5. Remainder Differential Algebras and Their Applications*

(Chapter of "Computational Differentiation: Techniques, Applications, and Tools",
Martin Berz, Christian Bischof, George Corliss, and Andreas Griewank, eds., SIAM, 1996.)

Kyoko Makino† Martin Berz†

Abstract

In many practical problems in which derivatives are calculated, their basic purpose is to be used in the modeling of a functional dependence, often based on a Taylor expansion to first or higher orders. While the practical computation of such derivatives is greatly facilitated and in many cases is possible only through the use of forward or reverse computational differentiation, there is usually no direct information regarding the accuracy of the functional model based on the Taylor expansion.

We show how, in parallel to the accumulation of derivatives, error bounds of all functional dependencies can be carried along the computation. The additional effort is minor, and the resulting bounds are usually rather sharp, in particular at higher orders. This Remainder Differential Algebraic Method is more straightforward and can yield tighter bounds than the mere interval bounding of the Taylor remainder's $(n+1)$st order derivative obtained via forward differentiation.

The method can be applied to various numerical problems: Here we focus on global optimization, where blow-up can often be substantially reduced compared with interval methods, in particular for the cases of complicated functions or many variables. This problem is at the core of many questions of nonlinear dynamics and can help facilitate a detailed, quantitative understanding.

Keywords: Remainder differential algebras, differential algebras, error bound, interval method, high-order derivatives, Taylor polynomial, Taylor remainder, beam physics, COSY INFINITY, Fortran precompiler.

1 Introduction

The significant advances in computer hardware that we have experienced particularly in the past decade allow the study of ever more complex problems. In many practical problems, one must locally model nonlinear functional dependencies, for example to study parameter sensitivity or to perform optimization. This is typically done through the computation of derivatives, and the forward and reverse modes of computational differentiation [Griewank1991e] have excelled in providing such derivatives accurately and inexpensively.

While the derivatives themselves are accurate except for computational errors that are typically very small, rigor is lost when the derivatives are used to model a functional dependence, because of the lack of information about the size of remainder terms. The method of interval arithmetic (see, for example, [Kulisch1981a]) often provides a means for keeping the mathematical rigor in the computation of model functions. However, the naive use of interval methods in large problems is prone to blow-up, which at times limits the practical usefulness of such methods.

*This research was financially supported by the Department of Energy, Grant No. DE-FG02-95ER40931, and the Alfred P. Sloan Foundation.

†Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824, USA, (makino@nscl.msu.edu, berz@pa.msu.edu).

In this paper, we discuss a combination of the techniques of computational differentiation and interval methods, in a way that uses the advantages and diminishes the disadvantages of either of the two methods. The new technique, the method of Remainder Differential Algebras, employs high-order computational differentiation [Berz1991a] to express the model function by a Taylor polynomial, and interval computation to evaluate the Taylor remainder error bound.

For example, in beam physics and weakly nonlinear dynamics in general, the Differential Algebraic technique [Berz1989a], [Berz1990e], [Berz1991d], [Berz1994b] has offered a remarkably robust way to study the nonlinear behavior of beams and has evolved into one of the essential tools. However, many difficult yet important questions remain, including the long-term stability of beams in circular accelerators or other dynamical systems, such as the planets in the solar system. In light of modern stability theories, this problem can be cast as an optimization problem [Berz1994c], [Berz1996b], which from the computational point of view is very complex. The method of Remainder Differential Algebras as an extension of conventional Differential Algebraic techniques may shed light on this and many other questions requiring verified computations of complicated problems [Berz1996b].

2 Remainder Differential Algebras

In this section, we provide an overview of the method of computing remainder terms along with the original function. The Taylor theorem plays an important role in this endeavor, and we briefly state it here.

THEOREM 2.1 (TAYLOR'S THEOREM). *Suppose that a function* $f : [\vec{a}, \vec{b}] \subset R^v \to R$ *is* $(n+1)$ *times partially differentiable on* $[\vec{a}, \vec{b}]$. *Assume* $\vec{x}_0 \in [\vec{a}, \vec{b}]$. *Then for each* $\vec{x} \in [\vec{a}, \vec{b}]$, *there is* $\theta \in R$ *with* $0 < \theta < 1$ *such that*

$$f(\vec{x}) = \sum_{\nu=0}^{n} \frac{1}{\nu!} \left((\vec{x} - \vec{x}_0) \cdot \vec{\nabla} \right)^{\nu} f(\vec{x}_0) + \frac{1}{(n+1)!} \left((\vec{x} - \vec{x}_0) \cdot \vec{\nabla} \right)^{n+1} f\left(\vec{x}_0 + (\vec{x} - \vec{x}_0)\theta\right),$$

where the partial differential operator $\left(\vec{h} \cdot \vec{\nabla}\right)^k$ *operates as*

$$\left(\vec{h} \cdot \vec{\nabla}\right)^k = \sum_{\substack{0 \le i_1, \cdots, i_v \le k \\ i_1 + \cdots + i_v = k}} \frac{k!}{i_1! \cdots i_v!} h_1^{i_1} \cdots h_v^{i_v} \frac{\partial^k}{\partial x_1^{i_1} \cdots \partial x_v^{i_v}}.$$

Depending on the situation at hand, the remainder term also can be cast into a variety of other forms. Taylor's theorem allows a quantitative estimate of the error that is to be expected when approximating a function by its Taylor polynomial. Furthermore, it even offers a way to obtain bounds for the error in practice, based on bounding the $(n+1)$st derivative, a method that has been employed in interval calculations.

Roughly speaking, Taylor's theorem suggests that in many cases the error decreases with the order as the width of the interval raised to the order being considered, and its practical use is often connected to this observation. However, certain examples illustrate that this behavior does not have to occur; one such example is in the following section.

For notational convenience, we introduce a parameter α to describe the details of a given Taylor expansion, namely, the order of the Taylor polynomial n, the reference point of expansion $\vec{x}_0$, and the domain interval $[\vec{a}, \vec{b}]$ on which the function is to be considered as

$$\alpha = (n, \vec{x}_0, [\vec{a}, \vec{b}]). \tag{1}$$

With the help of Taylor's theorem, any $(n+1)$ times partially differentiable function $f : [\vec{a}, \vec{b}] \subset R^v \to R$ can be expressed by the Taylor polynomial $P_{\alpha,f}$ of nth order and a remainder $\varepsilon_{\alpha,f}$. We write it symbolically as

$$f(\vec{x}) = P_{\alpha,f}(\vec{x} - \vec{x}_0) + \varepsilon_{\alpha,f}(\vec{x} - \vec{x}_0),$$

where $\varepsilon_{\alpha,f}(\vec{x} - \vec{x}_0)$ is continuous on the domain interval and thus bounded. Let the interval $I_{\alpha,f}$ be such that for any $\vec{x} \in [\vec{a}, \vec{b}]$, $\varepsilon_{\alpha,f}(\vec{x} - \vec{x}_0) \in I_{\alpha,f}$. Then

$$\forall \vec{x} \in [\vec{a}, \vec{b}], \quad f(\vec{x}) \in P_{\alpha,f}(\vec{x} - \vec{x}_0) + I_{\alpha,f}. \tag{2}$$

Because of the special form of the Taylor remainder term $\varepsilon_{\alpha,f}$, in practice the remainder usually decreases as $|\vec{x} - \vec{x}_0|^{n+1}$. Hence, if $|\vec{x} - \vec{x}_0|$ is chosen to be small, the interval $I_{\alpha,f}$, which from now on we refer to as the interval remainder bound, can become very small. The set $P_{\alpha,f}(\vec{x} - \vec{x}_0) + I_{\alpha,f}$ containing f consists of the Taylor polynomial $P_{\alpha,f}(\vec{x} - \vec{x}_0)$ and the interval remainder bound $I_{\alpha,f}$. We say a pair $(P_{\alpha,f}, I_{\alpha,f})$ of a Taylor polynomial $P_{\alpha,f}(\vec{x} - \vec{x}_0)$ and an interval remainder bound $I_{\alpha,f}$ is a Taylor model of f if and only if (2) is satisfied. In this case, we denote the Taylor model by

$$T_{\alpha,f} = (P_{\alpha,f}, I_{\alpha,f}).$$

We call n the order of the Taylor model, $\vec{x}_0$ the reference point of the Taylor model, $[\vec{a}, \vec{b}]$ the domain interval of the Taylor model, and α the parameter of the Taylor model.

In the following, we develop tools that allow us to calculate Taylor models for all functions representable on a computer.

2.1 Differential Algebras and Interval Arithmetic

In the preceding section, the concept of Remainder Differential Algebras and Taylor models was introduced. While the computational idea of Taylor models is unique, the two constituents of a Taylor model, namely, the Taylor polynomial and the bounding interval, are familiar concepts.

While computational differentiation techniques focus mostly on first- or low-order derivatives, some algorithms [Berz1991a] and codes allow the computation of very high-order derivatives [Berz1990d], [Berz1990a], [Berz1994b], [Berz1995a]. This application has received attention recently in the field of beam physics in a differential algebraic framework suited for treatment of differential equations [Berz1992a], [Berz1993a], [Berz1995a], [Berz1996b].

While these methods can provide the derivatives of the function to high orders, they fail to provide rigorous information about the range of the function. A simple example that dramatically illustrates this phenomenon is the function

$$f(x) = \begin{cases} \exp(-1/x^2) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

The value of the function and all the derivatives at $x = 0$ are 0. Thus the Taylor polynomial at the reference point $x = 0$ is just the constant 0. In particular, this also implies that the Taylor series of f converges everywhere, but it fails to agree with $f(x)$ everywhere but at $x = 0$.

In a situation such as this one, the methods of interval arithmetic make a contrast. Interval arithmetic carries the information of rigorous bounds of a function in computation, and the computation time is usually reasonably fast. However, the potential problem of blow-up always exists. To illustrate this phenomenon with a trivial example, we consider the interval $I = [a, b]$, which has the width $b - a$. We compute the addition of I to itself and its subtraction from itself: $I + I = [2a, 2b]$ and $I - I = [a - b, b - a]$. In both cases the resulting width is $2(b - a)$ and is twice the original width, although we know that regardless of what unknown quantity x is characterized by I, certainly $x - x$ should equal zero. Similar blow-up often poses a severe problem for interval methods, in particular when the underlying functions become very complex. In the case of Remainder Differential Algebras, however, the remainder bound intervals are kept so small that even the effect of considerable blow-up is not detrimental.

2.2 Addition and Multiplication of Taylor Models

In this section, we discuss how a Taylor model of a sum or product of two functions can be obtained from the Taylor models of the two individual functions. This represents the first step toward the computation of Taylor models for any function that can be represented on a computer.

Let the functions $f, g : [\vec{a}, \vec{b}] \subset R^v \to R$ have Taylor models

$$T_{\alpha,f} = (P_{\alpha,f}, I_{\alpha,f}) \quad \text{and} \quad T_{\alpha,g} = (P_{\alpha,g}, I_{\alpha,g}),$$

which entails that

$$\begin{array}{rrcll} \forall \vec{x} \in [\vec{a}, \vec{b}], & f(\vec{x}) & \in & P_{\alpha,f}(\vec{x} - \vec{x}_0) + I_{\alpha,f} & \quad \text{and} \\ & g(\vec{x}) & \in & P_{\alpha,g}(\vec{x} - \vec{x}_0) + I_{\alpha,g}. & \end{array}$$

Then it is straightforward to obtain a Taylor model for $f + g$; in fact, for any $\vec{x} \in [\vec{a}, \vec{b}]$,

$$\begin{array}{rcl} f(\vec{x}) + g(\vec{x}) & \in & (P_{\alpha,f}(\vec{x} - \vec{x}_0) + I_{\alpha,f}) + (P_{\alpha,g}(\vec{x} - \vec{x}_0) + I_{\alpha,g}) \\ & = & (P_{\alpha,f}(\vec{x} - \vec{x}_0) + P_{\alpha,g}(\vec{x} - \vec{x}_0)) + (I_{\alpha,f} + I_{\alpha,g}), \end{array}$$

so that a Taylor model $T_{\alpha,f+g}$ for $f + g$ can be obtained via

(3) $$P_{\alpha,f+g} = P_{\alpha,f} + P_{\alpha,g} \quad \text{and} \quad I_{\alpha,f+g} = I_{\alpha,f} + I_{\alpha,g}.$$

Thus we define

$$T_{\alpha,f} + T_{\alpha,g} = (P_{\alpha,f} + P_{\alpha,g}, I_{\alpha,f} + I_{\alpha,g}),$$

and we obtain that $T_{\alpha,f} + T_{\alpha,g} = (P_{\alpha,f+g}, I_{\alpha,f+g})$ is a Taylor model for $f + g$. Note that the above addition of Taylor models is both commutative and associative.

The goal in defining a multiplication of Taylor models is to determine a Taylor model for $f \cdot g$ from the knowledge of the Taylor models $T_{\alpha,f}$ and $T_{\alpha,g}$ for f and g. Observe that for any $\vec{x} \in [\vec{a}, \vec{b}]$,

$$\begin{array}{rcl} f(\vec{x}) \cdot g(\vec{x}) & \in & (P_{\alpha,f}(\vec{x} - \vec{x}_0) + I_{\alpha,f}) \cdot (P_{\alpha,g}(\vec{x} - \vec{x}_0) + I_{\alpha,g}) \\ & \subseteq & P_{\alpha,f}(\vec{x} - \vec{x}_0) \cdot P_{\alpha,g}(\vec{x} - \vec{x}_0) \\ & & + P_{\alpha,f}(\vec{x} - \vec{x}_0) \cdot I_{\alpha,g} + P_{\alpha,g}(\vec{x} - \vec{x}_0) \cdot I_{\alpha,f} + I_{\alpha,f} \cdot I_{\alpha,g}. \end{array}$$

Note that $P_{\alpha,f} \cdot P_{\alpha,g}$ is a polynomial of $(2n)$th order. We split it into the part of up to nth order, which agrees with the Taylor polynomial $P_{\alpha,f\cdot g}$ of order n of $f \cdot g$, and the extra polynomial P_e, so that we have

(4) $$P_{\alpha,f}(\vec{x} - \vec{x}_0) \cdot P_{\alpha,g}(\vec{x} - \vec{x}_0) = P_{\alpha,f\cdot g}(\vec{x} - \vec{x}_0) + P_e(\vec{x} - \vec{x}_0).$$

A Taylor model for $f \cdot g$ can now be obtained by finding an interval bound for all the terms except $P_{\alpha,f\cdot g}$. For this purpose, let $B(P)$ be a bound of the polynomial $P : [\vec{a},\vec{b}] \subset R^v \to R$, namely,

$$\forall \vec{x} \in [\vec{a},\vec{b}], \quad P(\vec{x}) \in B(P).$$

Apparently the efficient practical determination of $B(P)$ is not completely trivial; depending on the order and number of variables, different strategies may be employed, ranging from analytical estimates to interval evaluations. However, thanks to the specific circumstances, the occurring contributions are very small, and even moderate overestimation is not immediately critical.

Altogether, an interval remainder bound for $f \cdot g$ can be found via

$$I_{\alpha,f\cdot g} = B(P_e) + B(P_{\alpha,f}) \cdot I_{\alpha,g} + B(P_{\alpha,g}) \cdot I_{\alpha,f} + I_{\alpha,f} \cdot I_{\alpha,g}. \tag{5}$$

Thus we define $T_{\alpha,f} \cdot T_{\alpha,g} = (P_{\alpha,f\cdot g}, I_{\alpha,f\cdot g})$, and obtain that $T_{\alpha,f} \cdot T_{\alpha,g}$ is a Taylor model for $f \cdot g$. Note that commutativity of multiplication holds, $T_{\alpha,f} \cdot T_{\alpha,g} = T_{\alpha,g} \cdot T_{\alpha,f}$, while multiplication is not generally associative, and also distributivity does not generally hold.

While the idea of Taylor models of constant functions is almost trivial, we mention it for the sake of completeness. For a constant function $f(\vec{x}) \equiv t$, the Taylor model of f is

$$T_{\alpha,f} \equiv T_{\alpha,t} = (P_{\alpha,t}, I_{\alpha,t}) = (t, [0,0]).$$

Having introduced addition and multiplication as well as scalar multiplication, we can compute any polynomial of a Taylor model. Let $Q(f)$ be a polynomial of a function f, that is, $Q(f) = t_0 + t_1 f + t_2 f^2 + \cdots + t_k f^k$. In practice it is useful to evaluate $Q(f)$ via Horner's scheme,

$$Q(f) = t_0 + f \cdot \Big(t_1 + f \cdot \Big(t_2 + f \cdot (\cdots (t_{k-1} + f \cdot t_k) \cdots)\Big)\Big),$$

in order to minimize operations. Assume that we have already found the Taylor model of the function f to be $T_{\alpha,f} = (P_{\alpha,f}, I_{\alpha,f})$. Then, using additions and multiplications of Taylor models described above, we can compute a Taylor model for the function $Q(f)$ via

$$T_{\alpha,Q(f)} = \Big(P_{\alpha,Q(f)}, I_{\alpha,Q(f)}\Big).$$

2.3 Functions in Remainder Differential Algebras

In the preceding section, we showed how Taylor models for sums and products of functions can be obtained from those of the individual functions. The computation led to the definition of addition and multiplication of Taylor models. Here we study the computation of Taylor models for intrinsic functions, including the reciprocal applied to a given function f from the Taylor model of f.

The key idea is to employ Taylor's theorem of the function under consideration: However, in order to ensure that the resulting remainder term yields a small remainder interval and does not contribute anything to the Taylor polynomial, some additional manipulations are necessary.

Let us begin the study with the exponential function. Assume that we have already found the Taylor model of the function f to be $T_{\alpha,f} = (P_{\alpha,f}, I_{\alpha,f})$. Write the constant part of the function f around $\vec{x}_0$ as $c_{\alpha,f}$, which agrees with the constant part of the Taylor polynomial $P_{\alpha,f}$, and write the remaining part as $\bar{f}$, that is,

$$f(\vec{x}) = c_{\alpha,f} + \bar{f}(\vec{x}).$$

A Taylor model of $\bar{f}$ is then $T_{\alpha,\bar{f}} = (P_{\alpha,\bar{f}}, I_{\alpha,\bar{f}})$, where

$$P_{\alpha,\bar{f}}(\vec{x}-\vec{x}_0) = P_{\alpha,f}(\vec{x}-\vec{x}_0) - c_{\alpha,f} \qquad \text{and} \qquad I_{\alpha,\bar{f}} = I_{\alpha,f}.$$

Now we can write

$$\begin{aligned}
\exp(f(\vec{x})) &= \exp\left(c_{\alpha,f} + \bar{f}(\vec{x})\right) = \exp(c_{\alpha,f}) \cdot \exp\left(\bar{f}(\vec{x})\right) \\
&= \exp(c_{\alpha,f}) \cdot \left\{1 + \bar{f}(\vec{x}) + \frac{1}{2!}(\bar{f}(\vec{x}))^2 + \cdots + \frac{1}{k!}(\bar{f}(\vec{x}))^k \right. \\
&\qquad \left. + \frac{1}{(k+1)!}(\bar{f}(\vec{x}))^{k+1} \exp\left(\theta \cdot \bar{f}(\vec{x})\right)\right\},
\end{aligned}$$

where $0 < \theta < 1$. Taking $k \geq n$, where n is the order of Taylor model, the part

$$\exp(c_{\alpha,f}) \cdot \left\{1 + \bar{f}(\vec{x}) + \frac{1}{2!}(\bar{f}(\vec{x}))^2 + \cdots + \frac{1}{n!}(\bar{f}(\vec{x}))^n\right\}$$

is a polynomial of $\bar{f}$, of which we can obtain the Taylor model as outlined in the preceding section. The remainder part of $\exp(f(\vec{x}))$,

$$\exp(c_{\alpha,f}) \cdot \left\{\frac{1}{(n+1)!}(\bar{f}(\vec{x}))^{n+1} + \cdots + \frac{1}{(k+1)!}(\bar{f}(\vec{x}))^{k+1} \exp\left(\theta \cdot \bar{f}(\vec{x})\right)\right\}, \tag{6}$$

will be bounded by an interval. Since $P_{\alpha,\bar{f}}(\vec{x}-\vec{x}_0)$ does not have a constant part, $(P_{\alpha,\bar{f}}(\vec{x}-\vec{x}_0))^m$ starts from mth order. Thus, in the Taylor model computation, the remainder part (6) has vanishing polynomial part. The remainder bound interval for the Lagrange remainder term

$$\exp(c_{\alpha,f}) \frac{1}{(k+1)!}(\bar{f}(\vec{x}))^{k+1} \exp\left(\theta \cdot \bar{f}(\vec{x})\right)$$

can be estimated because, for any $\vec{x} \in [\vec{a}, \vec{b}]$, $P_{\alpha,\bar{f}}(\vec{x}-\vec{x}_0) \in B(P_{\alpha,\bar{f}})$, and $0 < \theta < 1$, and so

$$(\bar{f}(\vec{x}))^{k+1} \exp\left(\theta \cdot \bar{f}(\vec{x})\right) \in \left(B(P_{\alpha,\bar{f}}) + I_{\alpha,\bar{f}}\right)^{k+1} \exp\left([0,1] \cdot (B(P_{\alpha,\bar{f}}) + I_{\alpha,\bar{f}})\right).$$

Since the exponential function is monotonically increasing, the estimation of the interval bound of the part $\exp\left([0,1] \cdot (B(P_{\alpha,\bar{f}}) + I_{\alpha,\bar{f}})\right)$ is achieved by inserting the upper and lower bounds of the argument in the exponential.

A Taylor model for the logarithm of a function f can be computed in a similar manner from the Taylor model of the function. In this case, there is the limitation that it has to be ensured that the range of the function f lies entirely within the range of definition of the logarithm, which will be the case if, for any $\vec{x} \in [\vec{a}, \vec{b}]$, any element in the set $P_{\alpha,f}(\vec{x}-\vec{x}_0) + I_{\alpha,f}$ is positive. For the actual computation, we again split the constant part of the function f around $\vec{x}_0$ from the rest $f(\vec{x}) = c_{\alpha,f} + \bar{f}(\vec{x})$. Then we obtain

$$\begin{aligned}
\log(f(\vec{x})) &= \log\left(c_{\alpha,f} + \bar{f}(\vec{x})\right) = \log\left\{c_{\alpha,f} \cdot \left(1 + \frac{\bar{f}(\vec{x})}{c_{\alpha,f}}\right)\right\} \\
&= \log c_{\alpha,f} + \log\left(1 + \frac{\bar{f}(\vec{x})}{c_{\alpha,f}}\right) \\
&= \log c_{\alpha,f} + \frac{\bar{f}(\vec{x})}{c_{\alpha,f}} - \frac{1}{2}\frac{(\bar{f}(\vec{x}))^2}{c_{\alpha,f}^2} + \cdots + (-1)^{k+1}\frac{1}{k}\frac{(\bar{f}(\vec{x}))^k}{c_{\alpha,f}^k} \\
&\qquad + (-1)^{k+2}\frac{1}{k+1}\frac{(\bar{f}(\vec{x}))^{k+1}}{c_{\alpha,f}^{k+1}} \frac{1}{\left(1 + \theta \cdot \bar{f}(\vec{x})/c_{\alpha,f}\right)^{k+1}},
\end{aligned}$$

where $0 < \theta < 1$. Taking $k \geq n$, the part

$$\log c_{\alpha,f} + \frac{\bar{f}(\vec{x})}{c_{\alpha,f}} - \frac{1}{2}\frac{(\bar{f}(\vec{x}))^2}{c_{\alpha,f}^2} + \cdots + (-1)^{n+1}\frac{1}{n}\frac{(\bar{f}(\vec{x}))^n}{c_{\alpha,f}^n}$$

is again treated as a polynomial of $\bar{f}$ in the Taylor model computation. The Lagrange remainder part of $\log(f(\vec{x}))$ becomes part of the remainder bound interval of the Taylor model of $\log(f)$. The remainder term can be estimated as

$$(-1)^{k+2}\frac{1}{k+1}\frac{(B(P_{\alpha,\bar{f}}) + I_{\alpha,\bar{f}})^{k+1}}{c_{\alpha,f}^{k+1}}\frac{1}{\left(1 + [0,1]\cdot(B(P_{\alpha,\bar{f}}) + I_{\alpha,\bar{f}})/c_{\alpha,f}\right)^{k+1}}.$$

In a rather similar fashion, it is possible to determine Taylor models of square roots and trigonometric functions as soon as a Taylor model for the argument is known. As a last example, we determine a Taylor model for the multiplicative inverse from that of the function. This Taylor model can be computed if and only if, for any $\vec{x} \in [\vec{a}, \vec{b}]$, any element in the set $P_{\alpha,f}(\vec{x} - \vec{x}_0) + I_{\alpha,f}$ is nonzero. For the actual computation, we again split the constant part of the function f around $\vec{x}_0$ from the rest as before. Then we obtain

$$\begin{aligned}
(7) \qquad \frac{1}{f(\vec{x})} &= \frac{1}{c_{\alpha,f} + \bar{f}(\vec{x})} = \frac{1}{c_{\alpha,f}}\cdot\frac{1}{1 + \bar{f}(\vec{x})/c_{\alpha,f}} \\
&= \frac{1}{c_{\alpha,f}}\cdot\left\{1 - \frac{\bar{f}(\vec{x})}{c_{\alpha,f}} + \frac{(\bar{f}(\vec{x}))^2}{c_{\alpha,f}^2} - \cdots + (-1)^k\frac{(\bar{f}(\vec{x}))^k}{c_{\alpha,f}^k}\right\} \\
&\quad +(-1)^{k+1}\frac{(\bar{f}(\vec{x}))^{k+1}}{c_{\alpha,f}^{k+2}}\frac{1}{(1 + \theta\cdot\bar{f}(\vec{x})/c_{\alpha,f})^{k+2}},
\end{aligned}$$

where again $0 < \theta < 1$. By choosing $k \geq n$, the Taylor model computation for the multiplicative inverse function can be done as before.

Altogether, it is now possible to compute Taylor models for any function that can be represented in a computer environment by simple operator overloading, in much the same way as the mere computation of derivatives, Taylor polynomials, or interval bounds, along with the mere evaluation of the function.

For many practical problems, in particular the efficient solution of differential equations, it is actually important to complement the set of currently available operations by a derivation ∂ that allows the computation of a Taylor model of the derivative of a function from that of the original function. As in the case of the conventional Differential Algebraic method, in order to prevent loss of order in the differentiation process, the derivation ∂ can be evaluated only in the context of a Lie derivative $L_g = g\cdot\partial$, where $g(\vec{x}_0) = 0$. However, in the case of Taylor models, an additional complication is connected to the fact that from the Taylor model alone, it is impossible to determine a bound for the derivative, since nothing is known about the rate of change of the function $(f - P_{\alpha,f})$ within the remainder bound I_f. The situation can be remedied by a further extension of the Taylor model concept to contain not only bounds for the remainder, but also a low-parameter bounding sequence for all the higher derivatives that can occur. For reasons of space, we have to restrict ourselves to this outlook as to what is necessary to complete the algebra of Taylor models into a Remainder Differential Algebra.

3 Examples

Remainder Differential Algebras have many applications, including global optimization, quadrature, and solution of differential equations. We begin our discussion with the determination of a sharp bound for a simple example function using Remainder Differential Algebras. The sharpness of the resulting bound will be compared with the results that can be obtained in other ways. The function under consideration is

$$f(x) = \frac{1}{x} + x. \tag{8}$$

For an actual computation, we set the parameter α of (1) to $\alpha = (n, x_0, [a, b]) = (3, 2, [1.9, 2.1])$.

As in the case of conventional forward differentiation, the evaluation begins with the representation of the identity function, expressed in terms of a Taylor polynomial expanded at the reference point. This identity function i has the form

$$i(x) = x = x_0 + (x - x_0) = 2 + (x - 2).$$

Since this representation is exact, the remainder bound is $[0, 0]$. Hence, a Taylor model of the identity function i is

$$T_{\alpha,i} = (x_0 + (x - x_0), [0, 0]) = (2 + (x - 2), [0, 0]).$$

The constant part of i around $x_0 = 2$ is $c_{\alpha,i} = x_0 = 2$, and the nonconstant part of i is $\bar{i}(x) = x - x_0 = x - 2$. The Taylor model of $\bar{i}$ is

$$T_{\alpha,\bar{i}} = ((x - x_0), [0, 0]) = ((x - 2), [0, 0]) .$$

The computation of the inverse requires the knowledge of a bound of $P_{\alpha,\bar{i}}$, which here is readily obtained: $B(P_{\alpha,\bar{i}}) = B(x - x_0) = [a - x_0, b - x_0] = [-0.1, 0.1]$. We have furthermore $B(P_{\alpha,\bar{i}}) + I_{\alpha,\bar{i}} = [-0.1, 0.1] + [0, 0] = [-0.1, 0.1]$. Using (4) and (5), we have for the Taylor model of $(\bar{i})^2$

$$T_{\alpha,(\bar{i})^2} = \left((x - 2)^2, [0, 0]\right) .$$

The Taylor model of $(\bar{i})^3$ is computed similarly: $T_{\alpha,(\bar{i})^3} = \left((x - 2)^3, [0, 0]\right)$. As can be seen, so far all remainder intervals are of zero size. The first nonzero remainder interval comes from the evaluation of the Taylor remainder term, which is

$$\begin{aligned} \frac{(\bar{i}(x))^4}{c_{\alpha,i}^5} \frac{1}{\left(1 + \theta \cdot \bar{i}(x)/c_{\alpha,i}\right)^5} &\in \frac{(B(P_{\alpha,\bar{i}}) + I_{\alpha,\bar{i}})^4}{x_0^5 \cdot \left(1 + [0, 1] \cdot (B(P_{\alpha,\bar{i}}) + I_{\alpha,\bar{i}})/x_0\right)^5} \\ &\subseteq \frac{[0, 0.0001]}{2^5 \cdot ([0.95, 1.05])^5} \subseteq [0, 4.038 \times 10^{-6}]. \end{aligned} \tag{9}$$

As expected, this remainder term is "small of order four". According to (7), the Taylor model of $1/i$ is then

$$T_{\alpha,\frac{1}{i}} = \left(\frac{1}{2} - \frac{1}{2^2}(x - 2) + \frac{1}{2^3}(x - 2)^2 - \frac{1}{2^4}(x - 2)^3, [0, 4.038 \times 10^{-6}]\right),$$

TABLE 1

The remainder bound interval $I_{\alpha,1/i+i}$ for various orders; $x_0 = 2$, $[a,b] = [1.9, 2.1]$

Order	The remainder bound interval
1	$[\;0\;, 1.4579384 \times 10^{-3}\;]$
2	$[\;-7.6733603 \times 10^{-5}\;, 7.6733603 \times 10^{-5}\;]$
3	$[\;0\;, 4.0386107 \times 10^{-6}\;]$
4	$[\;-2.1255845 \times 10^{-7}\;, 2.1255845 \times 10^{-7}\;]$
5	$[\;0\;, 1.1187287 \times 10^{-8}\;]$
6	$[\;-5.8880459 \times 10^{-10}\;, 5.8880459 \times 10^{-10}\;]$
7	$[\;0\;, 3.0989715 \times 10^{-11}\;]$
8	$[\;-1.6310376 \times 10^{-12}\;, 1.6310376 \times 10^{-12}\;]$
9	$[\;0\;, 8.5844087 \times 10^{-14}\;]$
10	$[\;-4.5181098 \times 10^{-15}\;, 4.5181098 \times 10^{-15}\;]$
11	$[\;0\;, 2.3779525 \times 10^{-16}\;]$
12	$[\;-1.2515539 \times 10^{-17}\;, 1.2515539 \times 10^{-17}\;]$
13	$[\;0\;, 6.5871262 \times 10^{-19}\;]$
14	$[\;-3.4669085 \times 10^{-20}\;, 3.4669085 \times 10^{-20}\;]$
15	$[\;0\;, 1.8246887 \times 10^{-21}\;]$

and the remainder interval is indeed still very sharp. Using (3), we obtain as the final Taylor model of $1/i + i$

$$
\begin{aligned}
(10) \qquad & T_{\alpha,\frac{1}{i}+i} = T_{\alpha,\frac{1}{i}} + T_{\alpha,i} = \left(P_{\alpha,\frac{1}{i}+i}, I_{\alpha,\frac{1}{i}+i}\right) \\
= \; & \left(\left(2+\frac{1}{2}\right) + \left(1-\frac{1}{2^2}\right)(x-2) + \frac{1}{2^3}(x-2)^2 - \frac{1}{2^4}(x-2)^3, [0, 4.038 \times 10^{-6}]\right) \\
= \; & \left(2.5 + 0.75(x-2) + 0.125(x-2)^2 - 0.0625(x-2)^3, [0, 4.038 \times 10^{-6}]\right).
\end{aligned}
$$

Since the polynomial $P_{\alpha,\frac{1}{i}+i}$ is monotonically increasing in the domain $[a,b] = [1.9, 2.1]$, the bound interval of the polynomial is

$$
B\left(P_{\alpha,\frac{1}{i}+i}\right) = \left[P_{\alpha,\frac{1}{i}+i}(-0.1), P_{\alpha,\frac{1}{i}+i}(0.1)\right] = [2.42631, 2.57618].
$$

The width of the bound interval of the Taylor polynomial is 0.14987, and the width of the interval of the remainder bound is 4.038×10^{-6} in the third-order Taylor model evaluation; thus the remainder part is just a minor addition. The size of this remainder bound depends strongly on the order and decreases quickly with order. Table 1 shows the remainder bound interval for various orders in the Taylor model computation.

The Taylor model computation is assessed by noting the bound interval B of the original function (8), which is

$$
B\left(\frac{1}{x} + x\right) = \left[\frac{1}{a} + a, \frac{1}{b} + b\right] = [2.42631, 2.57619].
$$

It is illuminating to compare the sharpness of the bounding of the function with the sharpness that can be obtained from conventional interval methods. Evaluating the function with just one interval yields

$$
\frac{1}{[a,b]} + [a,b] = \frac{1}{[1.9, 2.1]} + [1.9, 2.1] \subseteq [2.37619, 2.62631].
$$

TABLE 2

The width of the bound interval of $f(x) = 1/x + x$ by various methods; $x_0 = 2$, $[a, b] = [1.9, 2.1]$

Method		Width of Bound Interval
Intervals	$n_d = 1$	0.25012531
	$n_d = 10$	0.15993589
	$n_d = 10^2$	0.15088206
	$n_d = 10^3$	0.14997543
	$n_d = 10^4$	0.14988476
	$n_d = 10^5$	0.14987569
	$n_d = 10^6$	0.14987478
	$n_d = 10^7$	0.14987469
	$n_d = 10^8$	0.14987468
Taylor models	1st order	0.15145793
	2nd order	0.15015346
	3rd order	0.14987903
	4th order	0.14987542
	5th order	0.14987469
	6th order	0.14987468
Exact		0.14987468

The width of the bound interval obtained by interval arithmetic is 0.25012, and so this simple example already shows a noticeable blow-up. By dividing the domain interval into many subintervals, the blow-up can be suppressed substantially. However, to achieve the sharpness of the third-order Taylor model, the domain has to be split into about $24,000$ subintervals. Table 2 shows a comparison of the widths of bound interval for the exact value, the method of Taylor models, and the divided interval method, where n_d indicates the number of division of the domain interval. Of course, sophisticated interval optimization methods [Hansen1979a], [Hansen1988a], [Ichida1979a], [Jansson1992a] can find sharp bounds for the function using substantially fewer interval evaluations.

Practically more important are optimization problems in several variables, and in this case, the situation becomes more dramatic. We wish first to illustrate the computational effort necessary for an accurate calculation of the result by estimating the required number of floating-point operations. We use a simple example function of six variables such as $f(\vec{x}) = \sum_j (1/x_j + x_j)$ to get a rough idea of the computational expense in the case of functions of many variables. In the one-dimensional case, one interval calculation $1/[a, b] + [a, b]$ requires two additions and two divisions. To compare with the third-order Taylor model computation, we divide the domain into 10^4 subintervals, on which additions and divisions total $\sim 10^5$ floating-point operations. Thus, in the multidimensional case with six independent variables, the number of floating-point operations explodes to $(10^4)^6 \times (\sim 10) = \sim 10^{25}$. Again, sophisticated interval optimization methods will be more favorable than these numbers suggest, but typically there is still a very noticeable growth of complexity. In the next section, we will encounter a realistic example from the area of nonlinear dynamics where all state-of-the-art interval optimization methods available to us fail to give a satisfactory answer.

To estimate the performance of the Taylor model approach, we note that the one-dimensional Taylor model in the third-order computation involves a total of about 35

Table 3
The total number of FP operations required to bound a simple function like $f(\vec{x}) = \sum_j (1/x_j + x_j)$

	One Dimensional	Six Dimensional
Interval	~ 10	~ 10
10^4 divided intervals	$\sim 10^5$	$\sim 10^{25}$
3rd order Taylor model	~ 10	$\sim 10^4$

additions, multiplications, and divisions, as counted in (9) and (10). As we use more variables, however, the total number of terms in the polynomial grows only modestly. For example, order three in six variables requires only a total of 84 terms. Thus in total, the number of floating-point operations of the third-order Taylor model is $\sim 10^4$. A summary of the number of floating-point operations is given in Table 3.

4 Applications to Beam Physics

Any language environment with object-oriented features or one that supports operator overloading can be utilized to implement the Remainder Differential Algebraic Method in a conceptually similar way as conventional forward differentiation. An implementation following the reverse approach to differentiation, on the other hand, appears to be substantially more involved. We pursued an implementation in the program COSY INFINITY [Berz1992a], [Berz1993a], [Berz1995a], [Berz1996b], which can be used both as a precompiler [Berz1990a] for Fortran code and within its own language environment, and which is a major vehicle for studies in beam physics.

Here and in many other practical applications, the problem of finding rigorous bounds on the extrema of functions has contributed to the development of many methods in numerical analysis. We have been working on the problem to estimate the long-term stability of weakly nonlinear systems in connection with the design of large storage rings in beam facilities [Berz1994c], [Hoffstätter1994a]. The problem involves rather complicated functions of about 10^5 floating-point operations, which are multidimensional polynomials with six variables and up to roughly 500th order. Many large terms cancel each other, and very small fluctuations have to be estimated. While the sharpness of the bounds in question is important in order to guarantee a large number of stable turns, in reality these functions have a very large number of local maxima. Hence, an exact estimate of their bounds requires careful treatment. To be useful, the maxima have to be sharp to about 10^{-6}, and for some applications to 10^{-12}.

We have tried the method of conventional interval optimization [Hansen1979a], [Hansen1988a], [Ichida1979a], [Jansson1992a] and optimization based on Remainder Differential Algebras. To get an idea about the quality of the upper bounds, we compared the results with the approximations by a rather tight rastering in real arithmetic. Because of the large number of local maxima, the method of rastering proved to be the most robust noninterval approach to estimate the absolute maxima of the functions in question.

In the case of the conventional interval bound optimization, the suppression of blow-up is the key issue. The whole domain is covered with many smaller subintervals. Then the subintervals that do not give a better bound are excluded. Furthermore, the restructuring of the evaluation of the function helps the situation, which involves a decrease in the number of elementary operations and the introduction of new elementary operations such as x^2. Even with these simplifications, however, the resulting objective functions tend to exhibit

TABLE 4

Comparison of the bound estimate in various methods (data from PhD thesis of Hoffstätter)

Example	Conventional Interval Bounding (guaranteed lower bound)	Remainder Differential Algebras (guaranteed lower bound)	Conventional Rastering (optimistic, nonguaranteed upper bound)
1. Physical pendulum	11,306	432,158,877,713	636,501,641,854
2. Henon map	1,671	192,650,961	263,904,035
3. Los Alamos PSR II	171	1,004,387	2,248,621

interval blow-up because of the complexity. With the existence of many local maxima, the exclusion of unnecessary subintervals becomes difficult and indeed, virtually impossible unless in the order of 10^4 subintervals per dimension are used. However, this large number makes it unrealistic to apply any refined method of local optimization in each remaining subinterval in multidimensional cases like ours. For this reason, we restricted ourselves to a simple scan of the objective function with a large number of subintervals of equal size [Hoffstätter1994a]. The results were obtained by using 630 subintervals for examples 1 and 2, which are two dimensional, and 1,000,188 subintervals for example 3, which is four dimensional [Hoffstätter1994a]. Several nonlinear systems were studied by using the methods mentioned above. The results of bounds on the number of stable turns in three examples are listed in Table 4 [Hoffstätter1994a]. In the case of the interval bounding and the method of Remainder Differential Algebras, the numbers show lower bounds, while in the case of the rastering, the numbers show upper bounds.

The first example is a one-dimensional physical pendulum with two independent variables. This case offers a good test because the nonlinear motion is permanently stable as a result of energy conservation. The second example is a Henon map with two independent variables. This is a standard test case for the analysis of nonlinear motion because it shows many of the phenomena encountered in nonlinear dynamics. These include stable and unstable regions, chaotic motion, and periodic elliptic fixed points and can even serve as a simplistic model of an accelerator in the presence of sextupoles for chromaticity correction. The third example is a realistic accelerator, the Los Alamos PSR II storage ring for the motion in a phase space of 100 mm mrad with four independent variables. To limit the computation time, the subintervals used for the optimization in the last example were five times as wide as the subintervals used for the other two examples. The suppression in numbers of turns in the case of the conventional interval method is connected to the unavoidable blow-up of intervals in the process of cancellation of large terms. The method of Remainder Differential Algebras gives a satisfactory result close to the optimistic estimate by the rastering in real arithmetic, thereby making this method far superior to interval bounding.

References

[Berz1989a] M. BERZ, *Differential algebraic description of beam dynamics to very high orders*, Particle Accelerators, 24 (1989), p. 109.

[Berz1990a] ——, *Differential algebra precompiler version 3 — Reference manual*, Tech. Report MSUCL – 755, National Superconducting Cyclotron Laboratory, Michigan State University, East Lansing, Mich., 1990.

[Berz1990d] ——, *COSY INFINITY, an arbitrary order general purpose optics code*, in Computer Codes and the Linear Accelerator Community, Los Alamos LA–11857–C, 1990, pp. 137 +.

[Berz1990e] ——, *Arbitrary order description of arbitrary particle optical systems*, Nuclear Instruments and Methods, A298 (1990), pp. 426 +.

[Berz1991a] ——, *Forward algorithms for high orders and many variables with application to beam physics*, in Automatic Differentiation of Algorithms: Theory, Implementation, and Application, A. Griewank and G. F. Corliss, eds., SIAM, Philadelphia, Penn., 1991, pp. 147–156.

[Berz1991d] ——, *High-order computation and normal form analysis of repetitive systems*, in Physics of Particle Accelerators, M. Month, ed., vol. AIP 249, American Institute of Physics, 1991, p. 456.

[Berz1992a] ——, *COSY INFINITY Version 6*, in Proc. Nonlinear Effects in Accelerators, M. Berz, S. Martin, and K. Ziegler, eds., IOP Publishing, 1992, p. 125.

[Berz1993a] ——, *New features in COSY INFINITY*, in AIP Conference Proceedings 297, Computational Accelerator Physics, R. Ryne, ed., New York, 1993, American Institute of Physics, pp. 267–278.

[Berz1994b] ——, *Modern map methods for charged particle optics*, Nuclear Instruments and Methods, 352 (1994).

[Berz1994c] M. Berz and G. H. Hoffstätter, *Exact estimates of the long term stability of weakly nonlinear systems applied to the design of large storage rings*, Interval Computations, 2 (1994), pp. 68–89.

[Berz1995a] ——, *COSY INFINITY Version 7 reference manual*, Tech. Report MSUCL–977, National Superconducting Cyclotron Laboratory, Michigan State University, East Lansing, Mich., 1995.

[Berz1996b] M. Berz, K. Makino, K. Shamseddine, G. H. Hoffstätter, and W. Wan, *COSY INFINITY and its applications to nonlinear dynamics*, in Computational Differentiation: Techniques, Applications, and Tools, M. Berz, C. Bischof, G. Corliss, and A. Griewank, eds., SIAM, Philadelphia, Penn., 1996, pp. 363–365.

[Griewank1991e] A. Griewank and G. F. Corliss, eds., *Automatic Differentiation of Algorithms: Theory, Implementation, and Application*, SIAM, Philadelphia, Penn., 1991.

[Hansen1979a] E. R. Hansen, *Global optimization using interval analysis – the one-dimensional case*, J. Optim. Theory and Appl., 29 (1979), pp. 331–334.

[Hansen1988a] ——, *An overview of global optimization using interval analysis*, in Reliability in Computing, R. E. Moore, ed., Academic Press, New York, 1988, pp. 289–307.

[Hoffstätter1994a] G. H. Hoffstätter, *Rigorous Bounds on Survival Times in Circular Accelerators and Efficient Computation of Fringe–Field Transfer Maps*, PhD thesis, Michigan State University, East Lansing, Mich., 1994. Also Deutsches Elektronen-Synchrotron report DESY 94–242, Notkestraße 85, 22603 Hamburg, Germany.

[Ichida1979a] K. Ichida and Y. Fujii, *An interval arithmetic method for global optimization*, Computing, 23 (1979), pp. 85–97.

[Jansson1992a] C. Jansson, *A global optimization method using interval arithmetic*, IMACS Annals of Computing and Applied Mathematics, (1992).

[Kulisch1981a] U. W. Kulisch and W. L. Miranker, *Computer Arithmetic in Theory and Practice*, Academic Press, New York, 1981.