

Interval Arithmetic Based Optimization in Nonlinear Regression

Antanas ŽILINSKAS Julius ŽILINSKAS

*Institute of Mathematics and Informatics
Akademijos 4, LT-08663 Vilnius, Lithuania
e-mail: antanasz@ktl.mii.lt, julius.zilinskas@mii.lt*

Received: April 2009; accepted: February 2010

Abstract. The optimization problems occurring in nonlinear regression normally cannot be proven unimodal. In the present paper applicability of global optimization algorithms to this problem is investigated with the focus on interval arithmetic based algorithms.

Keywords: global optimization, interval arithmetic, nonlinear regression, nonlinear least squares.

1. Introduction

Least squares regression (LSR) is among the most frequently used statistical methods. LSR algorithms are comprised of two main constituents: a minimization algorithm and a procedure of statistical analysis. In this paper we consider optimization problems occurring in LSR, specifically we focus on nonlinear LSR; optimization in the linear case is not a concern since it can be reduced to the solution of a system of linear equations. The theoretical and algorithmic achievements in nonlinear programming greatly influenced the development of software for solution of problems of nonlinear LSR. The contemporary statistical packages contain algorithms of nonlinear LSR including efficient subroutines for local minimization of sums of squared residuals. Besides of algorithms in commercial and shareware statistical packages there are also available local minimization algorithms for special problems of nonlinear LSR, e.g., related to dynamical system (Schittkowski, 2002). However, in applications the multimodal objective functions of residuals are not so rare to be ignored, moreover their unimodality can be proven only in exceptional cases. Therefore investigation of the applicability of the available global optimization algorithms to the problems of nonlinear LSR is an urgent problem.

We briefly discuss the advantages/disadvantages of various global optimization algorithms with respect to the properties of the objective functions in the problems of nonlinear LSR. The discussion is based on general properties of algorithms of global optimization (Törn and Žilinskas, 1989), experience of solution of nonlinear LSR problems by means of standard techniques (Seber and Wild, 2003), and several typical examples of application of global optimization algorithms for the solution of such problems (Dorsey and Mayer, 1995; Goffe *et al.*, 1994; Jerrell, 1997; Křivý *et al.*, 2000; Žilinskas and

Bogle, 2006). It is concluded that efficiency of interval arithmetic based global optimization algorithms are worth to be investigated. The subsequent part of the paper is devoted to the experimental investigation of the efficiency of well recognized interval arithmetic software in solution of proven difficult problems of nonlinear LSR.

2. Problem Statement

The optimization problem in nonlinear LSR is formulated as follows

$$\min_{X \in A} \sum_{i=1}^m (y_i - \varphi(X, Z_i))^2 = \min_{X \in A} f(X), \quad (1)$$

where the measurements y_i at the points $Z_i = (z_{1i}, z_{2i}, \dots, z_{pi})$ should be tuned by the nonlinear function $\varphi(X, \cdot)$. A typical example of the regression function is

$$\varphi(X, z) = x_1 \exp(-x_3 z) + x_2 \exp(-x_4 z),$$

which is used as a test problem (Křivý *et al.*, 2000), and as a model of a real world problem (Griffiths and Hunt, 1991).

The minimization problem (1) seems favourable for application of classical nonlinear programming techniques: normally the number of variables (equal to the number of model parameters to be estimated) is small, and the objective function is smooth. Indeed, many well developed nonlinear programming techniques can be applied to find a local minimizer of (1). A practical problem could be solved easily using algorithms from many available packages if a starting point for local descent would be known in the region of attraction of the global minimizer. However, such a starting point frequently is not known, and (1) should be considered as a global optimization problem. This general theoretical argumentation is supported by the experimental results (Křivý *et al.*, 2000) where 14 regression functions have been considered. Křivý *et al.* (2000) have shown that the standard algorithms from the statistical packages NCSS, SYSTAT, S-PLUS, SPSS for the large percentage of random starting points failed to find the global minimizer.

Neither minimax nor in average optimal algorithms are favourable for minimization of (1) since the intrinsic complexity of these algorithms is much higher than the complexity of computation of (1). The branch and bound type algorithms using underestimates based on global constants seem not competitive here since behaviour of $f(X)$ in different subsets of A can be very different, e.g., for small step in the parameter space the gradient norm can change more than by several orders. Adaptive versions of such algorithms can be appropriate for these problems, and hopefully they will be investigated in detail later. Randomized heuristics is a popular approach for various difficult optimization problems. There was a few attempts to attack the considered here optimization problems by randomized heuristics (Dorsey and Mayer, 1995; Goffe *et al.*, 1994; Křivý *et al.*, 2000; Žilinskas and Bogle, 2006). Although the latter results seem promising the considered methods

suffer from known general disadvantages of randomized heuristics. The strongest competitors of randomized heuristics for the considered problems seem interval arithmetic based global optimization methods as a natural complement to classical nonlinear programming methods for multimodal problems.

Let us note that the optimization problem related to the other technique of data analysis, namely to multidimensional scaling, is defined by the formula similar to (1). This optimization problem appeared indeed difficult, e.g., Mathar and Žilinskas (1994) proposed test functions based on this problem presenting real challenge to general global optimization algorithms. Our experience as well as experience of our colleagues shows that for the successful solution of these problems special sophisticated combinations of different techniques are needed (Karbauskaitė and Dzemyda, 2009; Žilinskas and Žilinskas, 2006; 2007; 2008; 2009; Žilinskas, 2008b). It seems likely that the hybridization idea can be appropriate also in construction of algorithms for nonlinear LSR. Therefore we are going not only to evaluate the efficiency of interval arithmetic based global optimization methods in solution of nonlinear LSR problems but also to investigate possibilities of the further development of hybrid algorithms for efficient solution of these problems.

3. Interval Arithmetic Based Algorithm

The objective function (1) seems favourable for application of interval arithmetic based global optimization algorithms: the function is defined by a rather simple analytical expression, the number of variables is small, and the formulas of derivatives are available. The extensive testing shows that for functions with such properties interval methods are efficient, and their application in such cases can be recommended (Hansen and Walster, 2004). The only difference of the problem (1) from the test problems used in experiments justifying high efficiency of interval methods is the number of summands: the number of summands in (1) typically is no less than 10. This at first glance inconsiderable difference, however can be crucial for the optimization efficiency because of increasing of the dependence (Hansen and Walster, 2004). Theoretical evaluation of the influence of dependence is difficult, and it should be investigated experimentally.

For the experiments the application of a well established software is desirable. A global optimization algorithm from the C++ toolbox for verified computing CToolbox (Hammer *et al.*, 1995) has been chosen. The standard for interval methods branch and bound approach is implemented using the branching tree constructed by bisecting (through the middle point of the longest side) the promising sub-regions (multidimensional interval boxes). Bounds of the function values are estimated using interval arithmetic in standard and centered forms. Monotonicity and concavity tests and interval Newton method (Hansen and Walster, 2004) are used to discard nonpromising sub-regions, where the objective function is monotone or concave, or where there are no stationary points. Intervals of first and second derivatives are evaluated using automatic differentiation. In the experiments the parameter of the algorithm – tolerance was set to 10^{-8} .

No criteria of performance of global optimization algorithms (time of optimization and number of function evaluations) are measured in CToolbox. The package has been

edited to provide these criteria. Time is measured in CPU time in seconds. The optimization is stopped if the time limit of 3600 s is exhausted. The interval objective function with automatic differentiation should be defined for the package. The number of evaluations of this function (NFE) is another measured criterion of performance.

The package uses interval library `fi_lib` (Hofschuster and Krämer, 1997) and C++ class library for extended scientific computing C-XSC (Hofschuster *et al.*, 2001). Žilinskas (2005) has shown that `fi_lib` library is the fastest and most accurate of freely available libraries implementing interval arithmetic.

Let us note that the function “interval to the power of interval” is not implemented as a function with automatic differentiation in CToolbox. Extended interval arithmetic allowing to evaluate interval function over an interval containing points where the corresponding real value function is not defined (division by interval containing 0 or logarithm of interval containing 0) is not supported as well. These deficiencies should be taken into account when designing the experiments.

4. Testing Results

The known difficult cases have been considered for testing. Fourteen regression functions with example data (Tvrđík, 2010) are used. Křivý *et al.* (2000) have shown that the standard algorithms from the statistical packages NCSS, SYSTAT, S-PLUS, SPSS are not reliable to find the solution from random starting points.

Because of the absence of power function and extended interval arithmetic in CToolbox, the feasible regions of the 10th, 12th and 14th test problems had to be changed to exclude the possibility for some variables to take the value 0. Therefore the original feasible intervals for the variables x_3 and x_4 in 10th test problem, for the variable x_3 in 12th test problem, and for the variables x_2 and x_3 in 14th test problem have been shifted to the right by 1; e.g., the initial interval $[0, 5]$ has been replaced with the interval $[1, 6]$.

The experiments have been performed on HP Compaq 6710b notebook computer with Intel Centrino processor, 2GB memory and Ubuntu LINUX. GNU C compiler 4.4.1. has been used to build `cxsc-2-4-0` package and developed programs. The results of optimization using CToolbox are presented in Table 1. Three test problems have been solved successfully: both two-dimensional problems and one three-dimensional problem. The feasible regions of these problems are relatively small, and variables are properly scaled. The other problems have not been solved with the prescribed accuracy during the allowed time limit of 3600 s. The guaranteed requested accuracy was not achieved because of too many active boxes obtained by bisection of the longest edge where the length of the initial longest edge in the not successful cases is much longer than in the successful cases.

The other algorithm used in our experiments was implementation of the interval global optimization method combined with searches implemented in real number arithmetic. This global unconstrained minimization method involves a combination of local search, branch-and-bound technique and interval arithmetic (Jansson and Knüppel, 1995). In this method derivatives are not required. The method has been implemented in

Table 1
Optimization of test problems using CToolbox algorithm

No.	Time (s)	NFE	Minimum
1	3600	718,757	Not finished
2	3600	1,723,972	Not finished
3	3600	1,186,617	Not finished
4	3600	1,191,193	Not finished
5	3600	789,384	Not finished
6	0.00	661	124.3622
7	3600	656,474	Not finished
8	3600	672,142	Not finished
9	0.54	13,570	8.896301×10^{-3}
10	3600	555,223	Not finished
11	3600	700,494	Not finished
13	3600	659,215	Not finished
14	3600	1,644,024	Not finished
12	1.29	17,454	4.375281×10^{-3}

PROFIL/BIAS (Knüppel, 1994; 1999). Numerical results for well-known problems and comparisons with other methods are available (Jansson and Knüppel, 1995).

Both tested interval global optimization algorithms did not find solutions for majority of problems with requested guaranteed accuracy during the allowed time limit. The detailed results of PROFIL/BIAS algorithm are not given because of limited space. CToolbox algorithm is faster than PROFIL/BIAS algorithm for simplest problems providing guaranteed solutions. Monotonicity and concavity tests and interval Newton method help to discard nonpromising sub-regions more early. The advantage of PROFIL/BIAS algorithm is its relatively good approximations found more often than by CToolbox algorithm, even when the global minimum is not found with prescribed accuracy. The use of local searches provides the opportunity for this. It seems that the algorithm using local searches as well as monotonicity and concavity tests and interval Newton method could have both the advantages. Objective function evaluation in PROFIL/BIAS algorithm is much faster, because interval objective function with automatic differentiation in CToolbox algorithm is much more expensive than the interval and real test functions without differentiation.

5. Testing Results for Modified Problems

The results of the previous section suggest that only the problems of low dimensionality can be expected to be solved with guaranteed accuracy. In case the regression function contains linear and nonlinear parameters the optimal values of the former can be found analytically reducing the dimensionality of the objective function. For example, the re-

gression function of cases 2 and 3

$$\varphi(X, Z) = x_3 (\exp(-x_1 z_1) + \exp(-x_2 z_2))$$

contains the linear parameter x_3 whose optimal value can be expressed as

$$b_3 = \frac{\sum y_i (\exp(-x_1 z_{1i}) + \exp(-x_2 z_{2i}))}{\sum (\exp(-x_1 z_{1i}) + \exp(-x_2 z_{2i}))^2},$$

implying the presentation of the regression function by the following formula

$$\varphi(X, Z) = b_3 (\exp(-x_1 z_1) + \exp(-x_2 z_2)).$$

Similarly, in the cases 7 and 11, the optimal values of the linear parameters x_1 and x_2 of the regression function

$$\varphi(X, z) = x_1 \exp(x_3 z) + x_2 \exp(x_4 z),$$

can be expressed as

$$b_1 = \frac{(\sum y_i \exp(x_3 z_i)) \times (\sum \exp(2x_4 z_i)) - (\sum y_i \exp(x_4 z_i)) \times (\sum \exp((x_3 + x_4) z_i))}{(\sum \exp(2x_3 z_i)) \times (\sum \exp(2x_4 z_i)) - (\sum \exp((x_3 + x_4) z_i))^2}$$

and

$$b_2 = \frac{(\sum y_i \exp(x_4 z_i)) \times (\sum \exp(2x_3 z_i)) - (\sum y_i \exp(x_3 z_i)) \times (\sum \exp((x_3 + x_4) z_i))}{(\sum \exp(2x_3 z_i)) \times (\sum \exp(2x_4 z_i)) - (\sum \exp((x_3 + x_4) z_i))^2}$$

with the subsequent reduction of the original four-dimensional problem to a two-dimensional problem:

$$\varphi(X, z) = b_1 \exp(x_3 z) + b_2 \exp(x_4 z).$$

Eleven problems have been modified in this way, except of 6th and 9th who have no linear parameters, and 13th who is too complicated for the considered modification.

The results of optimization of the modified test problems using CToolbox interval global optimization algorithm are shown in Table 2. The objective functions of all modified test problems have interval division. Optimization of several modified test problems cause interval errors because of division by interval containing 0. The results of these problems are not shown in Table 2. It seems that it would be possible to avoid symmetries of objective functions and sub-regions of feasible region containing division by zero in modified objective functions by setting linear constraints and using simplicial partitioning (Paulavičius and Žilinskas, 2007; 2008; Žilinskas, 2007; 2008a; Žilinskas and Žilinskas, 2002). However this requires further investigation.

The modified 3rd test problem has not been solved during allowed time limit seemingly because of different scales of variables. The modification of 12th test problem have made optimization twice faster by means of CPU time and 2.7 times faster by means of the number of objective function evaluations.

Table 2
Optimization of modified test problems using CToolbox algorithm

No.	Time (s)	NFE	Minimum
1	0.58	11,572	4.355266×10^{-5}
2	173.49	398,660	7.471221×10^{-5}
3	3600	1,054,775	Not finished
12	0.68	6507	4.375281×10^{-3}

6. Influence of Scaling

Different scales of variables make difficulties for interval global optimization algorithms especially when sub-regions are subdivided through the middle points of the longest sides. If the feasible interval of one variable is k times larger than of the others and sub-regions are bisected through the middle of the longest sides, $m = 2^{\lceil \log_2(k) \rceil} - 1$ bisections through this variable and $2m + 1$ evaluations of sub-regions should be done before any bisections through other variables. In the case of the 7th test problem, with very differently scaled variables $([0, 10^8], [0, 10^8], [-2, 0], [-5, 0])$, $2^{\lceil \log_2(5 \times 10^7) \rceil} - 1$ bisections through the first variable, the same number of bisections through the second variable and 3 bisections through the fourth variable – the total number $m = (2^{26} - 1) \times (2^{26} - 1) \times (2^2 - 1)$ bisections and approximately 2^{57} evaluations of sub-regions should be done before bisection through the third variable. Since this is not feasible practically, only bisections through two first variables is performed. To make the scales of variables similar we use feasible intervals $[0, 1]$ for all variables and use coefficients in the objective functions to scale variables back to get the original problems.

The results of optimization of the scaled test problems using CToolbox interval global optimization algorithm is shown in Table 3. Changing of scales have helped to solve 2nd test problem and 3rd modified test problem, which were not solved without scaling. Since 2nd test problem has been solved with scaling it is possible to estimate impact of modification for this test problem. The modification have made optimization at least 13 times faster by means of CPU time and 5 times faster by means of the number of objective function evaluations. Both modification and scaling of variables were required to solve 3rd test problem.

7. Conclusions

The results of experiments with difficult cases of nonlinear regression show that non-sophisticated application of interval arithmetic based global optimization methods is not promising. Only a couple of two-dimensional problems and one three-dimensional problem have been solved with guaranteed accuracy. The success of the guaranteed solution crucially depends on scaling of parameters and analytical reformulation of the problem in the case of presence of linear parameters. Analytical reformulation made solution of

Table 3
Optimization of scaled test problems using CToolbox algorithm

No.	Original formulation			Modified problems		
	Time (s)	NFE	Minimum	Time (s)	NFE	Minimum
1	3600	726,149	Not finished	0.59	11,572	4.355266×10^{-5}
2	72130	1,981,543	7.471221×10^{-5}	164.84	398,732	7.471221×10^{-5}
3	3600	871,086	Not finished	79.00	150,486	1.251892
4	3600	1,502,698	Not finished			
5	3600	778,363	Not finished			
6	0.00	661	124.3622			
7	3600	659,066	Not finished			
8	3600	646,724	Not finished			
9	0.56	13,570	8.896301×10^{-3}			
10	3600	541,437	Not finished			
11	3600	1,022,742	Not finished			
12	1.36	17,454	4.375281×10^{-3}	0.72	6507	4.375281×10^{-3}
13	3600	659,328	Not finished			
14	3600	1,659,838	Not finished			

problems from 2 to 13 times faster and together with scaling of variables enabled solution with guaranteed accuracy of three problems which were not solved in original formulation during reasonable time. The experience with interval methods can be valuable in development of other branch and bound type algorithms using adaptive underestimates of objective function.

Acknowledgments. The research is partially supported by the Agency for International Science and Technology Development Programmes in Lithuania through COST programme.

References

- Dorsey, R.E., Mayer, W.J. (1995). Genetic algorithms for estimation problems with multiple optima, nondifferentiability, and other irregular features. *Journal of Business and Economic Statistics*, 13(1), 53–66.
- Goffe, W., Ferrier, G., Rogers, J. (1994). Global optimization of statistical functions with simulated annealing. *Journal of Econometrics*, 60, 65–99.
- Griffiths, J.D., Hunt, J.G. (1991). Vehicle headways in urban areas. *Traffic Engineering and Control*, 32(10), 458–462.
- Hammer, R., Hocks, M., Kulish, U., Ratz, D. (1995). *C++ Toolbox for Verified Computing: Basic Numerical Problems*. Springer.
- Hansen, E., Walster, G.W. (2004). *Global Optimization Using Interval Analysis*, 2nd ed. Dekker, New York.
- Hofschuster, W., Krämer, W. (1997). A fast public domain interval library in ANSI C. In: Sydow, A. (Ed.), *Proceedings of the 15th IMACS World Congress on Scientific Computation, Modelling and Applied Mathematics*. Vol. 2. Berlin, pp. 395–400.
- Hofschuster, W., Krämer, W., Wedner, S., Wiethoff, A. (2001). *C-XSC 2.0 – A C++ Class Library for Extended Scientific Computing*. Preprint 2001/1, Universität Wuppertal.

- Jansson, C., Knüppel, O. (1995). A branch-and-bound algorithm for bound constrained optimization problems without derivatives. *Journal of Global Optimization*, 7(3), 297–331.
- Jerrell, M. (1997). Automatic differentiation and interval arithmetic for estimation of disequilibrium models. *Computational Economics*, 10, 295–316.
- Karbauskaitė, R., Dzemyda, G. (2009). Topology preservation measures in the visualization of manifold-type multidimensional data. *Informatica*, 20(2), 235–254.
- Knüppel, O. (1994). PROFIL/BIAS – A fast interval library. *Computing*, 53(3–4), 277–287.
- Knüppel, O. (1999). *PROFIL/BIAS V 2.0*. Report 99.1, Technische Universität Hamburg-Harburg.
- Křivý, I., Tvrđík, J., Krepec, R. (2000). Stochastic algorithms in nonlinear regression. *Computational Statistics and Data Analysis*, 33, 277–290.
- Mathar, R., Žilinskas, A. (1994). A class of test functions for global optimization. *Journal of Global Optimization*, 5, 195–199.
- Paulavičius, R., Žilinskas, J. (2007). Analysis of different norms and corresponding Lipschitz constants for global optimization in multidimensional case. *Information Technology and Control*, 36(4), 383–387.
- Paulavičius, R., Žilinskas, J. (2008). Improved Lipschitz bounds with the first norm for function values over multidimensional simplex. *Mathematical Modelling and Analysis*, 13(4), 553–563.
- Schittkowski, K. (2002). *Data Fitting in Dynamical Systems*. Kluwer, Dordrecht.
- Seber, G., Wild, C. (2003). *Nonlinear Regression*. Wiley, Hoboken.
- Törn, A., Žilinskas, A. (1989). Global Optimization. *Lecture Notes in Computer Science*, 350, 1–255.
- Tvrđík, J. (2010). *Global Optimization, Evolutionary Algorithms and Their Application to Computational Statistics*, http://albert.osu.cz/tvrdik/down/global_optimization.html. Last visited 2010-02-04.
- Žilinskas, A., Žilinskas, J. (2002). Global optimization based on a statistical model and simplicial partitioning. *Computers & Mathematics with Applications*, 44(7), 957–967.
- Žilinskas, A., Žilinskas, J. (2006). Parallel hybrid algorithm for global optimization of problems occurring in MDS based visualization. *Computers & Mathematics with Applications*, 52(1–2), 211–224.
- Žilinskas, A., Žilinskas, J. (2007). Two level minimization in multidimensional scaling. *Journal of Global Optimization*, 38(4), 581–596.
- Žilinskas, A., Žilinskas, J. (2008). A hybrid method for multidimensional scaling using city-block distances. *Mathematical Methods of Operations Research*, 68(3), 429–443.
- Žilinskas, A., Žilinskas, J. (2009). Branch and bound algorithm for multidimensional scaling with city-block metric. *Journal of Global Optimization*, 43(2–3), 357–372.
- Žilinskas, J. (2005). Comparison of packages for interval arithmetic. *Informatica*, 16(1), 145–154.
- Žilinskas, J. (2007). Reducing of search space of multidimensional scaling problems with data exposing symmetries. *Information Technology and Control*, 36(4), 377–382.
- Žilinskas, J. (2008a). Branch and bound with simplicial partitions for global optimization. *Mathematical Modelling and Analysis*, 13(1), 145–159.
- Žilinskas, J. (2008b). On dimensionality of embedding space in multidimensional scaling. *Informatica*, 19(3), 447–460.
- Žilinskas, J., Bogle, I.D.L. (2006). Balanced random interval arithmetic in market model estimation. *European Journal of Operational Research*, 175(3), 1367–1378.

A. Žilinskas is the chairman of Optimization Sector and a principal researcher at the Institute of Mathematics and Informatics, Lithuania. His scientific interests are statistical theory of global optimization, optimization based modelling and design, and analysis of multidimensional data by means of visualization. He is a member of editorial boards of Journal of Global Optimization, Control and Cybernetics, Informatica, The Open Cybernetics and Systemics Journal, and International Journal of Grid and High Performance Computing.

J. Žilinskas is a principal researcher in Systems Analysis Department at the Institute of Mathematics and Informatics, Lithuania. His research interests include global optimization, parallel computing, data analysis and visualization. He is a member of editorial boards of Computational Management Science, Informatica, Mathematical Modelling and Analysis, and Optimization Letters.

Intervalų aritmetika pagrįstų optimizavimo metodų taikymas netiesinėje regresijoje

Antanas ŽILINSKAS, Julius ŽILINSKAS

Optimizavimo uždaviniai kylantys sudarant netiesinės regresijos modelius yra sudėtingi ta prasme, kad jų unimodalumo paprastai neįmanoma įrodyti. Šiame straipsnyje nagrinėjama galimybė spręsti tokius uždavinius globaliojo optimizavimo metodais koncentruojant dėmesį į intervalų aritmetika pagrįstus metodus.