Taylor Series Models in Deterministic Global Optimization

R. Baker Kearfott
Alvard Arazyan

ABSTRACT Deterministic global optimization requires a global search with rejection of subregions. To reject a subregion, bounds on the range of the constraints and objective function can be used. Although often effective, simple interval arithmetic sometimes gives impractically large bounds on the ranges. However, Taylor models as developed by Berz et al may be effective in this context. Efficient incorporation of such models in a general global optimization package is a significant project. Here, we use the system COSY-INFINITY by Berz et al to study the bounds on the range of various order Taylor models for certain difficult test problems we have previously encountered. Based on that, we conclude that Taylor models may be useful for some, but not all, problems in verified global optimization. Forthcoming improvements in the COSY-INFINITY interface will help us reach stronger conclusions.

1 Deterministic Global Optimization

Deterministic global optimization involves exhaustive search over the domain. The domain is subdivided ("branching"), and those subdomains that cannot possibly contain global minimizers are rejected. For example, if the problem is the unconstrained problem

$$\boxed{\begin{array}{ll}\text{Enclose the minimizers of} & \phi(x)\\ \text{subject to} & x \in \boldsymbol{x},\end{array}} \tag{1.1}$$

then evaluating ϕ at a particular point x gives an upper bound for the global minimum of ϕ over the region $\boldsymbol{x}$. Some method is then used to bound the range of ϕ over subregions $\tilde{\boldsymbol{x}} \subset \boldsymbol{x}$. If the lower bound $\underline{\phi}$, so obtained, for ϕ over $\tilde{\boldsymbol{x}}$ has $\underline{\phi} > \phi(x)$, then $\tilde{\boldsymbol{x}}$ may be rejected as not containing any global optima; see Figure 1 for the situation in one dimension.

A related problem is that of finding all roots within a given region, that is,

$$\boxed{\begin{array}{ll}\text{Enclose all } x \text{ with} & f(x) = 0\\ \text{subject to} & x \in \boldsymbol{x}.\end{array}} \tag{1.2}$$

In equation (1.2), bounds on f over a subregion $\tilde{\boldsymbol{x}} \subset \boldsymbol{x}$ are obtained; denote the interval vector representing such bounds by $\boldsymbol{f}(\boldsymbol{x})$. If $0 \notin \boldsymbol{f}(\boldsymbol{x})$, that is, unless the lower bound for each component of f is less than zero and the upper bound is greater than zero, then there cannot be a solution of $f(x) = 0$ in $\tilde{\boldsymbol{x}}$, and $\tilde{\boldsymbol{x}}$ can be rejected.

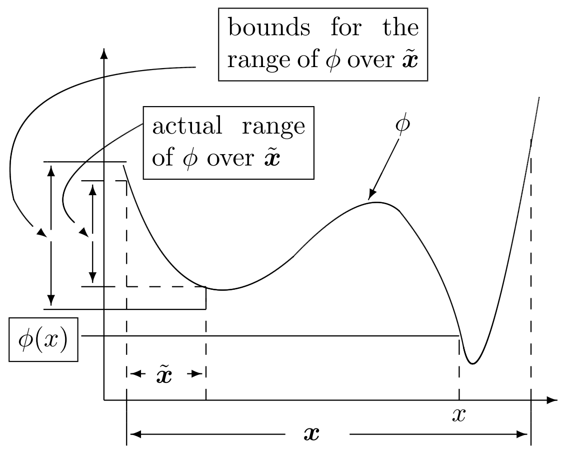

FIGURE 1. Rejecting $\tilde{\boldsymbol{x}}$ because of a high objective value

A simple interval evaluation $\boldsymbol{\phi}(\tilde{\boldsymbol{x}})$ (or $\boldsymbol{f}(\tilde{\boldsymbol{x}})$) of ϕ (or of f) over $\tilde{\boldsymbol{x}}$ is sometimes a practical way of obtaining the lower bound. (See [Han92], [Kea96], or a number of other introductory expositions.) However, there are some functions for which interval evaluation gives an extreme overestimation, and other techniques are necessary. One such function arises from Gritton's second problem, a chemical engineering model that J. D. Seader previously pointed out to us.

Example 1 *(Gritton's second problem) The eighteen real solutions of $f(x) = 0$ in $\boldsymbol{x} = [-12, 8]$ are sought, where f is defined by*

$$\begin{aligned} f(x) \quad = \quad & -371.93625x^{18} - 791.2465656x^{17} + 4044.944143x^{16} \\ & +978.1375167x^{15} - 16547.8928x^{14} + 22140.72827x^{13} \\ & -9326.549359x^{12} - 3518.536872x^{11} + 4782.532296x^{10} \\ & -1281.47944x^{9} - 283.4435875x^{8} + 202.6270915x^{7} \qquad (1.3) \\ & -16.17913459x^{6} - 8.88303902x^{5} + 1.575580173x^{4} \\ & +0.1245990848x^{3} - 0.03589148622x^{2} \\ & -0.0001951095576x + 0.0002274682229, \end{aligned}$$

(1.4)

This example can be treated by careful domain subdivision and use of point evaluations, as explained in [Kea97]. However, sharper bounds on the range would provide power that would make the code much simpler. Taylor models have shown promise for this.

2 Taylor Models and Global Optimization

An interval Taylor model in COSY-INFINITY for $\phi : \mathbb{R}^n \to \mathbb{R}$ is of the form

$$\phi(x) \in P_d(x - x_0) + \boldsymbol{I}_d, \tag{1.5}$$

where $P_d(x)$ is a degree-d polynomial in the n variables $x \in \mathbb{R}^n$, x_0 is a base point (often the midpoint of the interval vector $\boldsymbol{x}$), and $\boldsymbol{I}_d$ is an interval that encompasses the truncation error over the interval vector $\boldsymbol{x}$ and possible roundoff errors in computing the coefficients of P_d. Early work in interval computations did not indicate that Taylor models were promising. In particular, if one merely evaluated P_m with interval arithmetic over a box (i.e. over an interval vector) $\boldsymbol{x}$, then the difference between the width of $P_m(\boldsymbol{x}) + \boldsymbol{I}_d$ and the width of the actual range of ϕ over $\boldsymbol{x}$ decreases no faster than the square of the widths of $\boldsymbol{x}$, a rate that can already be achieved with $m = 2$. A higher convergence order can be achieved if the range of P_m can be estimated accurately, but computing such an estimation is NP-complete in the length of the expression defining P_m; see [KLRK98, Ch. 3 and Ch. 4].

However, Berz et al have found Taylor models to be highly effective at computing low-overestimation enclosures of the range of functions [MB99]. Berz' group has applied such models successfully to the analysis of stability of particle beams in accelerators [BH94], and has advocated its use for global optimization in general [MB99]; this, among other applications, is discussed in [MB00]. In an informal communication, Berz and Makino illustrated that the overestimation in Gritton's problem can be reduced by many orders of magnitude simply by approximating the degree-18 function with its degree-5 Taylor polynomial.

In summary, Taylor models, in principle, do not work, but, in practice, are effective. The effectiveness can be viewed as a type of symbolic preconditioning of the algebraic expression: If the domain widths are not excessive, then interval dependencies are reduced when the original expression is replaced by a Taylor model. This indicates that additional careful study of Taylor models in general deterministic global optimization algorithms is warranted.

3 Scope and Purpose of This Preliminary Study

During the past few years, with support from a SunSoft Cooperative Research and Development contract, we have gained experience with various practical problems within the GlobSol [Cor98, Kea96] interval global optimization package. Although successful with some problems, GlobSol, and another package Numerica [VHMD97], based on similar principles, cannot solve certain problems without subdividing the region $\boldsymbol{x}$ into an impractically large number of subregions. To solve such problems via verified global optimization, we need to identify the cause or causes of this algorithmic failure. These causes may include the following (among possibly other reasons).

(a) The stopping criteria for the subdivision process are inappropriate;

(b) the way that the boxes are subdivided (such as the method of selecting the coordinate to bisect in a bisection process) is inappropriate;

(c) the bounds on the ranges have an excessive amount of overestimation;

(d) there is something inherent in the mathematics of the equations, such as coupling between the components, that causes problems.

We have considered stopping criteria (item (a) above) in [Kea99], and feel we understand the mechanisms in most cases. We [Kea96, §4.3.2] and others (e.g. [Ber96, CR97, RC95]) have studied the criteria for subdividing the box (item (b) above). We have also determined that such subdivision criteria are best if consistent with the stopping criteria; such integrated subdivision and stopping criteria have already been implemented in GlobSol.

Certain inherent conditions, such as manifolds of solutions, can result in an impractically large number of subregions. In these cases, problem reformulation is probably necessary for higher dimensions, although use of more powerful equipment (for example, a sufficient number of parallel or distributed processors) may be appropriate if the number of variables is not too large. These cases may be hard to distinguish from overestimation on the bounds (item (c)), and probably need to be studied on a case-by-case basis.

Overestimation on bounds (item (c)) can potentially be handled using various tools within the algorithm itself. As mentioned above in §2, below in §4, and in [MB99, MB00], Taylor models can sometimes reduce overestimation considerably. However, the computational differentiation techniques in Taylor models, such as described in [Ber91, BH98], must be done efficiently to be practical in general global optimization algorithms. This differentiation process involves indexing schemes to reference the (generally) sparsely occurring non-zero terms from among the $(d+\nu)!/(d!\nu!)$ possible terms of a polynomial of degree d in ν variables [Ber91]. This and other implementation details make quality implementation of Taylor model computations a major task.

Berz et al have a good Taylor model implementation in COSY-INFINITY [BMS+96, Ber00]. Objective functions can be coded in the special COSY language, translated and interpreted within the package. In the experiments reported here, due to technical and licensing limitations, the objective had to be coded the COSY language, and called in a stand-alone mode. However, Jens Hoefkens has recently developed a Fortran 90 module for access to the COSY-INFINITY package. Future experiments will be easier and more comprehensive with this module.

A thorough test of Taylor arithmetic for general global optimization will need to integrate COSY-INFINITY computations within the global optimization algorithm, since the effect of range bound overestimation is different at different points (say, near a solution and far away from one), and since it is probably often advantageous to use local Taylor models specific to smaller subregions; this will be done with Hoefkens' module. However, because of the aforementioned limitations, we have proceeded in this paper as follows:

1. We have provided a simple translator that translates GlobSol's "code list" to the COSY-INFINITY language.
2. We have identified two interesting problems we have tried to solve with GlobSol.
3. We have computed Taylor model ranges for several expansion points and interval vector widths.

The goal of this study is to evaluate the potential usefulness of Taylor models in verified global optimization. In particular, we wish to know what order and degree are necessary. Can the same benefits be gotten by just implementing lower-order or is there a benefit of full generality? Also, how much of an impact is there on particular problems? Orders of magnitude difference in widths of range bounds for larger boxes would be useful, but small differences (perhaps less than a factor of 2) would be unimpressive, since Taylor arithmetic is more expensive than ordinary interval arithmetic.

4 Our Results

Our investigations to date have been with two examples.

4.1 Gritton's Second Problem

We initially tried Gritton's Example (1). One troublesome point is $\check{x} \approx 1.381$ with $f(\check{x}) = 0$. The range bounds over subintervals near this point should not contain zero, for such subintervals to be efficiently rejected. We tried base point $x_0 = 1.36$, and focused on the interval $\tilde{\boldsymbol{x}} = [1.35, 1.37]$, experimenting also with different widths. The result is Table 1.1. Here

TABLE 1.1. Widths of enclosing intervals for Example 1

Width	2	0.2	0.02
Degree 1	$3.46 \times 10^{+06}$	$1.76 \times 10^{+03}$	$1.27 \times 10^{+01}$
Degree 2	$1.08 \times 10^{+06}$	$7.03 \times 10^{+01}$	9.25×10^{-02}
Degree 3	$3.63 \times 10^{+05}$	$4.03 \times 10^{+00}$	4.29×10^{-02}
Degree 4	$1.16 \times 10^{+05}$	9.52×10^{-01}	4.27×10^{-02}
Degree 5	$4.68 \times 10^{+04}$	8.16×10^{-01}	4.27×10^{-02}
Degree 10	$3.20 \times 10^{+03}$	8.09×10^{-01}	4.27×10^{-02}
Simple interval	$4.10 \times 10^{+06}$	$3.22 \times 10^{+05}$	$2.65 \times 10^{+02}$
rastering	$2.64 \times 10^{+03}$	5.62×10^{-01}	4.03×10^{-02}

"rastering" is a heuristic method COSY-INFINITY uses to obtain inner bounds on the range: In these tables, the functions are evaluated at the end points of the component intervals and at three random values in the interior of each component interval; the minimum and maximum value so-obtained give a heuristic estimate for the range.

For Example (1), the Taylor model is definitely helpful. In particular, a straightforward interval calculation over $\tilde{\boldsymbol{x}} = [1.35, 1.37]$ gives $f(x) \in [-1381, 1384]$, whereas the Taylor model of degree 5 gives $f(x) \in [0.009023, 0.0431]$, close enough to the actual range of $[0.0111, 0.0431]$ to determine that there is no zero of f in $\tilde{\boldsymbol{x}}$. This contrasts sharply with the simple interval value, which is roughly 100,000 times too wide to be of use. In fact, the interval evaluation at $[1.3599995, 1.3600005]$, an interval of length 10^{-6}, contains the interval $[-.0436, .0939]$, whereas the interval evaluation at $[1.3599999, 1.3600001]$, an interval of length 2×10^{-7}, is contained in the interval $[.01137, .03882]$. Thus, intervals on the order of 2×10^{-7} are needed to reject portions of the region as far out as 10^{-2} from the root, whereas an interval of length 2×10^{-2} (or perhaps larger) can be rejected with a degree-5 Taylor model.

4.2 A Six-Dimensional Quartic

Neither GlobSol nor Numerica could solve the following six-dimensional polynomial system, whose components are of degree 4.

Example 2 *Find a_1, a_2, a_3, x_1, x_2, and x_3 such that $c_i = 0$, $i = 1, \ldots, 6$, where*

$$
\begin{aligned}
c_1 &= 0.08413r + 0.2163q_1 + 0.0792q_2 - 0.1372q_3, \\
c_2 &= -0.3266r - 0.57q_1 - 0.0792q_2 + 0.4907q_3
\end{aligned}
$$

$$
\begin{aligned}
c_3 &= 0.2704r + 0.3536\left(a_1(x_1 - x_3) + a_2(x_1^2 - x_3^2) + a_3(x_1^3 - x_3^3)\right. \\
&\qquad \left. + x_1^4 - x_3^4\right) \\
c_4 &= 0.02383p_1 - 0.01592r - 0.08295q_1 - 0.05158q_2 + 0.0314q_3 \\
c_5 &= -0.04768p_2 - 0.06774r - 0.1509q_1 + 0.1509q_3 \\
c_6 &= 0.02383p_3 - 0.1191r - 0.0314q_1 + 0.05158q_2 + 0.08295q_3, \quad \textit{where} \\
r &= a_1 + a_2 + a_3 + 1, \quad \textit{and} \\
p_i &= a_1 + 2a_2x_i + 3a_3x_i^2 + 4x_i^3, \quad i = 1, 2, 3, \\
q_i &= a_1x_i + a_2x_i^2 + a_3x_i^3 + x_i^4, \quad i = 1, 2, 3.
\end{aligned}
$$

In Example 2, we took center point $(a, x) = (1, 1, 1, 1, 1, 1)$, and took intervals of widths 1, 0.1, and 0.01 (equal widths in each direction) about this center point. The results appear in Table 1.2. In Table 1.2, we only

TABLE 1.2. Widths of enclosing intervals for c_1 for Example 2

Domain Width	1	0.1	0.01
Degree 1	9.6	0.54	0.0509
Degree 2	8.52	0.53	0.0508
Degree 3	8.52	0.53	0.0508
Degree 4	8.49	0.53	0.0508
Simple interval	6.85	0.59	0.0588
rastering	6.24	0.51	0.0505

display the results for c_1; the widths of the range bounds for the other five components behave similarly. From Table 1.2, it is clear for this problem that

- The simple interval computations do not have excessive overestimation, at least for this problem.
- There is probably not an advantage to using the Taylor model representations to compute interval range bounds on this particular problem, since they apparently do not lead to narrower-width enclosures.
- The Taylor model representations show order-1 convergence as the interval widths are decreased.

All in all, there is evidence that, in Example 2, the difficulties of GlobSol are probably not due to overestimation in function range bounds. Nonetheless application of Taylor models reveals behavior of Taylor models not apparent in Gritton's problem (Example 1) or in Makino's example [MB99].

Despite lack of advantage of interval computations in range estimation for the components in Example 2, Taylor models may still be useful for such problems. In particular, coupling between the equations (i.e. c_j and c_k both depend strongly on x_i for the same index i) plays a role in the difficulty. The equations are uncoupled with preconditioning. Hansen and others have pointed out that, for larger independent variable widths, the preconditioning is much more effective if it is done *symbolically*, before an interval evaluation is attempted. To symbolically precondition a system, the component functions need to be represented in terms of a basis and coefficients. The multivariate Taylor form provides such a basis.

Some Additional Remarks

We note that the public version of COSY-INFINITY at the time of this paper evaluates the polynomial part $P_d(x-x_0)$ by plugging the interval vector $\tilde{\boldsymbol{x}}$ into the expression and performing simple interval arithmetic. Also, since that version of COSY-INFINITY does not support the power function for all data types, x^2 is evaluated as $x \times x$, so that e.g. $[-1,1]^2$ evaluates to $[-1,1]$ rather than $[0,1]$. We have observed slightly sharper (but non-rigorous) Taylor bounds from computations with Mathematica compared with COSY-INFINITY; we attribute the difference to COSY's present treatment of the power function. Nonetheless, as evidenced in Makino's example (ibid.) and both examples presented here, in at least some cases, the Taylor model approach is either powerful without sharp bounding of the polynomial part or else sharply bounding the polynomial part would not help much. The COSY-INFINITY development group is presently testing better bounding procedures, including sharp treatment of even powers, use of linear trends, etc.

5 Some Conclusions

It is difficult to draw definitive conclusions from our still limited experience. However, it is apparent that Taylor models are sometimes helpful and sometimes not helpful in verified global optimization. Given a hard problem, a user of verified global optimization software should probably first determine whether or not the intractability is due to overestimation in range bounds or due to some other reason. Heuristic tools for this purpose include stand-alone Taylor model evaluators (such as COSY-INFINITY) and rastering schemes.

Although costing a significant implementation effort, Taylor model capability would be useful if bundled with verified global optimization software. However, it should either be a well-documented user-controlled option or automatically chosen with a good heuristic, since there is evidence that it sometimes will provide great benefit and sometimes could make the algorithm significantly slower.

Taylor models also may be helpful in putting systems of equations into a form in which we can symbolically precondition.

Actual implementation of Taylor models in a global optimization scheme may be somewhat different from the current COSY-INFINITY system, since models for first and second-order derivatives, in addition to the function itself, would be useful.

Acknowledgments: We wish to thank Martin Berz, who organized and supported a mini-conference at the National Superconducting Cyclotron Center for us to discuss and become familiar with COSY-INFINITY and its potential uses. Along these lines, we also wish to thank Kyoko Makino, who has been very helpful during our initial experiments with COSY-INFINITY.

6 References

[Ber91] M. Berz. Algorithms for higher derivatives in many variables with applications to beam physics. In A. Griewank and G. F. Corliss, editors, *Automatic Differentiation of Algorithms*, pages 147–156, 1991.

[Ber96] S. Berner. A parallel method for verified global optimization. In G. Alefeld, A. Frommer, and B. Lang, editors, *Scientific Computing and Validated Numerics*, Mathematical Research, volume 90, pages 200–206, Berlin, 1996. Akademie Verlag.

[Ber00] M. Berz. COSY INFINITY web page, 2000. `http://cosy.nscl.msu.edu/`.

[BH94] M. Berz and G. Hoffstätter. Exact bounds on the long term stability of weakly nonlinear systems applied to the design of large storage rings. *Interval Computations*, 1994(2):68–89, 1994.

[BH98] M. Berz and G. Hoffstätter. Computation and application of Taylor polynomials with interval remainder bounds. *Reliable Computing*, 4(1):83–97, February 1998.

[BMS+96] M. Berz, K. Makino, K. Shameiddine, G. H. Hoffstätter, and W. Wan. Cosy infinity and its applications in nonlinear dynamics. In M. Berz, C. Bischof, G. Corliss, and A. Griewank, editors, *Computational Differentiation, Techniques, Applications, and Tools*, pages 363–365, Philadelphia, 1996. SIAM.

[Cor98] G. F. Corliss. Globsol entry page, 1998. `http://www.mscs.mu.edu/~globsol/`.

[CR97] T. Csendes and D. Ratz. Subdivision direction selection in interval methods for global optimization. *SIAM J. Numer. Anal.*, 34(3):922–938, June 1997.

[Han92] E. R. Hansen. *Global Optimization Using Interval Analysis.* Marcel Dekker, Inc., New York, 1992.

[Kea96] R. B. Kearfott. *Rigorous Global Search: Continuous Problems.* Kluwer, Dordrecht, Netherlands, 1996.

[Kea97] R. B. Kearfott. Empirical evaluation of innovations in interval branch and bound algorithms for nonlinear algebraic systems. *SIAM J. Sci. Comput.*, 18(2):574–594, March 1997.

[Kea99] R. B. Kearfott. On stopping criteria in verified nonlinear systems or optimization algorithms, 1999. Accepted for publication in *ACM Trans. Math. Software.*

[KLRK98] V. Kreinovich, A. Lakeyev, J. Rohn, and P. Kahl. *Computational Complexity and Feasibility of Data Processing and Interval Computations.* Kluwer, Dordrecht, Netherlands, 1998.

[MB99] K. Makino and M. Berz. Efficient control of the dependency problem based on Taylor model methods. *Reliable Computing*, 5(1):3–12, 1999.

[MB00] K. Makino and M. Berz. New applications of Taylor model methods. In *this proceedings*, New York, 2000. Springer-Verlag.

[RC95] D. Ratz and T. Csendes. On the selection of subdivision directions in interval branch-and-bound methods for global optimization. *J. Global Optim.*, 7:183–207, 1995.

[VHMD97] P. Van Hentenryck, L. Michel, and Y. Deville. *Numerica: A Modeling Language for Global Optimization.* MIT Press, Cambridge, MA, 1997.