Multivariate Taylor Models in Global Optimization

by

R. Baker Kearfott, `rbk@louisiana.edu`
Department of Mathematics, University of Louisiana at Lafayette

Abstract

In deterministic global optimization algorithms, a systematic search is made of the entire domain. The domain is adaptively subdivided into smaller sub-domains. The subdomains are rejected, subdivided further, or accepted as containing solutions, based on bounds on the range of the objective function, constraints, and partial derivatives thereof.

For a global optimization algorithm to be practical, it is sometimes crucial that the bounds on the range be as sharp as possible. For some problems, straightforward interval evaluations or even mean-value extensions lead to bounds that are too pessimistic to be of use. In contrast, higher-order multivariate Taylor models can lead to much tighter bounds.

Talk Outline

Time permitting, we will

- briefly review deterministic global optimization,
- give examples where overestimation in straightforward interval evaluations causes problems,
- explain the basic idea underlying Taylor models,
- give historical background, theoretical results and previous use of Taylor models,
- show the performance of Taylor models on an example, and
- point out technical considerations in use of Taylor models.

The Global Optimization Problem

A basic problem is

$$\boxed{\begin{array}{ll} \text{minimize} & \phi(x) \\ \text{subject to} & \left\{ \begin{array}{rcl} c(x) & = & 0 \quad \text{and} \\ g(x) & \le & 0, \end{array} \right\} \end{array}} \qquad (1)$$

where $\phi : \boldsymbol{x} \subset \mathbb{R}^n \longrightarrow \mathbb{R}$, $c : \boldsymbol{x} \longrightarrow \mathbb{R}^{m_1}$, and $g : \boldsymbol{x} \longrightarrow \mathbb{R}^{m_2}$, where $\boldsymbol{x}$ is an interval vector

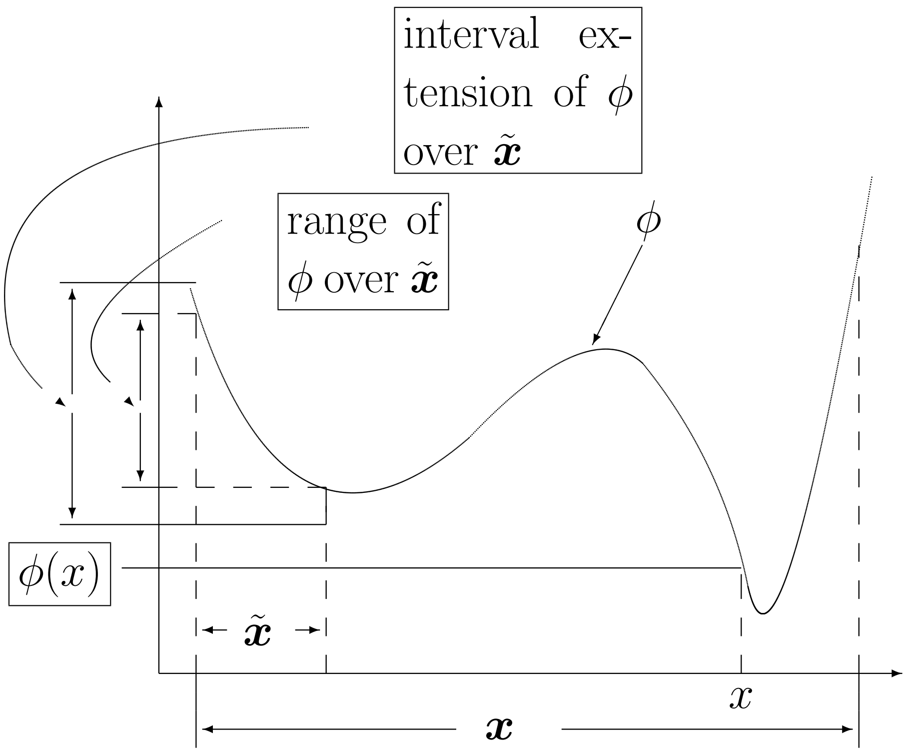
$$\boldsymbol{x} = ([\underline{x}_1, \overline{x}_1], \ldots, [\underline{x}_n, \overline{x}_n])^T.$$

A related problem is

$$\boxed{\begin{array}{ll} \text{Find all solutions to} & f(x) = 0 \\ \text{where } f : \boldsymbol{x} \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n & \end{array}} \qquad (2)$$

How Does It Work?

An Example Technique



The midpoint test: Rejecting $\tilde{\boldsymbol{x}}$ because of a high objective value

Example Computations

A Very Simple Illustration

- Suppose we want to find all solutions to $f(x) = x^2 - x - 1$ within $\boldsymbol{x} = [-2, 2]$.
- Suppose a subdivision process has produced the interval $[1.5, 2]$ to examine.
- Using interval arithmetic, $f([1.75, 2]) = [0.0625, 1.25]$, and $0 \notin [0.0625, 1.25]$. Thus, $[1.5, 2]$ can be discarded.
- Note that the actual range of f over $[1.75, 2]$ is $[0.3125, 1] \subset [0.0625, 1.25]$. However, the overestimation in the interval computation does not affect the result in this case.

An Example Where Overestimation Matters

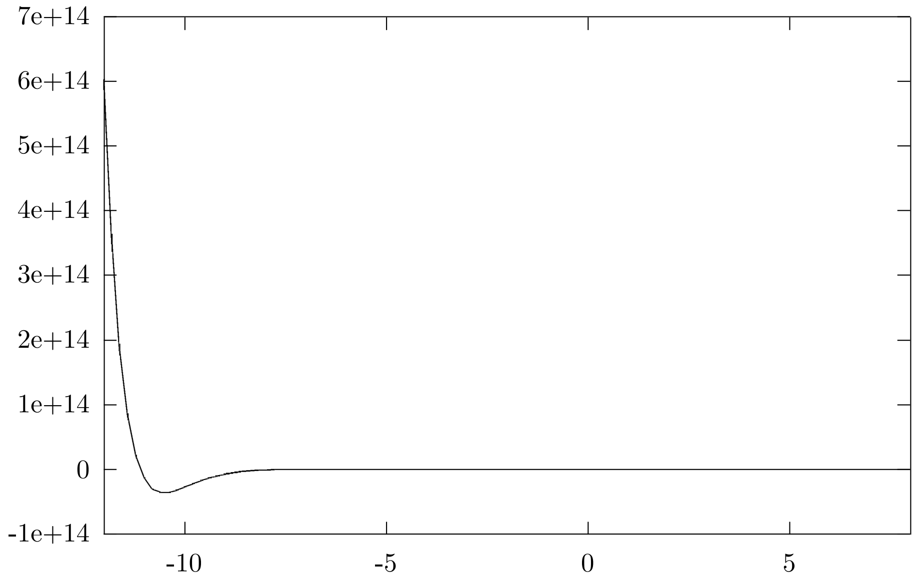
Gritton's Second Problem

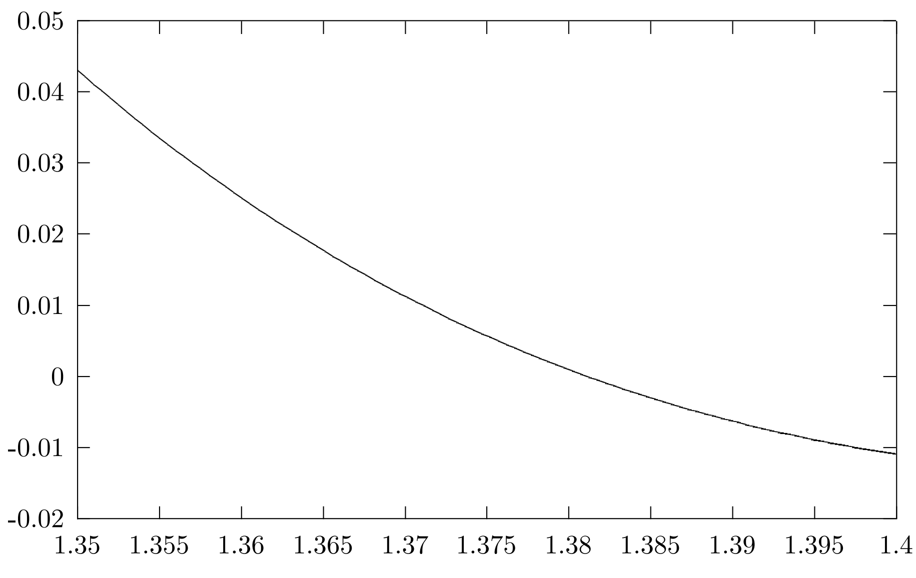
- An eighteenth degree polynomial arising from a chemical engineering problem.
- The problem has eighteen roots in the initial interval $[-12, 8]$.
- The root at $x \approx 1.381$ is difficult to isolate because of interval dependencies.

Gritton's Second Problem

$$
\begin{aligned}
p(x) \;=\; & -0.371936250000000 * 10^{+03} \\
& +x * (-0.791246565600000 * 10^{+03} \\
& +x * (0.404494414300000 * 10^{+04} \\
& +x * (0.978137516700000 * 10^{+03} \\
& +x * (-0.165478928000000 * 10^{+05} \\
& +x * (0.221407282700000 * 10^{+05} \\
& +x * (-0.932654935900000 * 10^{+04} \\
& +x * (-0.351853687200000 * 10^{+04} \\
& +x * (0.478253229600000 * 10^{+04} \\
& +x * (-0.128147944000000 * 10^{+04} \\
& +x * (-0.283443587500000 * 10^{+03} \\
& +x * (0.202627091500000 * 10^{+03} \\
& +x * (-0.161791345900000 * 10^{+02} \\
& +x * (-0.888303902000000 * 10^{+01} \\
& +x * (0.157558017300000 * 10^{+01} \\
& +x * (0.124599084800000 * 10^{+00} \\
& +x * (-0.358914862200000 * 10^{-01} \\
& +x * (-0.195109557600000 * 10^{-03} \\
& +x * (0.227468222900000 * 10^{-03}))))))))))))))))))
\end{aligned}
$$

Gritton's Second Problem





Gritton's Second Problem

- One possibly occurring computation is to reject the interval $\boldsymbol{x} = [1.35, 1.37]$, near the root near $x \approx 1.38$.
- A straightforward interval evaluation of $p(x)$ gives

  $$p(x) \in [-1381.74, 1383.98],$$

  not useful to determine $p(x) \neq 0$.
- A mean value extension gives

  $$p(x) \in [-6.29, 6.35],$$

  also not an adequate approximation.
- Because of monotonicity, the actual range (to 3 digits) is

  $$p(x) \in [0.0111, 0.0431],$$

  but an interval evaluation of $p'(\boldsymbol{x})$ gives $p'(x) \in [-6184.71, 6229.86]$, not indicative of monotonicity.

A Taylor Model for Gritton

- The Taylor polynomial of degree 5 about $x = 1.36$ can be computed as approximately $T_5(x) \approx 0.0251 - 1.582(-1.36 + x) + 20.2(-1.36 + x)^2 - 82.8(-1.36 + x)^3 + 42.8(-1.36 + x)^4 + 384.(-1.36 + x)^5$.

- The error term is of the form

$$E(x) = \frac{p^{(6)}([1.35, 1.37])}{6!}(x - 1.36)^6 + \left\{\begin{matrix}\text{Intervals enclosing roundoff} \\ \text{in the coefficients of } T_5\end{matrix}\right\}$$

- T_5 and E give the inclusion

$$\begin{aligned} p(x) &\in T_5([1.35, 1.37]) + E([1.35, 1.37]) \\ &\subseteq [0.00923, 0.0431] \end{aligned}$$

  This is sharp enough to conclude $p(x) \neq 0$, and is close to the actual range $[0.0111, 0.0431]$.

Prospects for Taylor Models

- Effectiveness of the model

$$f(x) \in T_k(\boldsymbol{x}) + E_k(\boldsymbol{x})$$

  in global optimization depends upon getting good bounds on the range of T_k.

- Negative aspects:
  - Naive interval computation for $T_k(\boldsymbol{x})$ can in principle lead to overestimation similarly to interval evaluation of f.
  - Computation of the range of a quadratic in n variables to within a specified accuracy is NP-complete in n.

Prospects for Taylor Models

Positive aspects

- In practice, estimation by a low-degree polynomial appears to be very effective, especially for functions defined by complicated expressions but with mild nonlinearities.
- Many hard global optimization problems have a small number of variables, within a range that has been handled well by Taylor method software.
- Polynomial models of degree 20 or more with orders of 10 or more have been successfully used in practice.

A Short History

Cornelius-Lohner (1984): Develop higher-order range estimation in one variable and suggest generalization

Berz (1987): Proposes Taylor arithmetic to integrate the differential equations involved in beam physics

Berz (1991): Publishes data structures and a software environment for efficient operations with multivariate Taylor models

Berz and Hoffstätter (1994): Publish incorporation of interval bounding process in beam theory computations

Berz and Makino (1996): Publish general method for Taylor arithmetic with remainder bounds.

Kreinovich et al (1998): Negative results for computational complexity in high dimensions

Makino and Berz (1999): Propose use in global optimization, and give an example showing promise

The Present State of the Art

- Software is available from the beam theory group at Michigan State. The Taylor arithmetic is part of the package COSY INFINITY (See `http://bt.nscl.msu.edu/cosy/`)
- COSY is oriented towards beam physics, but can be used for preliminary experiments in other areas.

Technical Aspects

- An automatic differentiation process is used to compute $T_n(\boldsymbol{x}) + E(\boldsymbol{x})$.
- Special data structures are used to make the multidimensional computations efficient.
- Various techniques can be used to bound $T_n(\boldsymbol{x})$ more precisely than with naive interval arithmetic.
  - Little is done at present.
  - In one variable, the range of T_n can be computed exactly if $n \leq 5$.
  - Heuristics have been proposed in the multivariate case.

Why is the Taylor model successful in beam computations?

1. The function modelled is only weakly nonlinear, so higher-order coefficients are small. The high-order Taylor model can thus give the required high accuracy.
2. Enclosures for the function value at *points*, in contrast to range enclosures over intervals, are sought. This leads to less overestimation.

How might the Taylor model work for global optimization?

1. The properties of interval arithmetic (*sub*-distributivity) cause different amounts of overestimation for different arrangements of the same algebraic expression.
2. Due to a basic theorem of interval arithmetic, the linear part of the Taylor polynomial has an exact range.
3. Basing the Taylor polynomial in the center of the interval of interest causes higher-order terms to be small corrections and minimizes overestimation.
4. The n-th degree model of an n-th degree expression seems to often give better results than the original expression.

Considerations for Global Optimization

1. There is a tradeoff between amount of computation to evaluate the function and amount of overestimation.
2. Is second or third order accurate enough in global optimization, where geometry, in addition to overestimation, drives subdivision?
3. The COSY software cannot be seamlessly incorporated into global optimization.
   (a) Much work is required to implement a general Taylor model.
   (b) Discovery of a simpler subset of the general model class would aid implementation.
4. Just how much is gained from considering Taylor models?

How to Proceed?

Proposed Discovery Path

1. Identify important problems not solvable without Taylor models.
2. Study ranges over intervals of interest, comparing naive interval evaluations to Taylor models of various orders.
3. Judge if it is useful to consider just low-order implementations.
4. Formulate the conditions (order of function, gradient, and Hessian approximations) necessary in for Global Optimization

Proposed Discovery Path

(continued)

5. Implement the simplified Taylor model in GlobSol.
6. Experiment with the previously identified problems.
7. Draw conclusions concerning the Taylor model's usefulness.
8. Summarize clearly the class of problems over which Taylor models work in Global optimization.