Nuclear Instruments and Methods in Physics Research A 363 (1995) 100–104

ELSEVIER

NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH Section A

Modern map methods for charged particle optics

Martin Berz

Department of Physics and National Superconducting Cyclotron Laboratory, Michigan State University, East Lansing, MI 48824, USA

Abstract

The differential algebraic methods, a natural tool for the description and determination of the solution of differential equations, have proven useful for the computation of aberrations of any desired order in any particle optical system. Besides conventional symplectic systems based on strongly or weakly focusing elements including fringe fields, they also readily allow the treatment of spin dynamics as well as classical synchrotron radiation. In recent years, a variety of codes have been written based on these techniques, including COSY INFINITY, which is currently used by approximately 150 registered users. Several examples of recent applications employing high order methods will be given.

While the method makes the problem of computation of Taylor maps straightforward and their manipulation and analysis convenient, for many applications it is important to have exact bounds of the truncation error. Recently it was shown how such information can be determined conveniently and with rather limited effort. Furthermore, it is often important to know the domain and the speed of convergence of the Taylor expansion. We will show that similar to the conventional DA approach, such information can be obtained by carrying the three elementary operations of addition, multiplication, and differentiation on the space of infinitely often differentiable functions to a suitable smaller space that can be described on a computer.

1. Introduction

Particle optical systems are characterized by differential equations $\boldsymbol{r}' = \boldsymbol{f}(\boldsymbol{r}, s)$ describing the evolution of the particle optical coordinates $\boldsymbol{r}$ as a function of the independent variable s, which is usually chosen to be the arc length along a reference trajectory. The function f is derived from the Cartesian equations of motion and contains information about the electromagnetic fields; the details of the form of f can be found in Refs. [1–3] and many other sources. The information about the optical system under consideration is then described by the map $\mathscr{M}$ relating initial conditions $\boldsymbol{r}_i$ at position s_i to final conditions $\boldsymbol{r}_f$ at position s_f via

$$\boldsymbol{r}_f = \mathscr{M}(\boldsymbol{r}_i, s_i, s_f).$$

In a very general sense, the solution of the differential equation defining $\mathscr{M}$ can be obtained by the manipulation of functions. In the simplest cases, exact solutions may be found; in other cases, it is necessary to resort to approximate methods. For example, if the functional dependence of $\boldsymbol{r}_f$ on s can be expanded in a Taylor series, the directional derivative $L_f = \boldsymbol{f} \cdot \boldsymbol{\nabla} + \partial_s$ (sometimes also referred to as "vector field" or "Lie derivative") can be used to relate final coordinates to initial coordinates via the propagator of the dynamical system

$$\boldsymbol{r}_f = \exp(\Delta s \cdot L_f)\boldsymbol{r}_i.$$

This propagator can be evaluated with high accuracy by keeping sufficiently many terms. Another way to provide a functional dependence between initial and final coordinates is via a numerical integration algorithm.

In all of these cases, it is required to perform manipulations in function spaces; the operations that are required are addition and multiplication (and their inverse), differentiation (and possibly its inverse) as well as possibly the application of elementary functions. A space consisting of a set A (here a set of functions) as well as an addition, multiplication, and scalar multiplication satisfying the usual conditions is called an algebra. If, furthermore, there is an operation ∂ satisfying

$$\partial(a+b) = \partial a + \partial b \quad \text{and} \quad \partial(a \cdot b) = (\partial a)b + a(\partial b)$$

for all $a, b \in A$, the structure is called a differential algebra [4,5]; so in a formal sense, what is necessary is to perform differential algebraic operations on function spaces.

Unfortunately, function spaces cannot easily be represented on the computer – in a similar sense as real numbers cannot be represented in their entirety. In the case of real numbers, it proved useful to approximate real numbers by floating point numbers, hoping that the retained information, the first n digits, are sufficient to describe whatever information is desired. To perform arithmetic, it is important that the floating point representation of sums and products can be obtained from those of the

SSDI 0168-9002(95)00361-4

respective numbers. This can be achieved by introducing a floating point addition $\oplus_F$ as well as a floating point multiplication $\odot_F$ to perform the required operations. Denoting the projection of a real number to the respective floating point number by F, we then demand

$$F(a+b) = F(a) \oplus_F F(b),$$
$$F(a \cdot b) = F(a) \odot_F F(b);$$

it is well known that the relations can only be satisfied approximately, and this is the source of the errors in practical computations. It is important to note that these errors can be accounted for in a mathematically rigorous way by using interval methods (see for example Ref. [6]).

A way to treat functions that is conceptually similar to the truncation of numbers to n digits is to retain the first n orders of their Taylor expansion. For the field of particle optics, this approach is particularly useful because in most cases it is not the map $\mathcal{M}$ proper that is needed, but rather its Taylor coefficients, the so-called aberrations. Similar to the case of floating point numbers, the goal is to compute the Taylor expansions for sum and products from those of the individual functions; in addition, to account for the differential algebraic structure, we need to do the same for the derivative. Formally this can be achieved by introducing an addition $\oplus_T$, a multiplication $\odot_T$, and a derivation ∂_T such that

$$T(f+g) = T(f) \oplus_T T(g),$$
$$T(f \cdot g) = T(f) \odot_T T(g),$$
$$T(\partial f) = \partial_T T(f).$$

First presented in Refs. [7,5], these concepts have been utilized in the code COSY INFINITY [8–10], which currently has about 150 registered users, and a variety of other codes [11–14]. COSY provides an object oriented language environment for Differential Algebraic (DA) operations, and all of the physics as well as the user commands are written in this environment. It allows the computation of aberrations of any desired order for any particle optical element, and has a large class of analysis features, some of which are discussed below.

Similar to the way in which the error performed in floating point operations can be rigorously estimated by interval methods, it is possible to obtain estimates for the error due to the approximation of the original functions due to Taylor expansion [15]; different from the interval methods for real number calculations, the interval bounds for the remainders have a tendency to become very tight. These methods are particularly useful for the determination of exact bounds for particle stability [16].

2. Maps for complicated elements

Because of their generality, DA methods are particularly useful for the computation of aberrations of complicated elements. In all of these cases, as soon as the fields of the element is known, its transfer map can be computed to arbitrary order. For many cases, it is possible to describe the fields of the particle optical elements under consideration with sufficient accuracy by certain models. For example, in the case of electric or magnetic multipoles, in many cases the fringe-field fall-off can be described rather accurately by the Enge formula, which describes the decrease of the multipole strength as a function of position via

$$E(s) = \frac{1}{1+\exp\left(a_0 + a_1\left(\frac{s}{d}\right) + \cdots + a_5\left(\frac{s}{d}\right)^5\right)}.$$

In the case of superimposed multipoles, the fall-off of each of the multipole terms can be described by an Enge function separately.

In a similar way, also the fringe field of bending magnets and electrostatic deflectors can be described. In the case of bending magnets, often the effective field boundary is tilted and curved, usually to deliberately obtain focusing and affect nonlinear terms, but sometimes also as an artifact of the construction process. In such a case of a rather complicated curve describing the effective boundary of the field, the fall-off of the fringe field is assumed to be governed by an Enge function depending on the distance to the effective field boundary, depicted in Fig. 1.

The result is a rather complicated algorithm describing the midplane field, which then has to be differentiated repeatedly to obtain the field in whole space. The computation of aberrations is complicated even more by the fact that the curvature of the reference orbit continuously changes, and unless the proper adjustments of position are made, will not line up with the desired reference orbit inside the magnet. This again entails that the actual deflection angle does not agree with the design angle between the effective field boundaries, and in addition there will be offsets in position and direction of the reference orbit.

Using the DA concept, COSY INFINITY can consider all these effects and compute accurate aberration coefficients for general dipole magnets. It is believed that it is the only code considering all the intricacies arising from the complicated field shape as well as the unusual motion of the reference orbit.

While the calculations performed by COSY are accurate to the precision of the integration, particularly at higher orders this level of sophistication has a price in computation time. For all situations but those in which the

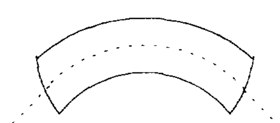

Fig. 1. A magnetic sector with tilted and curved effective field boundaries. The dashed line is the reference orbit.

highest level of sophistication is needed, it is useful to offer efficient approximate ways to treat fringe field effects. A new DA method based on the principle of symplectic scaling is introduced in a companion paper [17].

On the other hand, in some situations it is necessary to provide an even higher level of accuracy of the field description than what can be obtained from Enge models, and frequently it is advisable to use measured data directly, beyond their use to fit the proper parameters in the Enge model. The strategy for the representation of the field generated from measured data depends on whether measurements are available only in the midplane or for the whole space.

In case there are only measurements of the midplane, the challenges are on the one hand to smooth the data in a reasonable way compatible with the measurement accuracy. But perhaps more importantly, it is necessary to infer the out-of-plane information from them, a process that requires high-order differentiation [1,2,18]. It is well known that numerical differentiation of measured data is a subtle issue, and substantial thought has gone into the design of higher order differentiators suppressing spurious effects (see for example Ref. [19]).

We favor a method in which the measured data are actually interpolated by smooth functions, which then can be differentiated. As it turns out, for the objective of preserving the proper higher order derivatives, the method of Gaussian interpolation is very useful. For this purpose, Gaussian functions are placed at the measurement points, and their height is adjusted to fit the measured data; this approach is conceptually similar to the use of Gaussian beamlets for the simulation of space charge [20]. While generally being a least squares problem, in many cases a very good approximation to the optimal choices of heights can actually be obtained by simply scaling them with the value of the measured data at that point. By adjusting the width of the Gaussian, the amount of resulting smoothing can be adjusted. The resulting field in the midplane is then given by

$$B_y(\boldsymbol{r}) = \sum_i B_y(\boldsymbol{r}_i) \exp\left(- \left(\frac{\boldsymbol{r} - \boldsymbol{r}_i}{\sigma_i} \right)^2 \right);$$

because of the rapid decrease of the Gaussian, it is sufficient to restrict the sum to a few nearest grid points. The resulting field can be differentiated as often as necessary to perform the required out-of-plane expansion. As it turns out, the Gaussian method is particularly successful in preserving the high order derivatives of the functions. To judge the ability of the Gaussian method to extract higher derivatives properly, we interpolated various known functions by Gaussians and compared their known higher order derivatives with those found from the Gaussian method. The results are shown in Table 1.

The description of measured fields by Gaussian interpolation and their use for the computation of transfer maps are available directly within COSY, and have been used for a variety of projects, including the study of the S800 spectrograph described in the next section.

Table 1
Accuracy of derivative calculation for Gaussian interpolation

Function	Error in function	Error in third derivative	Error in fifth derivative
$f(x)=1$	10^{-12}	10^{-8}	10^{-6}
$f(x)=x$	10^{-12}	10^{-7}	10^{-4}
$f(x)=\cos x$	10^{-7}	10^{-5}	10^{-4}
$f(x)=\exp(-x^2)$	10^{-7}	10^{-5}	10^{-4}

In case field information is not only available inside the midplane but also in other planes, such information can be used beneficially for other field models. In case the conventional out-of-plane expansion based on a Gaussian midplane model does not reproduce the out-of-plane field data with sufficient accuracy, fully three-dimensional models are appropriate. In particular, we studied image charge models, where it proved advantageous to choose the distribution of the image charges in a Gaussian way. Such models can be used to model the total field [21], or they can be used to superimpose a weak correction field to the field predicted by the midplane method.

3. Spectrographs

The fact that DA methods allow a fully rigorous description of particle motion to even high orders makes them particularly suitable for the study of precision instruments such as spectrographs, electron microscopes, and systems for lithography. In all of these cases, the methods allow the determination of transfer maps as soon as the electromagnetic fields of the system are known.

We want to illustrate these methods for the case of the S800 [22], a high-resolution particle spectrograph under construction at NSCL at MSU. In this case, the desired energy resolution of 20000 as well as the rather large used aperture of the dipole magnets of ± 15 cm requires the consideration of the aberrations of the device to at least order five and possibly order seven. Due to saturation effects and overlapping fringe fields, it was decided to measure the fields in five planes. The information will then be used for a global field description as outlined in the last section.

Due to the multitude and size of the high-order aberrations occurring in the S800, a conventional correction of aberrations appears to be impossible. To account for the influence of aberrations, not only the positions of the particles are measured at the focal plane, but also their angles by means of a second detector separated by the first. The obtained information of horizontal positions and directions, together with the fact that the particle went through the target slit, implicitly determines the entire trajectory of the particle uniquely. Thus it also determines the particle's energy, and also the angles a and b, which shed additional light on the nuclear reaction at hand.

The use of DA methods allows a direct solution of this problem. In the transfer map $\mathscr{M}$ describing the system, exploit the fact that the initial horizontal coordinate of the particle is determined since it went through the target slit; furthermore, we ignore the parts of the map that are irrelevant for the current problem, the time of flight part, and the energy part, which here merely says that final energy equals initial energy. We obtain the reduced nonlinear relationship

$$\begin{pmatrix} x_f \\ a_f \\ y_f \\ b_f \end{pmatrix} = \mathscr{S} \begin{pmatrix} a_i \\ y_i \\ b_i \\ \delta_i \end{pmatrix},$$

which contains the measurable quantities on the left and the quantities of interest, namely δ as well as a and b, on the right as the arguments of the map. Thus, by inverting the nonlinear map $\mathscr{S}$, we can determine the required information. But the inversion of nonlinear maps is just one of several algorithms that can be performed conveniently within the DA framework; it is beyond the scope of this paper to provide details about the inversion algorithm, and we refer the reader to [18,23].

4. Spin dynamics

A rather recent problem in particle optics is the study of the dynamics of the spin of charged particles. This is relevant in several new areas, including the transport of polarized beams, the preservation of polarization during acceleration over a multitude of turns in circular accelerators, and finally the use of polarized electrons for contrast enhancement in electron microscopy.

Quite analogous to orbit dynamics, also spin dynamics can be treated in an elegant way using DA methods. The motion of magnetic moments is described in classical approximation by the BMT equation

$$\frac{\mathrm{d}s}{\mathrm{d}t} = w \times s,$$

where

$$w = k\bigg(-(1+G\gamma)\boldsymbol{B} + \frac{G}{1+\gamma}(\boldsymbol{P}\cdot\boldsymbol{B})\boldsymbol{P}\left(G + \frac{1}{1+\gamma}\right)\boldsymbol{P} \times \frac{\boldsymbol{E}}{c}\bigg),$$

and $k = e/\gamma m_0 c$, $G = (g-2)/g$, $\boldsymbol{P} = \boldsymbol{p}/m_0 c$. Due to the special linear structure, it follows that the solution of the motion is an orthogonal matrix $\hat{A}(\boldsymbol{r})$ describing the spin transformation, the coordinates of which depend on the orbital variables.

It turns out that the study of spin motion is best done by considering the evolution of the elements of the spin matrix; rephrasing the spin motion in these coordinates yields the equations of motion

$$\hat{A}'(z, s) = \hat{W}(z, s) \cdot \hat{A}(z, s);$$

this merely adds nine additional equations to the orbit equations of motion, but does not require introduction of any additional free variables beyond the six orbital quantities. The computation of the spin map can then be achieved by DA integration of motion. Similar to the orbital case, for the autonomous case it is also useful to express the problem by a propagation operator $\exp(L_f)$. However, in this case the exact form of the vector field L_f is substantially more complicated and will be discussed elsewhere [24].

5. Other applications

There are many other applications for which the use of DA methods has proven useful within the last years; while the scope of this paper may not allow their detailed discussion, we want to refer the reader to the relevant literature.

Using methods of computational theorem proving [25] and recently a new Lie-algebraic theory describing high-order achromats [26], it was possible to construct four-cell systems totally free of all aberrations up to a certain order. Various designs that were generated with COSY INFINITY based on these methods include third order [27], fourth order [28], and recently even fifth order [29] achromats.

For the study of repetitive systems like accelerators and storage rings, DA-based arbitrary order normal form methods have been developed [30]. These methods allow the computation of amplitude dependent tune shifts and shed light on nonlinear resonant behavior of repetitive systems; they have been used for the analysis and correction of the SSC low energy booster, the IUCF ring, and the PSRII ring.

Besides the study of resonances, normal form methods can also be used to perform fully rigorously guarantee long-term stability of repetitive systems utilizing arguments similar to those of the Ljapunov and Nekhoroshev stability theories. These methods, described in detail in Refs. [16,31], make use of the RDA approach discussed above and have been used for various stability estimates yielding guaranteed stability for up to approximately 10^{12} turns.

Finally, the DA methods are also useful for calculations in the field of glass optics. Recently, routines have been developed that allow the calculation of arbitrary order effects for glass optical systems, consisting of spherical and aspherical lenses, spherical, parabolic, and aspherical mirrors, as well as prisms [8].

Acknowledgements

For financial support I am grateful to the Alfred P. Sloan Foundation and the National Science Foundations. For substantial practical help and many discussions I am indebted to my students; in particular, Georg Hoffstätter for his expertise on long term stability; Weishi Wan for his knowledge of high-order achromats; Kyoko Fuchi for the development of the methods of Gaussian interpolation; and Meng Zhao for the work on glass optics.

References

[1] H. Wollnik, Charged Particle Optics (Academic Press, Orlando, Florida, 1987).

[2] P.W. Hawkes and E. Kasper, Principles of Electron Optics (Academic Press, London, 1989).

[3] M. Berz, Nucl. Instr. and Meth. A 298 (1990) 473.

[4] J.F. Ritt, Differential Algebra (American Mathematical Society, Washington, DC, 1950).

[5] M. Berz, Part. Accel. 24 (1989) 109.

[6] U.W. Kulisch and W.F. Miranker, Computer Arithmetic in Theory and Practice (Academic Press, New York, 1981).

[7] M. Berz, Nucl. Instr. and Meth. A 258 (1987) 431.

[8] M. Berz, COSY INFINITY Version 6 reference manual, Technical Report MSUCL-869, National Superconducting Cyclotron Laboratory, Michigan State University, East Lansing, MI 48824 (1993).

[9] M. Berz, Proc. 3rd Computational Accelerator Physics Conf., AIP Conf. Proc. 297 (1993) 267.

[10] M. Berz, Proc. Nonlinear Effects in Accelerators, eds. M. Berz, S. Martin and K. Ziegler (IOP Publishing, 1992) p. 125.

[11] J. van Zeijts and F. Neri, The arbitrary order design code TLIE 1.0, Proc. Workshop on Nonlinear Effects in Accelerators eds. M. Berz, S. Martin and K. Ziegler (IOP Publishing, 1993).

[12] Y. Yan, Ref. [9], p. 279.

[13] W.G. Davis, S.R. Douglas, G.D. Pusch and G.E. Lee-Whiting, The Chalk River differential algebra code DACYC and the role of differential and Lie algebras in understanding the orbit dynamics in cyclotrons, in Ref. [11].

[14] L. Michelotti, MXYZTPLK: A practical, user friendly c++ implementation of differential algebra, Technical report, Fermilab (1990).

[15] M. Berz and G. Hoffstätter, Computation and application of Taylor polynomials with interval remainder bounds, Interval Computations (1994).

[16] M. Berz and G. Hoffstätter, Exact estimates of the long term stability of weakly nonlinear systems applied to the design of large storage rings, Interval Computations, in press.

[17] G. Hoffstätter and M. Berz, these Proceedings (4th Int. Conf. on Charged Particle Optics, Tsukuba, Japan, 1994) Nucl. Instr. and Meth. A 363 (1995) 124.

[18] M. Berz, Nucl. Instr. and Meth. A 298 (1990) 426.

[19] Dong o Jeon, J. Comput. Phys., in press.

[20] M. Berz and H. Wollnik, Nucl. Instr. and Meth. A 267 (1988) 25.

[21] R. Degenhardt and M. Berz, High accuracy description of the fringe fields of particle spectrographs, Proc. 1993 Particle Accelerator Conf., Washington, DC, 1993.

[22] J. Nolen, A.F. Zeller, B. Sherrill, J.C. DeKamp and J. Yurkon, A proposal for construction of the S800 spectrograph, Technical Report MSUCL-694, National Superconducting Cyclotron Laboratory (1989).

[23] M. Berz, in: Physics of Particle Accelerators, ed. M. Month, volume AIP 249 (American Institute of Physics, 1991) p. 456.

[24] M. Berz, Differential algebraic description and analysis of spin motion, Proc. SPIN94, in press.

[25] E. Goldmann, W. Wan and M. Berz, Ref. [10], p. 201.

[26] W. Wan, A Theory of Arbitrary Order Achromats, PhD thesis, Michigan State University, East Lansing, Michigan, USA, in preparation.

[27] W. Wan, E. Goldmann and M. Berz, Ref. [9], p. 143.

[28] W. Wan and M. Berz, Design of a fourth order achromat, Proc. 1994 APS Spring Meeting, Washington, DC, 1994.

[29] W. Wan and M. Berz, Ref. [17], p. 142

[30] M. Berz, Ref. [10], p. 77.

[31] G.H. Hoffstätter, PhD thesis (1994).