to \mathbb{R}$ be the continuous function defined by

$$
f(x) = \begin{cases} (x-1)^2, & x \geq 1, \\ 1 - x^2, & x < 1, \end{cases}
$$

and $\boldsymbol{x} = [0, 2]$, $\check{x} = 1$. From Eq. (5) the slope set and the interval slope of f over $\boldsymbol{x}$ are $\mathbf{S}^\sharp(f, \boldsymbol{x}, \check{x}) = [-2, -1] \cup [0, 1]$, and $\square\mathbf{S}^\sharp(f, \boldsymbol{x}, \check{x}) = [-2, 1]$ respectively.

Automatic Slope Computation for Non-smooth Functions

The process of automatic slope computation (ASC) for non-smooth functions is very similar to the process of automatic differentiation. The ASC evaluates functions specified by algorithms or formulas according to the rules of a *slope arithmetic*. Through this subsection, $\mathbf{s}(f, \boldsymbol{x}, \check{\boldsymbol{x}})$ denotes the *automatic interval evaluation* of $\square\mathbf{S}^\sharp(f, \boldsymbol{x}, \check{\boldsymbol{x}})$, when f is continuous at $\check{\boldsymbol{x}}$. Introduced in [6] (p. 211-218), the function $x_p = \chi(x_s, x_q, x_r)$ is a device that generates code lists corresponding for the non-smooth functions χ (If-then-else branches), $|\cdot|$, and max.

The following formulas define the floating point and interval evaluations, and the slope evaluation of piecewise continuous functions. Example 3 illustrates their application.

Formula 1 *Floating point evaluation*

$$
\chi(x_s, x_q, x_r) = \begin{cases} x_q, & x_s < 0, \\ x_r, & otherwise. \end{cases} \tag{6}
$$

Formula 2 *Interval evaluation*

$$
\chi(\boldsymbol{x}_s, \boldsymbol{x}_q, \boldsymbol{x}_r) = \begin{cases} \boldsymbol{x}_q, & \boldsymbol{x}_s < 0; \\ \boldsymbol{x}_r, & \boldsymbol{x}_s > 0; \\ \boldsymbol{x}_q \underline{\cup} \boldsymbol{x}_r & 0 \in \boldsymbol{x}_s. \end{cases} \tag{7}
$$

Formula 3 *Interval evaluation of the slope* $\boldsymbol{s}(\chi, \boldsymbol{x}, \check{\boldsymbol{x}})$ *when* χ *is continuous in* x_s.

$$
\boldsymbol{s}(\chi(x_s, x_q, x_r), \boldsymbol{x}, \check{\boldsymbol{x}}) =
$$

$$
\begin{cases} \boldsymbol{s}(x_q, \boldsymbol{x}, \check{\boldsymbol{x}}), & \boldsymbol{x}_s \underline{\cup} \check{\boldsymbol{x}}_s < 0; \\ \boldsymbol{s}(x_r, \boldsymbol{x}, \check{\boldsymbol{x}}), & \boldsymbol{x}_s \underline{\cup} \check{\boldsymbol{x}}_s > 0; \\ \boldsymbol{s}(x_q, \boldsymbol{x}, \check{\boldsymbol{x}}) \underline{\cup} \boldsymbol{s}(x_r, \boldsymbol{x}, \check{\boldsymbol{x}}) & 0 \in \boldsymbol{x}_s \underline{\cup} \check{\boldsymbol{x}}_s. \end{cases} \tag{8}
$$

Example 3 Let $f : \mathbb{R} \to \mathbb{R}$ be the function defined by

$$
f(x) = \begin{cases} (x-1)(x-4), & x \leq 2, \\ \max\{2x - 8, |x - 2| - 2\}, & x > 2, \end{cases}
$$

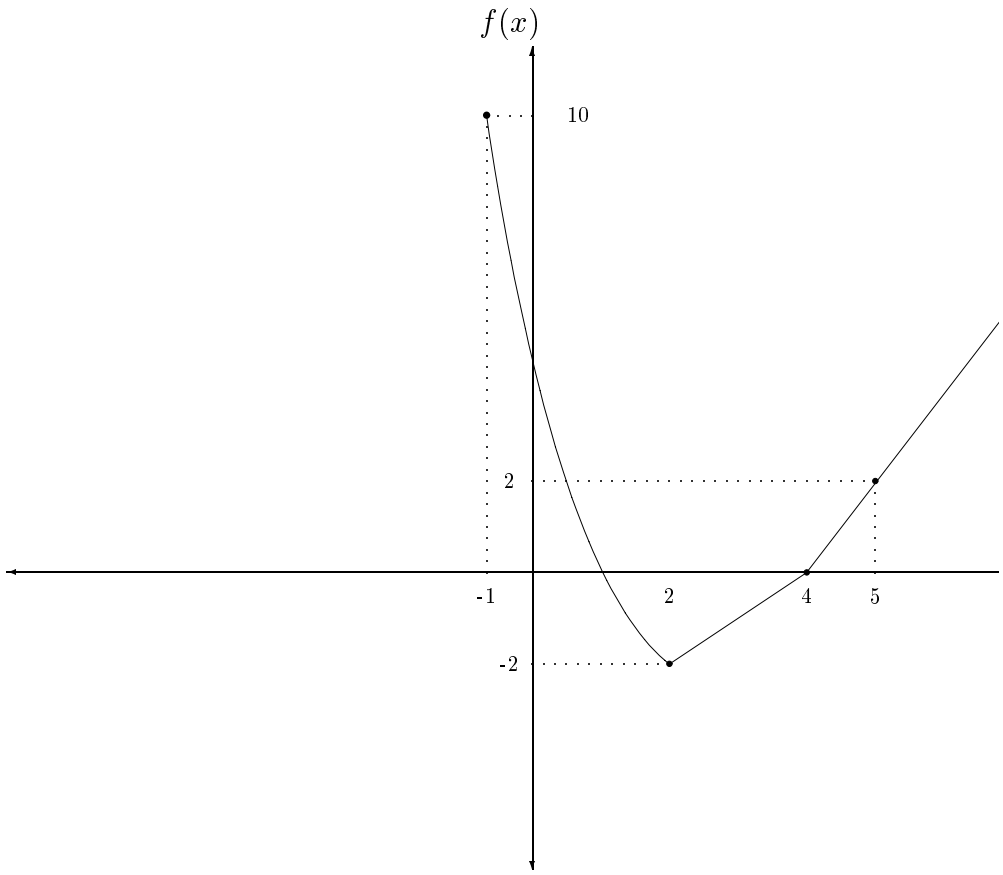

Figure 1: Graph of the function in Example 3

Since the graph of f is convex over the interval $\boldsymbol{x} = [-1, 5]$, the interval slope set is easily evaluated using Eq. (7) and (8), see Figure 1. From Eq. (7) an interval evaluation of f over $\boldsymbol{x} = [-1, 5]$ with $\check{x} = 2$, is $[-2.25, 10]$, while the exact range is $[-2, 10]$. From Eq. (8), the interval slope of f over $\boldsymbol{x} = [-1, 5]$ is

$$\begin{aligned} \mathbf{s}(f(x), \boldsymbol{x}, \check{x}) &= \mathbf{s}(\boldsymbol{x}_q, \boldsymbol{x}, \check{x}) \underline{\cup} \mathbf{s}(\boldsymbol{x}_r, \boldsymbol{x}, \check{x}) \\ &= [-4, 2] \underline{\cup} [-1, 2] = [-4, 2]. \end{aligned}$$

4. NUMERICAL RESULTS

This section presents an application of interval Newton techniques to a problem in thermodynamics, viz. the estimation from binary vapor liquid equilibrium (VLE) data of the energy parameters in the Wilson equation

$$\begin{aligned} \frac{g^E}{RT} &= -x_1 \ln(x_1 + \Lambda_{12} x_2) - x_2 \ln(x_2 + \Lambda_{21} x_1) \\ &= x_1 \ln \gamma_1 + x_2 \ln \gamma_2, \end{aligned} \tag{9}$$

where g^E is the molar excess Gibbs energy for a binary system, x_1 and x_2 the liquid-phase mole fractions.

From Eq. (9) expressions for the activity coefficients γ_1 and γ_2 are

$$\begin{aligned} \ln \gamma_1 &= x_2 \left[\frac{\Lambda_{12}}{x_1 + \Lambda_{12} x_2} - \frac{\Lambda_{21}}{x_2 + \Lambda_{21} x_1} \right], \\ &\quad - \ln(x_1 + \Lambda_{12} x_2) \end{aligned} \tag{10}$$

$$\begin{aligned} \ln \gamma_2 &= -x_1 \left[\frac{\Lambda_{12}}{x_1 + \Lambda_{12} x_2} - \frac{\Lambda_{21}}{x_2 + \Lambda_{21} x_1} \right], \\ &\quad - \ln(x_2 + \Lambda_{21} x_1) \end{aligned} \tag{11}$$

where the binary parameters Λ_{12} and Λ_{21} in Eq. (10) and Eq. (11) are given by

$$\Lambda_{12} = \frac{v_2}{v_1} exp\left[\frac{-\theta_1}{RT} \right], \tag{12}$$

$$\Lambda_{21} = \frac{v_1}{v_2} exp\left[\frac{-\theta_2}{RT} \right], \tag{13}$$

where 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

$$\gamma_{i,exp} = \frac{y_{i,exp} P_{exp}}{x_{i,exp} P_i^0}, \qquad i = 1, 2, \tag{14}$$

where P_i^0 is the vapor pressure of pure component i at the system temperature T. Two VLE models were considered, where the constants R and T in Eq. (12) and Eq. (13), and the experimental data in Eq. (14) were taken from *Gmehling et al.* [4]. By using interval arithmetic techniques, the minimization of the two objective functions for LS in Eq. (1) and LAV in Eq. (2) was compared, with $y_{ji} = \gamma_{ji,exp}$, $f_i(x_j, \theta) = \gamma_{ji,calc(\theta)}$, and the relative errors are

$$\frac{y_{ji} - f_i(x_j, \theta)}{y_{ji}} = \frac{\gamma_{ji,exp} - \gamma_{ji,calc(\theta)}}{\gamma_{ji,exp}}, \tag{15}$$

for $j = 1, \ldots, n$, and $i = 1, 2$, where $\gamma_{ji,calc(\theta)}$ are calculated from Eq. (9) at the same conditions (temperature, pressure and composition) used to measure $\gamma_{ji,exp}$.

Gau et al. [3] studied the relative errors in Eq. (15) with the LS estimate in Eq. (1) with interval arithmetic. We extend their procedure to non-smooth optimization by incorporating the LAV estimate in Eq. (2) using interval Newton techniques. The following two tables present estimates of the energy parameters, θ_1 and θ_2. $\theta^{(LS)}$ and $\theta^{(LAV)}$ denote the solutions for the objective functions defined by Eq. (1), and Eq. (2) respectively, and $\theta^{(D)}$ is the solution published in *Gmehling et al.* [4] by using the Newton method. The second and third columns show the evaluation of $\phi_{LAV}(\theta)$ and $\phi_{LS}(\theta)$ in the three solutions.

Table 1: Results for $T = 30°$ C

Solution	$\phi_{LAV}(\theta)$	$\phi_{LS}(\theta)$
$\theta^{(\mathbf{LAV})} = (\mathbf{-454.1, 1255})$	**0.3639**	0.0130
$\theta^{(LS)} = (-468.5, 1320)$	**0.3758**	0.0118
$\theta^{(D)} = (437, -437)$	**0.7139**	0.0383

Table 2: Results for $T = 50°$ C

Solution	$\phi_{LAV}(\theta)$	$\phi_{LS}(\theta)$
$\theta^{(\mathbf{LAV})} = (\mathbf{-388.2, 861.4})$	**0.2923**	0.0093
$\theta^{(LS)} = (-417.9, 969.3)$	**0.3130**	0.0081
$\theta^{(D)} = (342, -342)$	**0.6838**	0.0426

5. CONCLUSIONS

As both tables show, the evaluations of the objective functions $\phi_{LAV}(\theta)$ and $\phi_{LS}(\theta)$ at θ^{LAV} and θ^{LS} are lower than their corresponding evaluations at θ^D, the published result in [4]. Using interval methods, we found that the published solution corresponds to local optima, whereas interval methods enabled us to find the global optimal parameter values. Significantly, using Interval Arithmetic, both LS and LAV estimators can be used to obtain globally optimal solutions.

The approach presented here can also be used in connection with other objective functions, such as the maximum likelihood. Note that in the above example, differentiable functions (LS) and non-differentiable functions (LAV) were considered using this technique. Thus, it can be used in a wide range of applications.

We shall be applying interval arithmetic methods to more complex parameter estimation problems, such as those in fitness and health sciences or in the Hypoxia model for the Gulf Coast of Mexico [9], and derive more accurate estimations than those obtained by other means. We would like to know under what circumstances LS and LAV estimators perform best, and their sensitivity to the influence of different kinds of outliers.

ACKNOWLEDGMENTS

The authors would like to thank the editors and anonymous reviewers for their assistance. This work is partially supported by the Louisiana Board of Regents under Contract Number LEQSF (2006-2010)-ENH-TR-64 awarded to Southern University and A&M College, Baton Rouge, Louisiana.

REFERENCES

[1] F. H. Clarke, **Optimization and Nonsmooth Analysis**, Classics in Applied Mathematics, Philadelphia: Society for Industrial and Applied Mathematics, 1990.

[2] T. Dielman and E. Rose, "A Note on Hypothesis Testing in LAV Multiple Regression: A Small Sample Comparison", **Computational Statistics & Data Analysis**, Vol. 23, 1997, pp. 381–388.

[3] C. Y. Gau, J. F. Brennecke, and M. A. Stadtherr, "Reliable Nonlinear Parameter Estimation in EVL", **Fluid Phase Equilibria**, Vol. 168, No. 1, 2000, pp. 1–18.

[4] J. Gmehling, U. Onken, and W. Arlt, **Vapor-liquid Equilibrium Data Collection**, Chemistry Data Series, Vol. 1, Frankfurt/Main, Germany: Dechema, 1977–1996.

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

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

[7] R. E. Moore, **Interval Arithmetic and Automatic Error Analysis in Digital Computing**, PhD Dissertation, Stanford University, 1962.

[8] H. Muñoz and B. Kearfott, "Slope Intervals, Generalized Gradients, Semigradients, Slant Derivatives and Csets", **Reliable Computing**, Vol. 10, No. 3, 2004, pp. 163–193.

[9] H. Muñoz, D. Moreman, S. Bai, and N. Gwee, "A Generalized Hypoxia Model for the Gulf Coast of Mexico", **81st Annual Meeting of the Louisiana Academy of Sciences**, March 16, 2007.

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