Chapter 3
B-series and algebraic analysis

3.1 Introduction

This chapter is built on the broad introduction to numerical methods for initial value problems in Chapter 1 and the study of trees and related graph-theoretic structures in Chapter 2.

Initial value problems, and methods for obtaining numerical approximations to their solutions, are typically described in terms of triples (y_0, f, h), where

$$y_0 \in \mathbb{R}^N, \qquad f : \mathbb{R}^N \to \mathbb{R}^N, \qquad h \in \mathbb{R}.$$

For example, an initial value problem can be written as

$$y' = f(y), \qquad y(x_0) = y_0, \quad x_0 \in \mathbb{R}, \tag{3.1 a}$$

and the aim could be to evaluate $y(x_0 + h)$. A classical requirement is to find the "order" of a Runge–Kutta method. This single question can be viewed as a combination and a comparison of the solutions of two individual Taylor series questions.

1. Find the series expansion of the solution to (3.1 a) in the form

$$y(x_0 + h) = y_0 + hf(y_0) + \tfrac{1}{2}h^2 f'(y_0)f(y_0) + \cdots, \tag{3.1 b}$$

2. Find the series expansion of the numerical result computed by a Runge–Kutta method

$$\begin{array}{c|c} c & A \\ \hline & b^\intercal \end{array},$$

in the form

$$y_1 = y_0 + h(b^\intercal \mathbf{1})f(y_0) + h^2(b^\intercal c)f'(y_0)f(y_0) + \cdots. \tag{3.1 c}$$

J. C. Butcher, *B-Series*, Springer Series in Computational Mathematics 55,
https://doi.org/10.1007/978-3-030-70956-3_3

The term“Question” will be used in a technical sense much like “problem” is often used. It will always be concerned with a generic problem with the three ingredients: y_0, f and h. The questions discussed above are two of several questions that can be asked, for which the answers have a common pattern. That is, each of the solutions can be written as a formal series in which the terms are always the same, except for the assignment of different numerical coefficients. These are known as B-series and the coefficients referred to are “B-series coefficients”. What will be referred to as “B-series analysis”, which is the principal subject of this book, is the use of B-series coefficients as a means of analysing the underlying approximations.

The question of order is solved by equating the coefficient of h, h^2, $\dots$ in each of the two series (3.1 b) and (3.1 c), starting with

$$\begin{aligned} b^{\mathsf{T}}\mathbf{1} &= 1, \\ b^{\mathsf{T}}c &= \tfrac{1}{2}. \end{aligned}$$

A fuller suite of model questions is introduced in Section 3.2

Chapter outline

Section 3.2 begins with a discussion of covariance of possible questions that can be posed in terms of the two components of an initial value problem and a time unit. “Elementary differentials” are informally introduced and an elementary question set is introduced.

In Section 3.3, Taylor series will be reviewed in a multi-dimensional setting for use throughout the chapter. The basic components are written in terms of Fréchet derivatives $f^{(n)}(y)$, written as n-linear operators.

The central component of this chapter, Section 3.4, is concerned wih elementary differentials. These enable the terms in B-series to be constructed. Section 3.5 considers the derivation of B-series in two important cases. The principal techniques used ar repeated differentiation and evolution. The so called “elementary weights”, which arise as B-series coefficients for Runge–Kutta methods, and allow the order conditions to be expressed, are the subject of Section 3.6.

In Section 3.7, elementary differentials are found for a Kronecker product based on a tableau for a Runge–Kutta method, combined with a standard initial value problem. This leads to an alternative construction of “elementary weights”, and the order of Runge–Kutta methods. Constructibility of Runge–Kutta methods of arbitrary order, and the independence of elementary differentials follow from “attainable value” results given in Section 3.8. Finally, in Section 3.9, the composition of B-series is considered.

The principal references for this chapter are [7, 14] (Butcher, 1963, 1972) and [51, 52] (Hairer, Wanner, 1973, 1974).

3.2 Autonomous formulation and mappings

Given three ingredients

y_0 A point in the vector space $y_0 \in \mathbb{R}^N$,
f A mapping $f : \mathbb{R}^N \to \mathbb{R}^N$,
h A non-zero real number,

there are many calculations that can be made and questions that can be formulated. In this discussion, the dimension N will be an arbitrary positive integer.

In this chapter, "problem" will refer to a triple (y_0, f, h). This might be used to answer a "question" such as:

Given an initial-value problem

$$y'(x) = f(y(x)), \qquad y(x_0) = y_0,$$

what is the solution at $x = x_0 + h$?

Another question might be to evaluate the approximate solution to such a problem, using a specific numerical method.

We will distinguish between problems and questions.

Covariance of problems and questions

Given a problem formulated in terms of the triple (y_0, f, h), it is possible to rewrite the same problem in different ways by rescaling the time variable $h \mapsto sh$ or by carrying out an affine transformation $\eta \mapsto M\eta + d$, with M a nonsingular linear operator on $\mathbb{R}^N$, so that (y_0, f, h) is transformed to $(\widehat{y}_0, \widehat{f}, \widehat{h}) = (My_0 + d, Mf + d, sh)$. Consider, for example, the question of calculating a single step of the Euler method, with stepsize h. That is, evaluating y_1, where

$$y_1 = y_0 + hf(y_0). \tag{3.2 a}$$

What does it mean to transform to the $(\widehat{y}_0, \widehat{f}, \widehat{h})$ formulation?

This can be interpreted in two different ways:

- Solve the problem using transformed variables to give

$$\widehat{y}_1 = \widehat{y}_0 + \widehat{h}\widehat{f}(\widehat{y}_0), \tag{3.2 b}$$

- Solve the problem using the original variables and then rewrite in the new coordinates leading to

$$(M\widehat{y}_1 + d) = (M\widehat{y}_0 + d) + s\widehat{h}f(M\widehat{y}_0 + d),$$

which simplifies to

$$\widehat{y}_1 = \widehat{y}_0 + s\widehat{h}M^{-1}f(M\widehat{y}_0 + d). \tag{3.2 c}$$

The two results (3.2 b) and (3.2 c) are identical, if and only if

$$\widehat{f}(\widehat{y}) = sM^{-1}f(M\widehat{y} + d), \quad \text{for all } \widehat{y},$$

or, what is equivalent,

$$f(y) = s^{-1} M \widehat{f}\big(M^{-1}(y-d)\big), \quad \text{for all } y.$$

This example singles out the question of evaluating (3.2 a) as having the property referred to in Definition 3.2A below. But there are many questions also satisfying the same property and these will be the questions that are of significance in this book

Definition 3.2A A question $Q(y_0, h, f)$ is covariant if the following diagram commutes

$$\begin{array}{ccc} (y_0, h, f) & \xrightarrow{Q} & Q(y_0, h, f) \\ \Big\downarrow T & & \Big\downarrow T \\ (\widehat{y}_0, \widehat{h}, \widehat{f}) & \xrightarrow{Q} & Q(\widehat{y}_0, \widehat{h}, \widehat{f}) \end{array} \tag{3.2 d}$$

where the transformation $T : (y_0, h, f) \longmapsto (\widehat{y}_0, \widehat{h}, \widehat{f})$ satisfies

$$y_0 = M\widehat{y}_0 + d, \tag{3.2 e}$$
$$h = s\widehat{h}, \tag{3.2 f}$$
$$f(y) = s^{-1} M \widehat{f}\big(M^{-1}(y-d)\big). \tag{3.2 g}$$

In studying the properties of covariant questions in more detail, write corresponding mappings as $\mathfrak{a}_h$, $\mathfrak{b}_h$, ... and the mapped values as $\mathfrak{a}_h y_0$, $\mathfrak{b}_h y_0$, ...

Definition 3.2B If $\mathfrak{a}_0 y_0 = y_0$, then $\mathfrak{a}_h$ is a central mapping; if $\mathfrak{a}_0 y_0 = 0$, then $\mathfrak{a}_h$ is a differential mapping.

For example, $y_0 \mapsto y_0 + hf(y_0)$ is a central mapping and $y_0 \mapsto hf\big(y_0 + hf(y_0)\big)$ is a differential mapping. Note the factor h with every use of f because of the factors s and s^{-1} in (3.2 f) and (3.2 g) respectively. Furthermore, a factor of h is possible only in this context.

The factor M appearing in Definition 3.2A corresponds merely to a linear change of basis and many properties of covariance can be deduced using a simplified transformation in which $s = 1$ and $M = I$.

That is, we can argue from a form of the commutation diagram (3.2 d) in which (3.2 e)–(3.2 g) are replaced by

$$y_0 = \widehat{y}_0 + d, \tag{3.2 h}$$
$$h = \widehat{h},$$
$$f(y) = \widehat{f}(y-d). \tag{3.2 i}$$

Lemma 3.2C The mapping $a_h = hf(\theta y_0)$, for θ a real parameter, is covariant if and only if $\theta = 1$.

Proof. Apply the covariance criterion, using transformation rules (3.2 h) – (3.2 i), to a_h; it is found that

$$h\widehat{f}\big(\theta(\widehat{y}_0 + d) - d\big) = h\widehat{f}(\theta\widehat{y}_0),$$

implying that $\theta = 1$. □

Two introductory questions

Two questions given below, (3.2 m) and (3.2 n), will be solved to within $\mathscr{O}(h^4)$ using Taylor series. This discussion also has the role of a first introduction to "elementary differentials".

Approximations using Taylor series

It is interesting to ask what approximations of the form

$$E = C_1 h + C_2 h^2 + \cdots + C_p h^p + \mathscr{O}(h^{p+1})$$

can be found by iterated use of expressions of the form

$$hf(y_0 + E), \tag{3.2 j}$$

where E has a similar form. We will confine ourselves, in this informal discussion, to $p = 3$. We will start the iterations using the single term $E = \theta hf(y_0)$, where θ is a constant. By Taylor's theorem we obtain the result

$$hf(y_0) + \theta h^2 f'(y_0)f(y_0) + \theta^2 \tfrac{1}{2!} h^3 f''(y_0)\big(f(y_0), f(y_0)\big) + \mathscr{O}(h^4), \tag{3.2 k}$$

where the expression $f''(y_0)\big(f(y_0), f(y_0)\big)$ means the evaluation of the bilinear operator $f''(y_0)$ acting on two copies of $f(y_0)$. Taking linear combinations of (3.2 k), for several distinct values of θ, the individual terms are found, to within $\mathscr{O}(h^4)$. By carrying out additional iterations of (3.2 j), further terms can be found, such as

$$h^3 f'(y_0)f'(y_0)f(y_0), \quad h^4 f'(y_0)f'(y_0)f'(y_0)f(y_0), \quad h^4 f'(y_0)f''(y_0)f(y_0)^2$$

The first four elementary differentials

In Section 3.4, we will introduce quantities, known as elementary differentials, of which those up to order 3 are

$$\begin{aligned}
F_1 &:= f(y_0),\\
F_2 &:= f'(y_0)f(y_0),\\
F_3 &:= f''(y_0)f(y_0)^2,\\
F_4 &:= f'(y_0)f'(y_0)f(y_0).
\end{aligned}$$

Our aim, in this preliminary discussion, is to use series of the form

$$y_0 + a_1 hF_1 + a_2 h^2 F_2 + \tfrac{1}{2} a_3 h^3 F_3 + a_4 h^3 F_4, \tag{3.2l}$$

to answer the two important questions:

Question 1: Find $y(x_0+h)$, where y satisfies the initial value problem:

$$y'(x) = f\big(y(x)\big), \quad y(x_0) = y_0. \tag{3.2 m}$$

Question 2: Find y_1, satisfying the functional equation:

$$y_1 = y_0 + hf(y_1). \tag{3.2 n}$$

In each case the answer is intended to be a series approximation to within $\mathscr{O}(h^4)$. Note that the factor $\frac{1}{2}$, in the term $\frac{1}{2}a_3h^3F_3$ in (3.2l), will be seen to be an instance of a systematic scheme to simplify the manipulations.

As a preliminary to answering the two questions, evaluate the first terms of $hf(\overline{y})$, where $\overline{y}$ is given by (3.2l). We have

$$\begin{aligned}
&hf(y_0 + a_1 hF_1 + a_2 h^2 F_2 + \tfrac{1}{2} a_3 h^3 F_3 + a_4 h^3 F_4)\\
&= hf(y_0) + hf'(y_0)\big(a_1 hF_1 + a_2 h^2 F_2\big) + \tfrac{1}{2} hf''(y_0)\big(a_1 hF_1\big)^2 + \mathscr{O}(h^4)\\
&= hF_1 + a_1 h^2 F_2 + \tfrac{1}{2} a_1^2 h^3 F_3 + a_2 h^3 F_4 + \mathscr{O}(h^4).
\end{aligned}$$

For Question 1, we write the two sides of (3.2 m) separately. The left-hand side is

$$\begin{aligned}
&h(\mathrm{d}/\mathrm{d}h)(y_0 + a_1 hF_1 + a_2 h^2 F_2 + \tfrac{1}{2} a_3 h^3 F_3 + a_4 h^3 F_4)\\
&= a_1 hF_1 + 2a_2 h^2 F_2 + \tfrac{3}{2} a_3 h^3 F_3 + 3a_4 h^3 F_4 + \mathscr{O}(h^4),
\end{aligned}$$

with the right-hand side equal to

$$\begin{aligned}
&hf(y_0 + a_1 hF_1 + a_2 h^2 F_2 + \tfrac{1}{2} a_3 h^3 F_3 + a_4 h^3 F_4)\\
&= hF_1 + a_1 h^2 F_2 + \tfrac{1}{2} a_1^2 h^3 F_3 + a_2 F_4 + \mathscr{O}(h^4),
\end{aligned}$$

so that matching coefficients gives

$$a_1 = 1, \quad a_2 = \tfrac{1}{2}, \quad a_3 = \tfrac{1}{3}, \quad a_4 = \tfrac{1}{6}.$$

For Question 2, (3.2l) becomes, to within $\mathscr{O}(h^4)$,

$$\begin{aligned}
&a_1 hF_1 + a_2 h^2 F_2 + \tfrac{1}{2} a_3 h^3 F_3 + a_4 h^3 F_4\\
&= hf(y_0 + a_1 hF_1 + a_2 h^2 F_2 + \tfrac{1}{2} a_3 h^3 F_3 + a_4 h^3 F_4)\\
&= hF_1 + a_1 h^2 F_2 + \tfrac{1}{2} a_1^2 h^3 F_3 + a_2 h^3 F_4,
\end{aligned}$$

and we deduce that

$$a_1 = a_2 = a_3 = a_4 = 1.$$

Questions 1 and 2 that have been considered are flow_h and $\mathit{implicit}_h$ respectively, from the easy question set, which we now introduce. These will be used from time to time as examples of various results as they are introduced.

The easy question set

The questions in this set, which range from trivial to challenging, are as follows, where in each case, a_h is acting on y_0.

id_h	Give the value of y_0
slope_h	Calculate $hf(y_0)$
Euler_h	Calculate $y_1 = y_0 + hf(y_0)$
$\mathit{implicit}_h$	Calculate y_1, where $y_1 = y_0 + hf(y_1)$
$\mathit{mid\text{-}point}_h$	Calculate y_1, where $y_1 = y_0 + hf\big(\frac{1}{2}(y_0 + y_1)\big)$
flow_h	Evaluate $y_1 = y(x_0 + h)$, where $y(x)$ satisfies the initial value problem $y'(x) = f\big(y(x)\big)$, $y(x_0) = y_0$
$\mathit{flow\text{-}slope}_h$	Evaluate $hy'(x_0 + h)$, where $y(x)$ satisfies the initial value problem $y'(x) = f\big(y(x)\big)$, $y(x_0) = y_0$
$\mathit{Runge\text{-}I}_h$	The numerical method given by (1.5 c) (p. 20)
$\mathit{Runge\text{-}II}_h$	The numerical method given by (1.5 d) (p. 20)

Exercise 32 Find the Taylor series for $\mathit{mid\text{-}point}_h$ to within $\mathscr{O}(h^4)$ in terms of F_1, F_2, F_3, F_4.

As an example, we consider $\mathit{Runge\text{-}I}_h$ We have in turn, working to within $\mathscr{O}(h^4)$,

$$\begin{aligned}
Y_1 &= y_0 &&= y_0\\
hF_1 &= hf(Y_1) &&= \qquad hF_1,\\
Y_2 &= y_0 + hF_1 &&= y_0 + hF_1,\\
hF_2 &= hf(Y_2) &&= \qquad hF_1 + \tfrac{1}{2}h^2F_2 + \tfrac{1}{6}h^3F_3,\\
y_1 &= y_0 + \tfrac{1}{2}hF_1 + \tfrac{1}{2}hF_2 &&= y_0 + hF_1 + \tfrac{1}{4}h^2F_2 + \tfrac{1}{12}h^3F_3,
\end{aligned}$$

giving a result, identical with flow_h to within $\mathscr{O}(h^3)$.

Exercise 33 Carry out a similar analysis for $\mathit{Runge\text{-}II}_h$.

Exercise 34 Carry out a similar analysis for $\mathit{flow\text{-}slope}_h$.

3.3 Fréchet derivatives and Taylor series

Use of Polish notation

In this section, and throughout this book, we will use a Polish notation for writing the result of applying a linear or multilinear operator to a vector or sequence of vectors [70] (Łukasiewicz,Tarski,1930). For example, if $\boldsymbol{d}_1$ is a linear operator, and $\boldsymbol{d}_2$ is a bilinear operator, then $\boldsymbol{d}_1 v$ and $\boldsymbol{d}_2 vw$ will denote $\boldsymbol{d}_1(v)$and $\boldsymbol{d}_2(v,w)$, respectively. A similar notation has been adopted in [2] (Azamov, Bekimov, 2016).

Fréchet derivatives

The principal reference for this section is [81] (Rudin, 1976).

If f is Fréchet differentiable at $y = y_0$, then a linear operator $f'(y_0)$ exists so that the approximation

$$f(y_0+\varepsilon) \approx f(y_0) + f'(y_0)\varepsilon \tag{3.3 a}$$

holds to within $\mathrm{o}(\|\varepsilon\|)$. That is, (3.3 a) should hold so accurately that

$$\frac{f(y_0+\varepsilon) - f(y_0) - f'(y_0)\varepsilon}{\|\varepsilon\|}$$

can be made arbitrarily small by making $\|\varepsilon\|$ small enough but non-zero.

In terms of individual components,

$$f^i(y_0+\varepsilon) = f^i(y_0) + \sum_{j=1}^{N} f^i_j(y_0)\varepsilon^j + \mathrm{o}(\|\varepsilon\|), \qquad i = 1,2,\dots,N,$$

where f^i_j denotes the partial derivative

$$f^i_j = \frac{\partial f^i}{\partial y^j}.$$

Second and higher derivatives

If, for arbitrary δ, the function $f'(y)\delta$ is also Fréchet differentiable at $y = y_0$, with derivative $f''(y_0)\delta$ then, for a vector ε with small norm, it holds that

$$f'(y_0+\varepsilon)\delta = f'(y_0)\delta + f''(y_0)\delta\varepsilon + \mathrm{o}(\|\varepsilon\|).$$

Repeating this process, we can, for a multiply-differentiable function, define a sequence of linear and multi-linear operators

$$f'(y_0), \quad f''(y_0), \quad f^{(3)}(y_0), \quad \dots$$

Acting with these operators on a sequence of arbitrary vectors $\delta_1, \delta_2, \ldots$, we obtain

$$f'(y_0)\delta_1, \quad f''(y_0)\delta_1\delta_2, \quad f^{(3)}(y_0)\delta_1\delta_2\delta_3, \quad \ldots$$

The differentiability of each member of this sequence provides the definition of the next term in the sequence of high-order derivatives. That is, assuming the estimates exist,

$$\begin{aligned}&f^{(n)}(y_0)\delta_1\delta_2,\ldots\delta_{n-1}\varepsilon\\&= f^{(n-1)}(y_0+\varepsilon)\delta_1\delta_2,\ldots\delta_{n-1} - f^{(n-1)}(y_0)\delta_1\delta_2,\ldots\delta_{n-1} + \mathrm{o}(\varepsilon).\end{aligned}$$

Write individual components in partial derivative form and we have

$$\begin{aligned}(e^i)^\intercal f'(y_0)\delta_1 &= \sum_j f^i_j \delta^j_1,\\ (e^i)^\intercal f''(y_0)\delta_1\delta_2 &= \sum_{jk} f^i_{jk}\delta^j_1\delta^k_2,\\ (e^i)^\intercal f^{(3)}(y_0)\delta_1\delta_2\delta_3 &= \sum_{jk\ell} f^i_{jk\ell}\delta^j_1\delta^k_2\delta^\ell_3,\end{aligned}$$

and similarly for higher order derivatives.

This chapter will work largely with Taylor expansions about a standard base point y_0 and, hence, it will be convenient to have a special way of writing Fréchet derivatives evaluated at this point. Accordingly we will write

$$\begin{aligned}\mathbf{f} &:= f(y_0),\\ \mathbf{f}' &:= f'(y_0),\\ \mathbf{f}'' &:= f''(y_0),\\ &\vdots \quad \vdots\\ \mathbf{f}^{(n)} &:= f^{(n)}(y_0).\end{aligned} \tag{3.3 b}$$

If y_0 is replaced by some other argument, such as y_1, then we will write

$$\mathbf{f}^{(n)}(y_1) := f^{(n)}(y_1).$$

However, it is emphasized, that $\mathbf{f}^{(n)}$, without an argument, will always denote $f^{(n)}(y_0)$.

Taylor series

The approximation of a function of a real variable by its Taylor expansion

$$f(y_0+\delta) = f(y_0) + f'(y_0)\delta + \tfrac{1}{2!}f''(y_0)\delta^2 + \cdots + \tfrac{1}{n!}f^{(n)}(y_0)\delta^n + \mathrm{o}(\delta^n), \tag{3.3 c}$$

is at the heart of numerical analysis but we will need to use it in a multi-dimensional context. If $f : \mathbb{R}^N \to \mathbb{R}^N$, then, as we have seen, $f'(y_0)$ becomes a linear operator characterized by

$$f(y_0+\delta)=f(y_0)+f'(x_0)\delta+\mathrm{o}(\|\delta\|)$$

or, in terms of individual components,

$$f^i(y_0+\delta)=f^i(y_0)+\sum_{j=1}^{N}f^i_j(y_0)\delta^j+\mathrm{o}(\|\delta\|),\qquad i=1,2,\dots,N,$$

where f^i_j denotes $\partial f^i/\partial y_0^j$. We will adopt the "summation convention" so that (3.3 c) can be written more compactly as

$$f^i(y_0+\delta)=f^i(y_0)+f^i_j(y_0)\delta^j+\mathrm{o}(\|\delta\|).$$

In this notation, higher order Taylor series can also be written

$$f^i(y_0+\delta)=f^i(y_0)+f^i_j(y_0)\delta^j+\tfrac{1}{2!}f^i_{jk}(y_0)\delta^j\delta^k+\mathrm{o}(\|\delta\|^2),$$
$$f^i(y_0+\delta)=f^i(y_0)+f^i_j(y_0)\delta^j+\tfrac{1}{2!}f^i_{jk}(y_0)\delta^j\delta^k+\tfrac{1}{3!}f^i_{jk\ell}(y_0)\delta^j\delta^k\delta^\ell+\mathrm{o}(\|\delta\|^3).$$

To write the multi-dimensional case in a similar style to (3.3 c), we will use the sequence defined by (3.3 b), where we note that $\mathbf{f}$ denotes the vector $f(y_0)$ and $\mathbf{f}'$ denotes the linear operator $\mathbf{f}^{(1)}(y_0)$. Similarly $\mathbf{f}''=\mathbf{f}^{(2)}$ denotes the bilinear operator that gives the result

$$\mathbf{f}''\delta_1\delta_2=\mathbf{f}''\delta_2\delta_1,$$

with component i equal to

$$f^i_{jk}(y_0)\delta_1^j\delta_2^k,$$

and a similar pattern for higher order derivatives, evaluated at y_0.

The designation of specific components denoted by superscripts will be avoided as much as possible and will not be used without any warning. This enables us to use the notation δ^n to denote the juxtaposition of n copies of δ. Hence, the multi-dimensional version of Taylor's theorem can be written

$$f(y_0+\delta)=\mathbf{f}+\sum_{n=1}^{p}\frac{1}{n!}\mathbf{f}^{(n)}\delta^n+\mathrm{o}(\|\delta\|^p). \tag{3.3 d}$$

A typical term in this series, containing the factor $\mathbf{f}^{(n)}\delta^n$, is traditionally written

$$\mathbf{f}^{(n)}(\delta,\delta,\dots,\delta).$$

However, writing this instead as

$$\mathbf{f}^{(n)}\delta\delta\cdots\delta \text{ or } \mathbf{f}^{(n)}\delta^n,$$

within the Taylor formula (3.3 d), gives an advantage and an added clarity because of the linearity inherent in the notation. It is also advantageous to adopt the Polish notation in understanding terms like this. When the operations are performed in order from the left, the first step gives $\mathbf{f}^{(n)}\delta$ which becomes an $n-1$-fold linear operator. This, in turn, operates on a second copy of δ to yield an $n-2$-fold linear operator and so on, until $\mathbf{f}^{(n)}\delta^n$ becomes a constant vector. Alternatively, we can simply regard $\mathbf{f}^{(n)}$ as an operator requiring n operands which, in this case, are each equal to δ.

Symmetry of higher derivatives

Under sufficient smoothness assumptions the Fréchet derivative $\mathbf{f}^{(n)}$, as an n-linear operator is symmetric. That is, if π is any permutation of $\{1,2,\ldots,n\}$, then

$$\mathbf{f}^{(n)}\delta_1\delta_2\cdots\delta_n = \mathbf{f}^{(n)}\delta_{\pi 1}\delta_{\pi 2}\cdots\delta_{\pi n}.$$

This is a standard result [81] (Rudin, 1976).

Taylor's series for a sum of perturbations

If δ in (3.3 d) is replaced by a sum of the form

$$\delta = \delta_1 + \delta_2 + \cdots,$$

then terms like δ^n can be evaluated using the multinomial theorem. For example, if $\delta = \delta_1 + \delta_2 + \delta_3$, then

$$\begin{aligned}
\delta^2 &= \delta_1^2 + 2\delta_1\delta_2 + 2\delta_1\delta_3 + \delta_2^2 + 2\delta_2\delta_3 + \delta_3^2,\\
\delta^3 &= \delta_1^3 + 3\delta_1^2\delta_2 + 3\delta_1^2\delta_3 + 3\delta_1\delta_2^2 + 6\delta_1\delta_2\delta_3 + 3\delta_1\delta_3^2\\
&\quad + \delta_2^3 + 3\delta_2^2\delta_3 + 3\delta_2\delta_3^2 + \delta_3^3,
\end{aligned}$$

which leads to the Taylor series up to cubic terms:

$$\begin{aligned}
f(y_0+\delta_1+\delta_2+\delta_3) = \mathbf{f} &+ \mathbf{f}'\delta_1 + \mathbf{f}'\delta_2 + \mathbf{f}'\delta_3\\
&+ \tfrac{1}{2}\mathbf{f}''\delta_1^2 + \mathbf{f}''\delta_1\delta_2 + \mathbf{f}''\delta_1\delta_3 + \tfrac{1}{2}\mathbf{f}''\delta_2^2 + \mathbf{f}''\delta_2\delta_3 + \tfrac{1}{2}\mathbf{f}''\delta_3^2\\
&+ \tfrac{1}{6}\mathbf{f}^{(3)}\delta_1^3 + \tfrac{1}{2}\mathbf{f}^{(3)}\delta_1^2\delta_2 + \tfrac{1}{2}\mathbf{f}^{(3)}\delta_1^2\delta_3\\
&+ \tfrac{1}{2}\mathbf{f}^{(3)}\delta_1\delta_2^2 + \mathbf{f}^{(3)}\delta_1\delta_2\delta_3 + \tfrac{1}{2}\mathbf{f}^{(3)}\delta_1\delta_3^2\\
&+ \tfrac{1}{6}\mathbf{f}^{(3)}\delta_2^3 + \tfrac{1}{2}\mathbf{f}^{(3)}\delta_2^2\delta_3 + \tfrac{1}{2}\mathbf{f}^{(3)}\delta_2\delta_3^2 + \tfrac{1}{6}\mathbf{f}^{(3)}\delta_3^3.
\end{aligned}$$

Arbitrary number of terms

In a more general case, where there is an arbitrary number of terms in the sum $\delta = \sum \delta_i$, the coefficient of $\mathbf{f}^{(n)}\delta_1^{k_1}\delta_2^{k_2}\cdots$ is equal to

$$\frac{1}{k_1!k_2!\cdots} = \frac{1}{n!}\binom{n}{k_1,k_2,\ldots}, \qquad (3.3\,\mathrm{e})$$

in accordance with the multinomial theorem.

Introduce the set $\mathbf{N}$ of sequences of non-negative integers, such that for $\mathbf{n} \in \mathbf{N}$, $\mathbf{n}_i$ is eventually zero. Conventionally, an infinite sequence of zeros can be omitted from the notation. For given $\mathbf{n}$, $\mathbf{n}! := \prod_{i=1}^{\infty} \mathbf{n}_i!$. We are now in a position to write Taylor's theorem in the convenient form

Theorem 3.3A

$$f(\mathbf{y}_0 + \delta) = \sum_{\mathbf{n}\in\mathbf{N}} \frac{1}{\mathbf{n}!} \mathbf{f}^{(|\mathbf{n}|)} \delta^{\mathbf{n}}. \tag{3.3 f}$$

Proof. This is a consequence of the remarks following (3.3 e). □

The set $\mathbf{N}$ is denumerable; that is, it is possible to write its members in a sequence. A convenient way of doing this is to list the members in increasing values of $\sum_{i=0}^{\infty} i\, \mathbf{n}_i!$, with equally ranked items sorted arbitrarily amongst themselves. This would give a sequence beginning with the members

$$[0,0,0,0,0],\ [1,0,0,0,0],\ [2,0,0,0,0],\ [0,1,0,0,0],\ [3,0,0,0,0],\ [1,1,0,0,0],$$
$$[0,0,1,0,0],\ [4,0,0,0,0],\ [2,1,0,0,0],\ [1,0,1,0,0],\ [0,0,0,1,0]$$

and the sum in (3.3 f) becomes

$$\begin{aligned}&\mathbf{f} + \mathbf{f}'\delta_1 + \tfrac{1}{2}\mathbf{f}''\delta_1^2 + \mathbf{f}'\delta_2 + \tfrac{1}{6}\mathbf{f}^{(3)}\delta_1^3 + \mathbf{f}''\delta_1\delta_2 + \mathbf{f}'\delta_3\\ &+ \tfrac{1}{24}\mathbf{f}^{(4)}\delta_1^4 + \tfrac{1}{2}\mathbf{f}^{(3)}\delta_1^2\delta_2 + \mathbf{f}''\delta_1\delta_3 + \mathbf{f}'\delta_4 + \cdots\end{aligned}$$

3.4 Elementary differentials and B-series

B-series to order 4

We will reconsider the questions in the easy question set, not for a particular one-dimensional problem, but for a general N-dimensional autonomous problem of the form

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f(y), \quad y(x_0) = y_0.$$

Of the given problems, the first which might create difficulties is $\mathit{implicit}_h$ and we will attempt to solve this by iteration, starting from $y_1 \approx y_0$. The next iteration will be

$$y_1 \approx y_0 + hf\big(y_0 + hf(y_0)\big).$$

To explore this approximation in greater detail we need to remind ourselves of the nature of $f(y_0 + \delta)$, where $\delta = hf(y_0)$. We will adopt a notation $\mathbf{f} := f(y_0)$. Assuming f is analytic, Taylor's theorem can be used to obtain

$$y_0 + hf(y_0 + h\mathbf{f}) = y_0 + h\mathbf{f} + h^2\mathbf{f}'\mathbf{f} + \tfrac{1}{2}h^3\mathbf{f}''\mathbf{f}^2 + \tfrac{1}{6}h^4\mathbf{f}^{(3)}\mathbf{f}^3 + \mathscr{O}(h^5).$$

For a second iteration, use Theorem 3.2C with

$$\begin{aligned}\delta_1 &= h\mathbf{f},\\ \delta_2 &= h^2\mathbf{f}'\mathbf{f},\\ \delta_3 &= \tfrac{1}{2}h^3\mathbf{f}''\mathbf{f}^2.\end{aligned}$$

Note that, because we are developing the series only to h^4 terms, the term $\frac{1}{6}h^4\mathbf{f}^{(3)}\mathbf{f}^3$ is omitted from $hf(y_0+h\mathbf{f}+h^2\mathbf{f}'\mathbf{f}+\frac{1}{2}h^3\mathbf{f}''\mathbf{f}^2+\frac{1}{6}h^4\mathbf{f}^{(3)}\mathbf{f}^3)$ in the second iteration. For the same reason, we need only a limited number of $\mathbf{n}$ terms; these are

$$\mathbf{n} = \left\{\begin{array}{l}[0,0,0],\\ [1,0,0],\\ [2,0,0],\\ [3,0,0],\\ [0,1,0],\\ [1,1,0],\\ [0,0,1],\end{array}\right.$$

and the second iteration gives the series

$$\begin{aligned}&y_0+hf(y_0+h\mathbf{f}+h^2\mathbf{f}'\mathbf{f}+\tfrac{1}{2}h^3\mathbf{f}''\mathbf{f}^2)\\ &= y_0+\overset{[0,0,0]}{h\mathbf{f}}+\overset{[1,0,0]}{h^2\mathbf{f}'\mathbf{f}}+\overset{[2,0,0]}{\tfrac{1}{2}h^3\mathbf{f}''\mathbf{f}^2}+\overset{[3,0,0]}{\tfrac{1}{6}h^4\mathbf{f}^{(3)}\mathbf{f}^3}+\overset{[0,1,0]}{h^3\mathbf{f}'\mathbf{f}'\mathbf{f}}+\overset{[1,1,0]}{h^4\mathbf{f}''\mathbf{f}\mathbf{f}'\mathbf{f}}+\overset{[0,0,1]}{\tfrac{1}{2}h^4\mathbf{f}'\mathbf{f}''\mathbf{f}^2}+\mathcal{O}(h^5),\end{aligned}$$

with the corresponding components of $\mathbf{n}$ written above the terms.

A final iteration adds one more term $h^4\mathbf{f}'\mathbf{f}'\mathbf{f}'\mathbf{f}$ leading to the final approximation to this order; which we show with the terms arranged in a more natural order.

$$\begin{aligned}y_1 = y_0+h\mathbf{f}+h^2\mathbf{f}'\mathbf{f}+\tfrac{1}{2}h^3\mathbf{f}''\mathbf{f}^2+h^3\mathbf{f}'\mathbf{f}'\mathbf{f}&\\ +\tfrac{1}{6}h^4\mathbf{f}^{(3)}\mathbf{f}^3+h^4\mathbf{f}''\mathbf{f}\mathbf{f}'\mathbf{f}+\tfrac{1}{2}h^4\mathbf{f}'\mathbf{f}''\mathbf{f}^2+h^4\mathbf{f}'\mathbf{f}'\mathbf{f}'\mathbf{f}+\mathcal{O}(h^5)&.\end{aligned}$$

We can write the Taylor series for all members of the easy question set, up to order 4, in a similar way to $\textit{implicit}_h$ with only the numerical coefficients in the various terms altered.

For $\textit{mid-point}_h$ the result can be derived by replacing h by $h/2$ in the series for $\textit{implicit}_h$ to obtain a series for the solution to

$$\widehat{y} = y_0+hf\big(y_0+\tfrac{1}{2}h\widehat{y}\big),$$

and then writing $y = 2\widehat{y}-y_0$ for the series for $\textit{mid-point}_h$. For $\textit{flow}_h$ the quoted result can be checked by first evaluating $\textit{flow-slope}$ and comparing this with the h derivative of $\textit{flow}$. For $\textit{Runge-I}_h$ and $\textit{Runge-II}_h$ the calculations are straightforward. The results we have already found for $\textit{implicit}_h$ will be used as a base and the other

Table 12 Coefficients for the easy question set, up to h^4 terms

Problem	y_0	$h\mathbf{f}$	$h^2\mathbf{f}'\mathbf{f}$	$\frac{1}{2}h^3\mathbf{f}''\mathbf{f}^2$	$h^3\mathbf{f}'\mathbf{f}'\mathbf{f}$	$\frac{1}{6}h^4\mathbf{f}^{(3)}\mathbf{f}^3$	$h^4\mathbf{f}''\mathbf{f}\mathbf{f}'\mathbf{f}$	$\frac{1}{2}h^4\mathbf{f}'\mathbf{f}''\mathbf{f}^2$	$h^4\mathbf{f}'\mathbf{f}'\mathbf{f}'\mathbf{f}$
id_h	1	0	0	0	0	0	0	0	0
slope_h	0	1	0	0	0	0	0	0	0
Euler_h	1	1	0	0	0	0	0	0	0
$\mathit{implicit}_h$	1	1	1	1	1	1	1	1	1
$\mathit{mid\text{-}point}_h$	1	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
flow_h	1	1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{12}$	$\frac{1}{24}$
$\mathit{flow\text{-}slope}_h$	0	1	1	1	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$
$\mathit{Runge\text{-}I}_h$	1	1	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}$	0	0	0
$\mathit{Runge\text{-}II}_h$	1	1	$\frac{1}{2}$	$\frac{1}{4}$	0	$\frac{1}{8}$	0	0	0

coefficients will be written in terms of these. This information will be presented in Table 12.

Elementary differentials

The limited results we have considered up to order 4 suggest a general form for the terms appearing in the Taylor expansions for arbitrary orders.

Definition 3.4A The elementary differential associated with a tree $\mathrm{t} = [\mathrm{t}_1\mathrm{t}_2\cdots\mathrm{t}_n]$ is

$$\mathbf{F}(\mathrm{t}) = \mathbf{f}^{(n)}\mathbf{F}(\mathrm{t}_1)\mathbf{F}(\mathrm{t}_2)\cdots\mathbf{F}(\mathrm{t}_n), \tag{3.4 a}$$

with $\mathbf{F}(\tau) = \mathbf{f}$.

Note that t can be written as $\tau_n\mathrm{t}_1\mathrm{t}_2\cdots\mathrm{t}_n$, closely matching the structure of (3.4 a).

To see how (3.4 a) is used, evaluate the elementary differentials up to order 4 in detail:

$$\begin{aligned}
&\mathbf{F}(\mathbf{t}_1) && = \mathbf{f}, && \\
&\mathbf{F}(\mathbf{t}_2) = \mathbf{F}([\mathbf{t}_1]) && = \mathbf{f}'\mathbf{F}(\mathbf{t}_1) && = \mathbf{f}'\mathbf{f}, \\
&\mathbf{F}(\mathbf{t}_3) = \mathbf{F}([\mathbf{t}_1^2]) && = \mathbf{f}''\mathbf{F}(\mathbf{t}_1)\mathbf{F}(\mathbf{t}_1) && = \mathbf{f}''\mathbf{f}\mathbf{f}, \\
&\mathbf{F}(\mathbf{t}_4) = \mathbf{F}([\mathbf{t}_2]) && = \mathbf{f}'\mathbf{F}(\mathbf{t}_2) && = \mathbf{f}'(\mathbf{f}'\mathbf{f}) \quad = \mathbf{f}'\mathbf{f}'\mathbf{f}, \\
&\mathbf{F}(\mathbf{t}_5) = \mathbf{F}([\mathbf{t}_1^3]) && = \mathbf{f}^{(3)}\mathbf{F}(\mathbf{t}_1)\mathbf{F}(\mathbf{t}_1)\mathbf{F}(\mathbf{t}_1) && = \mathbf{f}^{(3)}\mathbf{f}\mathbf{f}\mathbf{f}, \\
&\mathbf{F}(\mathbf{t}_6) = \mathbf{F}([\mathbf{t}_1\mathbf{t}_2]) && = \mathbf{f}''\mathbf{F}(\mathbf{t}_1)\mathbf{F}(\mathbf{t}_2) && = \mathbf{f}''\mathbf{f}(\mathbf{f}'\mathbf{f}) = \mathbf{f}''\mathbf{f}\mathbf{f}'\mathbf{f}, \\
&\mathbf{F}(\mathbf{t}_7) = \mathbf{F}([\mathbf{t}_3]) && = \mathbf{f}'\mathbf{F}(\mathbf{t}_3) && = \mathbf{f}'(\mathbf{f}''\mathbf{f}\mathbf{f}) = \mathbf{f}'\mathbf{f}''\mathbf{f}\mathbf{f}, \\
&\mathbf{F}(\mathbf{t}_8) = \mathbf{F}([\mathbf{t}_4]) && = \mathbf{f}'\mathbf{F}(\mathbf{t}_4) && = \mathbf{f}'(\mathbf{f}'\mathbf{f}'\mathbf{f}) = \mathbf{f}'\mathbf{f}'\mathbf{f}'\mathbf{f}.
\end{aligned}$$

Elementary differentials up to order 6 are shown, for reference, in Table 13 (p. 114).

B-series

B-series are a formalism for expressing the Taylor series for the solution to questions written in terms of the triple (y_0, h, f). They are always written in terms of elementary differentials which in turn are indexed by the trees introduced in Chapter 2 (p. 39).

We will now look at some possible choices of coefficients. Let B^* denote the set of all mappings from $\varnothing \cup T$ to $\mathbb{R}$. Also write B^0 for the linear subspace of B^* such that if $a \in \mathsf{B}^0$ then $a(\varnothing) = 0$. Similarly write B for the subset of B^* such that if $a \in \mathsf{B}$ then $a(\varnothing) = 1$. Before we consider these properties systematically, we will look at some examples.

Given $a \in \mathsf{B}^*$, we want to use series covered by

Definition 3.4B The B-series $(\mathbf{B}_h y_0)a$ is a formal series defined by

$$(\mathbf{B}_h y_0)a = a(\varnothing)y_0 + \sum_{\mathrm{t}\in\mathrm{T}} \frac{h^{|t|}a(\mathrm{t})}{\sigma(\mathrm{t})}\mathbf{F}(\mathrm{t}). \tag{3.4 b}$$

Recall the use of the symmetry $\sigma(\mathrm{t})$ in (3.4 b), defined in Definition 2.5A (p. 58). Those series for which $a \in \mathsf{B}$ have a special role because, for such a series, $hf\big((\mathbf{B}_h y_0)a\big)$ also exists as a B-series. These correspond to the central mappings referred to in Definition 3.2B (p. 102).

Definition 3.4C A B-series $(\mathbf{B}_h y_0)a$ is said to be a central series if $a \in \mathsf{B}$.

B-series as a formal inner product

The sequence of elementary differentials, together with the value of the base point, usually y_0, when suitably scaled by $h^{|\mathrm{t}|}/\sigma(\mathrm{t})$, can be written as an infinite row vector

$$\mathbf{B}_h y_0 = \begin{bmatrix} y_0 & h\mathbf{f} & h^2\mathbf{f}'\mathbf{f} & \tfrac{1}{2}h^3\mathbf{f}''\mathbf{f}\mathbf{f} & h^3\mathbf{f}'\mathbf{f}'\mathbf{f} & \cdots & \sigma(\mathrm{t})^{-1}h^{|\mathrm{t}|}\mathbf{F}(\mathrm{t}) & \cdots \end{bmatrix}.$$

Similarly, the B-series coefficients are conveniently written as an infinite column vector

$$a = \begin{bmatrix} a(\varnothing) & a(\mathbf{t}_1) & a(\mathbf{t}_1) & a(\mathbf{t}_3) & a(\mathbf{t}_4) & \cdots & a(\mathrm{t}) & \cdots \end{bmatrix}^{\mathsf{T}}.$$

Table 13 Elementary differentials to order 6

$\lvert t \rvert$	t		$\mathbf{F}(t)$	$\lvert t \rvert$	t		$\mathbf{F}(t)$
1	$\mathbf{t}_1$		$\mathbf{f}$	6	$\mathbf{t}_{18}$		$\mathbf{f^{(5)}f^5}$
				6	$\mathbf{t}_{19}$		$\mathbf{f^{(4)}f^3f'f}$
2	$\mathbf{t}_2$		$\mathbf{f'f}$	6	$\mathbf{t}_{20}$		$\mathbf{f^{(3)}f^2f''f^2}$
				6	$\mathbf{t}_{21}$		$\mathbf{f^{(3)}f^2f'f'f}$
3	$\mathbf{t}_3$		$\mathbf{f''f^2}$	6	$\mathbf{t}_{22}$		$\mathbf{f^{3)}ff'ff'f}$
3	$\mathbf{t}_4$		$\mathbf{f'f'f}$	6	$\mathbf{t}_{23}$		$\mathbf{f''ff^{(3)}f^3}$
				6	$\mathbf{t}_{24}$		$\mathbf{f''ff''ff'f}$
4	$\mathbf{t}_5$		$\mathbf{f^{(3)}f^3}$	6	$\mathbf{t}_{25}$		$\mathbf{f''ff'f''f^2}$
4	$\mathbf{t}_6$		$\mathbf{f''ff'f}$	6	$\mathbf{t}_{26}$		$\mathbf{f''ff'f'f'f}$
4	$\mathbf{t}_7$		$\mathbf{f'f''f^2}$	6	$\mathbf{t}_{27}$		$\mathbf{f''f'ff''f^2}$
4	$\mathbf{t}_8$		$\mathbf{f'f'f'f}$	6	$\mathbf{t}_{28}$		$\mathbf{f''f'ff'f'f}$
				6	$\mathbf{t}_{29}$		$\mathbf{f'f^{(4)}f^4}$
5	$\mathbf{t}_9$		$\mathbf{f^{(4)}f^4}$	6	$\mathbf{t}_{30}$		$\mathbf{f'f^{(3)}f^2f'f}$
5	$\mathbf{t}_{10}$		$\mathbf{f^{(3)}f^2f'f}$	6	$\mathbf{t}_{31}$		$\mathbf{f'f''ff''f^2}$
5	$\mathbf{t}_{11}$		$\mathbf{f''ff''f^2}$	6	$\mathbf{t}_{32}$		$\mathbf{f'f''ff'f'f}$
5	$\mathbf{t}_{12}$		$\mathbf{f''ff'f'f}$	6	$\mathbf{t}_{33}$		$\mathbf{f'f''f'ff'f}$
5	$\mathbf{t}_{13}$		$\mathbf{f''f'ff'f}$	6	$\mathbf{t}_{34}$		$\mathbf{f'f'f'f^{(3)}f^3}$
5	$\mathbf{t}_{14}$		$\mathbf{f'f^{(3)}f^3}$	6	$\mathbf{t}_{35}$		$\mathbf{f'f'f''ff'f}$
5	$\mathbf{t}_{15}$		$\mathbf{f'f''ff'f}$	6	$\mathbf{t}_{36}$		$\mathbf{f'f'f'f''f^2}$
5	$\mathbf{t}_{16}$		$\mathbf{f'f'f''f^2}$	6	$\mathbf{t}_{37}$		$\mathbf{f'f'f'f'f'f}$
5	$\mathbf{t}_{17}$		$\mathbf{f'f'f'f'f}$				

Up to trees of order 3, these can be written in the form

$$(\mathbf{B}_h y_0)a = \begin{bmatrix} y_0 & h\mathbf{F}(\bullet) & h^2\mathbf{F}(\overset{\bullet}{\bullet}) & \frac{1}{2}h^3\mathbf{F}({}^{\bullet}\!\vee\!{}^{\bullet}) & h^3\mathbf{F}(\overset{\bullet}{\overset{\bullet}{\bullet}}) & \cdots \end{bmatrix} \begin{bmatrix} a(\varnothing) \\ a(\bullet) \\ a(\overset{\bullet}{\bullet}) \\ a({}^{\bullet}\!\vee\!{}^{\bullet}) \\ a(\overset{\bullet}{\overset{\bullet}{\bullet}}) \\ \vdots \end{bmatrix}.$$

B-series for the easy question set

Our aim will now be to extend the results given in Table 12 (p. 112) to all trees so that we know the exact B-series for all members of the easy question set.

id_h, slope_h and Euler_h

We can immediately write down the B-series for these three questions

> Theorem 3.4D The B-series coefficients for the questions referred to are:
>
> $$\begin{aligned} \mathit{id}_h:&\ a(\varnothing)=1, & a(\mathrm{t})&=0, & & & \mathrm{t}&\in \mathrm{T},\\ \mathit{slope}_h:&\ a(\varnothing)=0, & a(\tau)&=1, & a(\mathrm{t})&=0, & |\mathrm{t}|&>1,\\ \mathit{Euler}_h:&\ a(\varnothing)=1, & a(\tau)&=1, & a(\mathrm{t})&=0, & |\mathrm{t}|&>1. \end{aligned}$$

Some special members of B^*: $\mathrm{E} \in \mathsf{B}$

Write E for the B-series for for unit step h. That is

$$\begin{aligned} \mathrm{E}(\varnothing) &= 1,\\ \mathrm{E}(\mathrm{t}) &= \frac{1}{\mathrm{t}!}, \qquad \mathrm{t}\in\mathrm{T}. \end{aligned}$$

For the flow through a distance θh, we write

$$\begin{aligned} \mathrm{E}^{(\theta)}(\varnothing) &= 1,\\ \mathrm{E}^{(\theta)}(\mathrm{t}) &= \frac{\theta^{|\mathrm{t}|}}{\mathrm{t}!}, \qquad \mathrm{t}\in\mathrm{T}. \end{aligned}$$

We can also write $\mathrm{E}^{\theta} = \mathrm{E}^{(\theta)}$.

Some special members of B^*: $\mathrm{D} \in \mathsf{B}^0$

Scaled differentiation to produce $hf(y_0) = hF(\tau)(y_0)$, that is *slope*, is given by

$$\begin{aligned} \mathrm{D}(\varnothing) &= 0, \\ \mathrm{D}(\tau) &= 1, \\ \mathrm{D}(\mathrm{t}) &= 0, \qquad |\mathrm{t}| > 1. \end{aligned}$$

In Section 3.9 (p. 133) we will introduce compositions of B-series. We will give a single example here. If $a \in \mathsf{B}$, then the central series $(\mathbf{B}_h y_0)a$ can be substituted for y_0 in $(\mathbf{B}_h y_0)\mathrm{D}$ to give an expression for $hf\big((\mathbf{B}_h y_0)a\big)$. This expression can be expanded as a B-series in its own right, which will be written as $(\mathbf{B}_h y_0)a\mathrm{D}$. The notation $a\mathrm{D}$ will later be seen to be part of the general and self-consistent terminology introduced in Section 3.9.

Theorem 3.4E If $a \in \mathsf{B}$, then

$$(a\mathrm{D})(\varnothing) = 0, \tag{3.4 c}$$
$$(a\mathrm{D})(\tau) = 1, \tag{3.4 d}$$
$$(a\mathrm{D})(\mathrm{t}) = \prod_{i=1}^{m} a(\mathrm{t}_i), \qquad \mathrm{t} = [\mathrm{t}_1\mathrm{t}_2\cdots\mathrm{t}_m]. \tag{3.4 e}$$

Proof. Use Theorem 3.3A with $\delta = \mathbf{B}(a, y_0, h) - y_0$. First verify (3.4 c,3.4 d), from

$$hf\big(y_0 + \mathscr{O}(h)\big) = hf(y_0) + \mathscr{O}(h^2).$$

To prove (3.4 e) with

$$\mathrm{t} = [\mathrm{t}_1^{k_1}\mathrm{t}_2^{k_2}\cdots\mathrm{t}_m^{k_m}], \qquad \mathrm{t}_1, \mathrm{t}_2, \ldots, \mathrm{t}_m \text{ distinct},$$

write $\mathbf{n} = [k_1, k_2, \ldots, k_m]$. The $\mathbf{n}$ term in (3.3 f) becomes

$$h^{1+\sum k_i|\mathrm{t}_i|}\prod_{i=1}^{m}\frac{a(\mathrm{t}_i)^{k_i}}{(k_i)!\sigma(\mathrm{t}_i)^{k_i}}\mathbf{f}^{(\sum k_i|\mathrm{t}_i|)}\mathbf{F}(\mathrm{t}_1)^{k_1}\mathbf{F}(\mathrm{t}_2)^{k_2}\cdots\mathbf{F}(\mathrm{t}_m)^{k_m} = (a\mathrm{D})(\mathrm{t})\frac{h^{|\mathrm{t}|}}{\sigma(t)}\mathbf{F}(\mathrm{t}).$$

□

Written another way, Theorem 3.4E states that

$$hf\big((\mathbf{B}_h y_0)a\big) = \sum_{\mathrm{t}\in T} h^{|\mathrm{t}|}\sigma(\mathrm{t})^{-1}(a\mathrm{D})(\mathrm{t})\mathbf{F}(\mathrm{t}). \tag{3.4 f}$$

The following result is proved in the same way as Theorem 3.4E and expressed as a generalization of (3.4 f)

Corollary 3.4F For $n = 1, 2, 3, \ldots,$

$$h^n f^{(n)}\big((\mathbf{B}_h y_0)a\big) = \sum_{\mathrm{t}\in T} h^{|\mathrm{t}|+n-1}\sigma(\mathrm{t})^{-1}(a\mathrm{D})(\mathrm{t})\mathbf{F}(\mathrm{t}*^{n-1}).$$

Exercise 35 Show that

$$(\mathrm{ED})(\mathrm{t}) = \frac{|\mathrm{t}|}{\mathrm{t}!}.$$

3.5 B-series for flow_h and $\mathit{implicit}_h$

B-series based on repeated differentiation

Recall the discussion on partitions introduced in Chapter 2, Section 2.6 (p. 65). Our aim in this section is to derive the B-series coefficients for flow_h and $\mathit{implicit}_h$ based on repeated differentiation, starting from the two equation systems

$$y'(x) = f(y(x)), \qquad y(x_0) = y_0, \tag{3.5 a}$$

$$y(x) = y_0 + (x - x_0) f\big(y(x)\big), \tag{3.5 b}$$

respectively, and making use of partitions.

B-series for flow_h

Starting from the initial value problem (3.5 a) further differentiations give in turn

$$\begin{aligned} y''(x) &= f'\big(y(x)\big)y'(x), \\ y^{(3)}(x) &= f''\big(y(x)\big)y'(x)y'(x) + f'\big(y(x)\big)y''(x), \\ y^{(4)}(x) &= f^{(3)}\big(y(x)\big)y'(x)y'(x)y'(x) \\ &\qquad + 3f''\big(y(x)\big)y'(x)y''(x) + f'\big(y(x)\big)y^{(3)}(x). \end{aligned} \tag{3.5 c}$$

Evaluating the Taylor coefficients at $x = x_0$, we find the equations given in (3.5 d), where we have included also $\mathbf{y}^{(5)}$

$$\begin{aligned} \mathbf{y}' &= \mathbf{f}, \\ \mathbf{y}'' &= \mathbf{f}'\mathbf{y}', \\ \mathbf{y}^{(3)} &= \mathbf{f}''\mathbf{y}'\mathbf{y}' + \mathbf{f}'\mathbf{y}'', \\ \mathbf{y}^{(4)} &= \mathbf{f}^{(3)}\mathbf{y}'\mathbf{y}'\mathbf{y}' + 3\mathbf{f}''\mathbf{y}'\mathbf{y}'' + \mathbf{f}'\mathbf{y}^{(3)}, \\ \mathbf{y}^{(5)} &= \mathbf{f}^{(4)}\mathbf{y}'\mathbf{y}'\mathbf{y}'\mathbf{y}' + 6\mathbf{f}^{(3)}\mathbf{y}'\mathbf{y}'\mathbf{y}'' + 4\mathbf{f}''\mathbf{y}'\mathbf{y}^{(3)} + 3\mathbf{f}''\mathbf{y}''\mathbf{y}'' + \mathbf{f}'\mathbf{y}^{(4)}. \end{aligned} \tag{3.5 d}$$

It is not necessary to refer back to (3.5 c) to go from one line to the next. It is sufficient to note that each term in (3.5 d) corresponds to a specific term in (3.5 c) which is differentiated using the product rule and mapped back to specific terms with $h = 0$. This is illustrated below for the two types of factor.

$$\begin{tikzcd} f^{(m)}(y(x)) \arrow[r] & f^{(m+1)}(y(x))y'(x) \arrow[d] \\ \mathbf{f}^{(m)} \arrow[u] \arrow[r, dashed] & \mathbf{f}^{(m+1)}\mathbf{y}' \end{tikzcd} \qquad \begin{tikzcd} y^{(n)}(x) \arrow[r] & y^{(n+1)}(x) \arrow[d] \\ \mathbf{y}^{(n)} \arrow[u] \arrow[r, dashed] & \mathbf{y}^{(n+1)} \end{tikzcd}$$

We will illustrate the derivation of $\mathbf{y}^{(4)}$ from $\mathbf{y}^{(3)}$ in (3.5 d).

$$\begin{aligned} \mathbf{f}''\mathbf{y}'\mathbf{y}' + \mathbf{f}'\mathbf{y}'' &\mapsto \left(\mathbf{f}^{(3)}\mathbf{y}'\mathbf{y}'\mathbf{y}' + \mathbf{f}''\mathbf{y}''\mathbf{y}' + \mathbf{f}''\mathbf{y}'\mathbf{y}''\right) + \left(\mathbf{f}''\mathbf{y}'\mathbf{y}'' + \mathbf{f}'\mathbf{y}^{(3)}\right) \\ &= \mathbf{f}^{(3)}\mathbf{y}'\mathbf{y}'\mathbf{y}' + 3\mathbf{f}''\mathbf{y}'\mathbf{y}'' + \mathbf{f}'\mathbf{y}^{(3)}. \end{aligned}$$

Exercise 36 Verify the formula given for $\mathbf{y}^{(5)}$ in (3.5 d).

Comparison with tree evolution

Now compare the derivation in (3.5 d), shown in Figure 9, with the evolution of partitions shown in Figure 2.6 (p. 71). In Figure 9 every node in column number

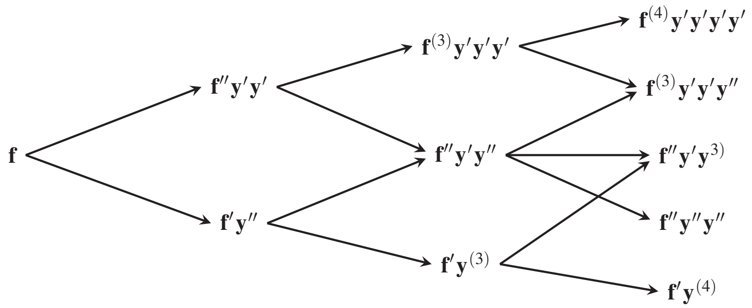


Figure 9 Succesive differentiation of $f(y(x))$ with evaluation at $x = x_0$

$n = 1, 2, \ldots$ is labelled with a term of the form

$$\mathbf{f}^{(m)}\mathbf{y}^{(k_1)}\mathbf{y}^{(k_2)}\cdots\mathbf{y}^{(k_m)},$$

where $\sum_{i=1}^{m} k_i = n$. This term evolves to

$$\mathbf{f}^{(m+1)}\mathbf{y}'\mathbf{y}^{(k_1)}\mathbf{y}^{(k_2)}\cdots\mathbf{y}^{(k_m)}+\mathbf{f}^{(m)}\mathbf{y}^{(k_1+1)}\mathbf{y}^{(k_2)}\cdots\mathbf{y}^{(k_m)}+\cdots+\mathbf{f}^{(m)}\mathbf{y}^{(k_1)}\mathbf{y}^{(k_2)}\cdots\mathbf{y}^{(k_m+1)}.$$

In exactly the same way, the partition $k_1+k_2+\cdots+k_m$ evolves to the sum of the $m+1$ partitions

$$\begin{array}{lllll}
1+ & k_1 & +\ k_2 & +\cdots+ & k_m,\\
 & (k_1+1) & +\ k_2 & +\cdots+ & k_m,\\
 & k_1 & +(k_2+1) & +\cdots+ & k_m,\\
 & & \vdots & & \\
 & k_1 & +\ k_2 & +\cdots+ & (k_m+1).
\end{array}$$

Define

$$\begin{aligned}
z(p)(x) &= f^{(\mathrm{card}(p))}(y(x))y^{(p_1)}(x)y^{(p_2)}(x)\cdots y^{(p_{\mathrm{card}(p)})}(x),\\
\mathbf{z}(p) := z(p)(x_0) &= \mathbf{f}^{(\mathrm{card}(p))}\mathbf{y}^{(p_1)}\mathbf{y}^{(p_2)}\cdots\mathbf{y}^{(p_{\mathrm{card}(p)})}.
\end{aligned}$$

This notation can be generalized to a linear combination of partitions

$$\begin{aligned}
\Big(\sum_{i=1}^{j}C_i\widehat{p}_i(x)\Big) &= \sum_{i=1}^{j}C_i\mathbf{z}(\widehat{p}_i)(x),\\
\mathbf{z}\Big(\sum_{i=1}^{j}C_i\widehat{p}_i\Big) &= \sum_{i=1}^{j}C_i\mathbf{z}(\widehat{p}_i),
\end{aligned}$$

and, in a special case,

$$\frac{\mathrm{d}}{\mathrm{d}x}z(p)(x)=z\big(\mathrm{evolve}(p)(x)\big).$$

This means that each term corresponds to a partition of n which can be written

$$\mathbf{z}(p) := \mathbf{f}^{(\mathrm{card}(p))}\mathbf{y}^{(p_1)}\mathbf{y}^{(p_2)}\cdots\mathbf{y}^{(p_{\mathrm{card}(p)})}.$$

For convenience in this section, define

$$\begin{aligned}
z^{(p)}(x) &= f^{(\mathrm{card}(p))}(y(x))y^{(p_1)}(x)y^{(p_2)}(x)\cdots y^{(p_{\mathrm{card}(p)})}(x),\\
\mathbf{z}^{(p)} &= z^{(p)}(x_0),\\
&= \mathbf{f}^{(\mathrm{card}(p))}\mathbf{y}^{(p_1)}\mathbf{y}^{(p_2)}\cdots\mathbf{y}^{(p_{\mathrm{card}(p)})}.
\end{aligned}$$

With this notation, the formulae for $\mathbf{y}^{(i)}$, for $i=1,2,3,4$, are

$$\begin{aligned}
\mathbf{y}^{(1)} &= \mathbf{z}^{()},\\
\mathbf{y}^{(2)} &= \mathbf{z}^{(1)},\\
\mathbf{y}^{(3)} &= \mathbf{z}^{(2)}+\mathbf{z}^{(1+1)},\\
\mathbf{y}^{(4)} &= \mathbf{z}^{(3)}+3\mathbf{z}^{(1+2)}+\mathbf{z}^{(1+1+1)},
\end{aligned}\tag{3.5 e}$$

or, in terms of partitions of a set using the function ϕ, which maps partitions of a set to partitions of a number,

$$\begin{aligned}
\mathbf{y}^{(1)} &= \mathbf{z}^{(\phi())},\\
\mathbf{y}^{(2)} &= \mathbf{z}^{(\phi(\mathsf{1}))},\\
\mathbf{y}^{(3)} &= \mathbf{z}^{(\phi(\mathsf{12}))} + \mathbf{z}^{(\phi(\mathsf{1+2}))},\\
\mathbf{y}^{(4)} &= \mathbf{z}^{(\phi(\mathsf{123}))} + \mathbf{z}^{(\phi(\mathsf{1+23}))} + \mathbf{z}^{(\phi(\mathsf{2+13}))} + \mathbf{z}^{(\phi(\mathsf{3+12}))} + \mathbf{z}^{(\phi(\mathsf{1+2+3}))},\\
\mathbf{y}^{(4)} &= \mathbf{z}^{(\phi(\mathsf{123}))} + \mathbf{z}^{(\phi(\mathsf{1+23}))} + \mathbf{z}^{(\phi(\mathsf{2+13}))} + \mathbf{z}^{(\phi(\mathsf{3+12}))} + \mathbf{z}^{(\phi(\mathsf{1+2+3}))}.
\end{aligned} \tag{3.5 f}$$

***Informal explanation of* (3.5 e) *and* (3.5 f)**

A partition in $\mathscr{P}(n)$ is written as a sum of positive integer with total n. For $n = 3$ as in the final line of (3.5 e), the partitions are 3, $1+2$ and $1+1+1$, and correspond to $\mathbf{f}'\mathbf{y}^{(3)}$, $\mathbf{f}''\mathbf{y}^{(1)}\mathbf{y}^{(2)}$, and $\mathbf{f}'''\mathbf{y}^{(1)}\mathbf{y}^{(1)}\mathbf{y}^{(1)}$, respectively. The five partitions in $\mathscr{P}[\{\mathsf{1},\mathsf{2},\mathsf{3}\}]$ are $\mathsf{123}$, a compact expression for $\{\mathsf{1},\mathsf{2},\mathsf{3}\}$, and four further partitions. The way ϕ interrelates the number and set partitions in (3.5 e) and (3.5 f) is shown in (3.5 g).

$$\begin{aligned}
\phi(\mathsf{1}) &= 1,\\
\phi(\mathsf{12}) &= 2,\\
\phi(\mathsf{1}+\mathsf{1}) &= 1+1,\\
\phi(\mathsf{123}) &= 3,\\
\left.\begin{matrix}\phi(\mathsf{1}+\mathsf{23})\\ \phi(\mathsf{2}+\mathsf{13})\\ \phi(\mathsf{3}+\mathsf{12})\end{matrix}\right\} &= 1+2,\\
\phi(\mathsf{1}+\mathsf{2}+\mathsf{3}) &= 1+1+1.
\end{aligned} \tag{3.5 g}$$

In the case of the empty set, corresponding to $n = 0$, there are no partitions, but for the current application, an empty partition with zero cardinality is conventionally used to represent $\mathbf{y}^{(1)} = \mathbf{z}^{(\phi())}$ as $z(x_0) = \mathbf{f}$.

Combining partitions

It is convenient to combine the partitions of n in the formula for $y^{(n)}$ into a single term. Let

$$Z(x) = f\big(y(x)\big),$$

so that

$$Z^{(n)}(x) = \sum_{p\in\mathscr{P}(n)} \text{p-weight}(p)\, z^{(p)}(x)$$

with

$$\begin{aligned}\mathbf{Z}^{(n)} &= \sum_{p\in\mathscr{P}(n)} \text{p-weight}(p)\,\mathbf{z}^{(p)}\\ &= \sum_{p\in\mathscr{P}(n)} \text{p-weight}(p)\mathbf{f}^{(\text{card}(p))}\mathbf{y}^{(p_1)}\mathbf{y}^{(p_2)}\cdots\mathbf{y}^{(p_{\text{card}(p)})}.\end{aligned} \tag{3.5 h}$$

B-series for *implicit*$_h$

Starting from (3.5 b), and carrying out further differentiations, we havet

$$\begin{aligned}y'(x) &= Z(x) + (x-x_0)Z''(x),\\ y''(x) &= 2Z''(x) + (x-x_0)Z''(x),\\ y^{(3)(x)} &= 3Z''(x) + (x-x_0)Z^{(3)}(x).\end{aligned}$$

Substituting $x = x_0$, we obtain

$$\begin{aligned}\mathbf{y}' &= \mathbf{Z} &&= \mathbf{f},\\ \mathbf{y}'' &= 2\mathbf{Z}' &&= 2\mathbf{f}'\mathbf{y}',\\ \mathbf{y}^{(3)} &= 3\mathbf{Z}'' &&= 3\big(\mathbf{f}''\mathbf{y}'\mathbf{y}' + \mathbf{f}'\mathbf{y}''\big).\end{aligned}$$

This suggests a general result which we now present.

Theorem 3.5A The Taylor coefficients for *implicit*$_h$ are

$$\mathbf{y}^{(n)} = n\mathbf{Z}^{(n-1)}. \qquad n = 1,2,\ldots \tag{3.5 i}$$

Proof. Write (3.5 b) in the form

$$y(x) = y_0 + (x-x_0)Z(x),$$

leading to

$$y'(x) = Z(x) + (x-x_0)Z'(x).$$

This is the case $n = 1$ of

$$y^{(n)}(x) = nZ^{(n-1)}(x) + (x-x_0)Z^{(n)}(x), \quad n = 1,2,\ldots$$

For $n > 1$, use induction

$$\begin{aligned}y^{(n)}(x) &= \frac{\mathrm{d}}{\mathrm{d}x}\big((n-1)Z^{(n-2)}(x) + (x-x_0)Z^{(n-1)}(x)\big)\\ &= (n-1)Z^{(n-1)}(x) + \big((x-x_0)Z^{(n)}(x) + Z^{(n-1)}(x)\big)\\ &= nZ^{(n-1)}(x) + (x-x_0)Z^{(n)}(x).\end{aligned}$$

Substituting $x = x_0$, (3.5 i) follows. □

Taylor series written as B-series

B-series for flow_h

Starting from $\mathbf{y}' = \mathbf{f} = \mathbf{F}(\mathbf{t}_1)$, we can in turn derive formulae for $\mathbf{y}''$ and $\mathbf{y}^{(3)}$ and continue. In this successive derivation, the details are omitted for $\mathbf{y}^{(4)}$ and $\mathbf{y}^{(5)}$.

$$
\begin{aligned}
\mathbf{y}' &= \mathbf{f} = \mathbf{F}(\mathbf{t}_1),\\
\mathbf{y}'' &= \mathbf{f}'\mathbf{F}(\mathbf{t}_1) = \mathbf{F}([\mathrm{t}_1]) = \mathbf{F}(\mathbf{t}_2),\\
\mathbf{y}^{(3)} &= \mathbf{f}''\mathbf{F}(\mathbf{t}_1)\mathbf{F}(\mathbf{t}_1) + \mathbf{f}'\mathbf{F}(\mathbf{t}_2) = \mathbf{F}([\mathbf{t}_1^2]) + \mathbf{F}([\mathbf{t}_2]) = \mathbf{F}(\mathrm{t}_3) + \mathbf{F}(\mathrm{t}_4),\\
\mathbf{y}^{(4)} &= \mathbf{F}(\mathbf{t}_5) + 3\mathbf{F}(\mathbf{t}_6) + \mathbf{F}(\mathbf{t}_7) + \mathbf{F}(\mathbf{t}_8),\\
\mathbf{y}^{(5)} &= \mathbf{F}(\mathbf{t}_9) + 6\mathbf{F}(\mathbf{t}_{10}) + 4\mathbf{F}(\mathbf{t}_{11}) + 3\mathbf{F}(\mathbf{t}_{12}) + 4\mathbf{F}(\mathbf{t}_{13})\\
&\qquad + \mathbf{F}(\mathbf{t}_{14}) + 3\mathbf{F}(\mathbf{t}_{15}) + \mathbf{F}(\mathbf{t}_{16}) + \mathbf{F}(\mathbf{t}_{17}).
\end{aligned}
\tag{3.5 j}
$$

Referring back to Chapter 2 Section 2.5 (p. 58) and the function α introduced there, with value given in Theorem 2.5F (p. 61), we observe that, as far as we have carried out the calculations, the coefficient of $\mathbf{F}(\mathrm{t})$ in the expression for $\mathbf{y}^{(|\mathrm{t}|)}$ is identical with $\alpha(\mathrm{t})$. This observation is true in general.

Lemma 3.5B For flow_h,

$$\mathbf{y}^{(n)} = \sum_{|\mathrm{t}|=n} \alpha(\mathrm{t})\mathbf{F}(\mathrm{t}).$$

Note the close relationship between this result and Theorem 2.6F (p. 73).

Proof. Use induction starting with $n = 1$, for which the result holds. To find the coefficient of $\mathbf{F}(\mathrm{t})$ in $\mathbf{y}^{(n)}$, where $\mathrm{t} = [\mathrm{t}_1\mathrm{t}_2\cdots\mathrm{t}_m]$, and $|\mathrm{t}| = n$ and $\sum_{i=1}^m n_i = n-1$, where $n_i = |\mathrm{t}_i|$, assume that the result is true for lower orders so that the coefficient of $\mathbf{F}(\mathrm{t}_i)$ in $\mathbf{y}^{(|\mathrm{t}_i|)}$ is $\alpha(\mathrm{t}_i)$. The term $\mathbf{F}(\mathrm{t})$ appears in the term of (3.5 h) for which $p = \sum_{i=1} |\mathrm{t}_i|$. The required coefficient is

$$
\begin{aligned}
\text{p-weight}(p)\prod_{i=1}^m \alpha(\mathrm{t}_i) &= \frac{(n-1)!}{\prod_{i=1}^m n_i!(i!)^{n_i}} \prod_{i=1}^m \frac{n_i!}{\sigma(\mathrm{t}_i)\mathrm{t}_i!}\\
&= \frac{n!}{(\prod_{i=1}^m (i!)^{n_i}\sigma(\mathrm{t}_i))(n\prod_{i=1}^m \mathrm{t}_i!)}\\
&= \frac{n!}{\sigma(\mathrm{t})\mathrm{t}!}\\
&= \alpha(\mathrm{t}).
\end{aligned}
\tag{3.5 k}
$$

□

We can now rewrite the B-series coefficients

Theorem 3.5C The B-series for flow_h is given by $(\mathbf{B}_h y_0)^{\intercal} a$, where

$$a(\varnothing) = 1,$$
$$a(\mathrm{t}) = \frac{1}{\mathrm{t}!}, \quad \mathrm{t} \in \mathrm{T}.$$

Proof. From Lemma 3.5B the coefficient of $\mathbf{F}(\mathrm{t})/|\mathrm{t}|!$ is $\alpha(\mathrm{t})/|\mathrm{t}|! = 1/\sigma(\mathrm{t})\mathrm{t}!$. □

B-series for implicit$_h$

In a similar way to the flow, we start with $\mathbf{y}' = \mathbf{f} = \mathbf{F}(\mathbf{t}_1)$ and carry out similar calculations for $\mathbf{y}''$ and $\mathbf{y}^{(3)}$, showing details, and state the results, without details of the derivation, for $\mathbf{y}^{(4)}$ and $\mathbf{y}^{(5)}$.

$$\begin{aligned}
\mathbf{y}' &= \mathbf{f} = \mathbf{F}(\mathbf{t}_1),\\
\mathbf{y}'' &= 2\mathbf{f}'\mathbf{F}(\mathbf{t}_1) = 2\mathbf{F}([\mathrm{t}_1]) = 2\mathbf{F}(\mathbf{t}_2),\\
\mathbf{y}^{(3)} &= 3\mathbf{f}''\mathbf{F}(\mathbf{t}_1)\mathbf{F}(\mathbf{t}_1) + 3\mathbf{f}'\mathbf{F}(\mathbf{t}_2) = 3\mathbf{F}([\mathbf{t}_1^2]) + 6\mathbf{F}([\mathbf{t}_2]) = 3\mathbf{F}(\mathrm{t}_3) + 6\mathbf{F}(\mathrm{t}_4),\\
\mathbf{y}^{(4)} &= 4\mathbf{F}(\mathbf{t}_5) + 24\mathbf{F}(\mathbf{t}_6) + 12\mathbf{F}(\mathbf{t}_7) + 24\mathbf{F}(\mathbf{t}_8),\\
\mathbf{y}^{(5)} &= 5\mathbf{F}(\mathbf{t}_9) + 60\mathbf{F}(\mathbf{t}_{10}) + 60\mathbf{F}(\mathbf{t}_{11}) + 120\mathbf{F}(\mathbf{t}_{12}) + 60\mathbf{F}(\mathbf{t}_{13})\\
&\qquad + 20\mathbf{F}(\mathbf{t}_{14}) + 120\mathbf{F}(\mathbf{t}_{15}) + 60\mathbf{F}(\mathbf{t}_{16}) + 120\mathbf{F}(\mathbf{t}_{17}).
\end{aligned} \tag{3.51}$$

The examples in (3.51) give results identical to those in (3.5 j) except that for each tree t, $\alpha(\mathrm{t})$ is replaced by $\beta(\mathrm{t})$. The general result holds in the form

Lemma 3.5D For $\mathit{implicit}_h$,

$$\mathbf{y}^{(n)} = \sum_{|\mathrm{t}|=n} \beta(\mathrm{t})\mathbf{F}(\mathrm{t}).$$

Proof. The proof is as for Theorem 3.5B except for the consequences of the factor n in (3.5 i). The manipulations in (3.5 k) are replaced by

$$\begin{aligned}
n\,\text{p-weight}(p)\prod_{i=1}^{m}\beta(\mathrm{t}_i) &= \frac{n(n-1)!}{\prod_{i=1}^{m} n_i!(i!)^{n_i}}\prod_{i=1}^{m}\frac{n_i!}{\sigma(\mathrm{t}_i)}\\
&= \frac{n!}{(\prod_{i=1}^{m}(i!)^{n_i}\sigma(\mathrm{t}_i))}\\
&= \frac{n!}{\sigma(\mathrm{t})}\\
&= \beta(\mathrm{t}).
\end{aligned}$$

□

Theorem 3.5E The B-series for $\mathit{implicit}_h$ is given by $(\mathbf{B}_h y_0)^\intercal a$, where

$$\begin{aligned} a(\varnothing) &= 1, \\ a(\mathrm{t}) &= 1, \quad \mathrm{t} \in \mathrm{T}. \end{aligned}$$

Proof. From Lemma 3.5D the coefficient of $\mathbf{F}(\mathrm{t})/|\mathrm{t}|!$, is $\beta \mathrm{t})/|\mathrm{t}|! = 1/\sigma(\mathrm{t})$. □

B-series for $\mathit{implicit}_h$ ***by*** antipode

We introduce an alternative approach to finding the result of Theorem 3.5E.

Reversing y_0 and y_1 in

$$y_1 = y_0 + hf(y_1),$$

we find the Euler method with negative stepsize

$$\begin{array}{c|c} 0 & \\ \hline & -1 \end{array}.$$

Hence, the B-series for $\mathit{implicit}_h$ is a^{-1}, where

$$\begin{aligned} a(\varnothing) &= 1, \\ a(\tau) &= -1, \\ a(\mathrm{t}) &= 0. \qquad |\mathrm{t}| \geq 2. \end{aligned}$$

To find the antipode, use Definition 2.9B. For any particular tree t, the only partition that contributes to the antipode is the complete partition into single vertices, because $a(\mathrm{t}') = 0$ when $|\mathrm{t}'| > 1$. Hence,

$$a^{-1}(\mathrm{t}) = \big(-a(\tau)\big)^{|\mathrm{t}|} = 1.$$

3.6 Elementary weights and the order of Runge–Kutta methods

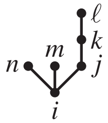
To express the B-series of a Runge–Kutta method we need to use “elementary weights”. Consider a generic method

$$\begin{array}{c|c} c & A \\ \hline & b^\intercal \end{array}. \tag{3.6 a}$$

Definition 3.6A The elementary weight $\Phi(\mathrm{t})$ and the stage weight $varphi_i(\mathrm{t})$ $(i = 1, 2, \ldots, s)$ of (3.6 a) corresponding to t are

$$\begin{aligned}
\varphi_i(\tau) &= c_i,\\
\Phi(\tau) &= b^{\mathsf{T}}\mathbf{1},\\
\varphi_i([\mathrm{t}_1\mathrm{t}_2\cdots\mathrm{t}_m]) &= \sum_{j=1}^{s} a_{ij}\prod_{k=1}^{m}\varphi_j(\mathrm{t}_k),,\\
\Phi([\mathrm{t}_1\mathrm{t}_2\cdots\mathrm{t}_m]) &= \sum_{j=1}^{s} b_j\prod_{k=1}^{m}\varphi_j(\mathrm{t}_k).
\end{aligned} \tag{3.6 b}$$

It is convenient to expand the recursions in 3.6A, as in the example

t = 

$$\Phi(\mathrm{t}) = \sum_{i,j,k,\ell,m,n=1}^{s} b_i a_{in} a_{im} a_{ij} a_{jk} a_{k\ell}.$$

Because the summations over the leaves of the tree can all be carried out explicitly, to give $\sum_n a_{in} = \sum_m a_{im} = c_i$, $\sum_\ell a_{k\ell} = c_k$, we can also write

$$\Phi(\mathrm{t}) = \sum_{i,j,k=1}^{s} b_i c_i^2 a_{ij} a_{jk} c_k.$$

It is also convenient to write the values of Φ in a combination of matrix-vector notation and component-by-component products. Thus we can write

$$\Phi(\mathrm{t}) = b^{\mathsf{T}}(c^2 A^2 c).$$

In this notation, pointwise powers of vectors have first priority, conformable vector and matrix multiplications have second priority and pointwise vector products have third priority. Parentheses are used where necessary to overrule these priorities.

The values of $\Phi(\mathrm{t})$, in this compact notation are given in Table 14 up to order 6.

In the following theorem, the meaning of elementary weights is extended for the sake of convenience by defining

$$\varphi_i(\varnothing) = \Phi(\varnothing) = 1.$$

Theorem 3.6B The B-series for stage i of (3.6 a) is $(\mathbf{B}_h\varphi_i)y_0$. The B-series for the output of (3.6 a) is $(\mathbf{B}_h\Phi)y_0$.

Proof. The stages Y_i and output y_1 are defined by

Table 14 Elementary weights to order 6

$\lvert t \rvert$	t		$\Phi(t)$
1	$\mathbf{t}_1$	τ	$b^\mathsf{T}\mathbf{1}$
2	$\mathbf{t}_2$	$[\tau]$	$b^\mathsf{T}c$
3	$\mathbf{t}_3$	$[\tau^2]$	$b^\mathsf{T}c^2$
3	$\mathbf{t}_4$	$[_2\tau]_2$	$b^\mathsf{T}Ac$
4	$\mathbf{t}_5$	$[\tau^3]$	$b^\mathsf{T}c^3$
4	$\mathbf{t}_6$	$[\tau[\tau]]$	$b^\mathsf{T}(cAc)$
4	$\mathbf{t}_7$	$[_2\tau^2]_2$	$b^\mathsf{T}Ac^2$
4	$\mathbf{t}_8$	$[_3\tau]_3$	$b^\mathsf{T}A^2c$
5	$\mathbf{t}_9$	$[\tau^4]$	$b^\mathsf{T}c^4$
5	$\mathbf{t}_{10}$	$[\tau^2[\tau]]$	$b^\mathsf{T}(c^2Ac)$
5	$\mathbf{t}_{11}$	$[\tau[\tau^2]]$	$b^\mathsf{T}(cAc^2)$
5	$\mathbf{t}_{12}$	$[\tau[_2\tau]_2]$	$b^\mathsf{T}(cA^2c)$
5	$\mathbf{t}_{13}$	$[[\tau]^2]$	$b^\mathsf{T}(Ac)^2$
5	$\mathbf{t}_{14}$	$[_2\tau^3]_2$	$b^\mathsf{T}Ac^3$
5	$\mathbf{t}_{15}$	$[_2\tau[\tau]]_2$	$b^\mathsf{T}A(cAc)$
5	$\mathbf{t}_{16}$	$[_3\tau^2]_3$	$b^\mathsf{T}A^2c^2$
5	$\mathbf{t}_{17}$	$[_4\tau]_4$	$b^\mathsf{T}A^3c$
6	$\mathbf{t}_{18}$	$[\tau^5]$	$b^\mathsf{T}c^5$
6	$\mathbf{t}_{19}$	$[\tau^3[\tau]]$	$b^\mathsf{T}(c^3Ac)$
6	$\mathbf{t}_{20}$	$[\tau^2[\tau^2]]$	$b^\mathsf{T}(c^2Ac^2)$
6	$\mathbf{t}_{21}$	$[\tau^2[_2\tau]_2]$	$b^\mathsf{T}(c^2A^2c)$
6	$\mathbf{t}_{22}$	$[\tau[\tau]^2]$	$b^\mathsf{T}(c(Ac)^2)$
6	$\mathbf{t}_{23}$	$[\tau[\tau^3]]$	$b^\mathsf{T}(cAc^3)$
6	$\mathbf{t}_{24}$	$[\tau[\tau[\tau]]]$	$b^\mathsf{T}(cA(cAc))$
6	$\mathbf{t}_{25}$	$[\tau[_2\tau^2]_2]$	$b^\mathsf{T}(cA^2c^2)$
6	$\mathbf{t}_{26}$	$[\tau[_3\tau]_3]$	$b^\mathsf{T}(cA^3c)$
6	$\mathbf{t}_{27}$	$[[\tau][\tau^2]]$	$b^\mathsf{T}(AcAc^2)$
6	$\mathbf{t}_{28}$	$[[\tau][_2\tau]_2]$	$b^\mathsf{T}(AcA^2c)$
6	$\mathbf{t}_{29}$	$[\tau[_2\tau^2]_2]$	$b^\mathsf{T}Ac^4$
6	$\mathbf{t}_{30}$	$[\tau[_3\tau]_3]$	$b^\mathsf{T}A(c^2Ac)$
6	$\mathbf{t}_{31}$	$[[\tau][\tau^2]]$	$b^\mathsf{T}A(cAc^2)$
6	$\mathbf{t}_{32}$	$[[\tau][_2\tau]_2]$	$b^\mathsf{T}A(cA^2c)$
6	$\mathbf{t}_{33}$	$[_2[\tau]^2]_2$	$b^\mathsf{T}A(Ac)^2$
6	$\mathbf{t}_{34}$	$[_3\tau^3]_3$	$b^\mathsf{T}A^2c^3$
6	$\mathbf{t}_{35}$	$[_3\tau[\tau]]_3$	$b^\mathsf{T}A^2(cAc)$
6	$\mathbf{t}_{36}$	$[_4\tau^2]_4$	$b^\mathsf{T}A^3c^2$
6	$\mathbf{t}_{37}$	$[_5\tau]_5$	$b^\mathsf{T}A^4c$

$$Y_i = y_0 + h\sum_{j=1}^{s} a_{ij} f(Y_j), \quad i = 1, 2, \ldots, s,$$

$$y_1 = y_0 + h\sum_{j=1}^{s} b_j f(Y_j).$$

Let $\eta_i \in \mathsf{B}$ and $\xi \in \mathsf{B}$ represent Y_i and y_1 resepctively, so that

$$\eta_i = 1 + \sum_{j=1}^{s} a_{ij}(\eta_j \mathrm{D}),$$

$$\xi = 1 + \sum_{j=1}^{s} b_j(\eta_j \mathrm{D}).$$

For $\mathrm{t} = [\mathrm{t}_1\mathrm{t}_2\cdots\mathrm{t}_m]$, $i = 1,2,\dots,s$, use $(\eta_i\mathrm{D})(\mathrm{t}) = \prod_{k=1}^{m} \eta_i(\mathrm{t}_k)$, so that

$$
\begin{aligned}
\eta_i(\tau) &= c_i,\\
\xi(\tau) &= \sum_{i=1}^{s} b_i = b^{\mathsf{T}}\mathbf{1},\\
\eta_i([\mathrm{t}_1\mathrm{t}_2\cdots\mathrm{t}_m]) &= \sum_{j=1}^{s} a_{ij}\prod_{k=1}^{m}\eta_j(\mathrm{t}_k),\\
\xi([\mathrm{t}_1\mathrm{t}_2\cdots\mathrm{t}_m]) &= \sum_{j=1}^{s} b_j\prod_{k=1}^{m}\eta_j(\mathrm{t}_k).
\end{aligned}
\tag{3.6 c}
$$

. By comparing (3.6 b) and (3.6 c), we see that, by induction, $\eta_i = \varphi_i$, $i = 1,2,\dots,s$, and $\xi = \Phi$. □

Theorem 3.6C For an initial value problem of arbitrary dimension, a Runge–Kutta method (A, b^{T}, c) has order p if and only if

$$\Phi(\mathrm{t}) = \frac{1}{\mathrm{t}!}$$

for all trees such that $|\mathrm{t}| \le p$.

3.7 Elementary differentials based on Kronecker products

Given a Runge–Kutta method $M = (A, b^{\mathsf{T}}, c)$, let

$$\boldsymbol{A} = \begin{bmatrix} A & \mathbf{0} \\ b^{\mathsf{T}} & 0 \end{bmatrix}, \tag{3.7 a}$$

and write $n = s + 1$. For a direct approach to finding the Taylor expansions of the stages and the output of M, define the problem $(\widetilde{y}_0, \widetilde{f}, h)$. where $\widetilde{f} : \mathbb{R}^{nN} \to \mathbb{R}^{nN}$ and $\widetilde{y}_0 \in \mathbb{R}^{nN}$ are defined by

$$\widetilde{f}(\widetilde{y}) = \widetilde{f}\left(\begin{bmatrix} \widetilde{y}^1 \\ \widetilde{y}^2 \\ \vdots \\ \widetilde{y}^n \end{bmatrix}\right) = (\boldsymbol{A} \otimes I)\left(\begin{bmatrix} f(\widetilde{y}^1) \\ f(\widetilde{y}^2) \\ \vdots \\ f(\widetilde{y}^n) \end{bmatrix}\right), \qquad \widetilde{y}_0 = \begin{bmatrix} y_0 \\ y_0 \\ \vdots \\ y_0 \end{bmatrix}$$

The results we want are embedded in the B-series for *implicit*$_h$ applied to $(\widetilde{y}_0, \widetilde{f}, h)$, That is the series for $\widetilde{y}_1$, given by

$$\widetilde{y}_1 = \widetilde{y}_0 + h\widetilde{f}(\widetilde{y}_1), \quad \widetilde{y}(x_0) = \widetilde{y}_0.$$

It will not be necessary to evaluate the partial derivatives, and hence the elementary differentials, of $\widetilde{f}$ at all points in $\mathbb{R}^{nN} \to \mathbb{R}^{nN}$, but only where

$$\widetilde{y}^1 = \widetilde{y}^2 = \cdots = \widetilde{y}^n = y_0. \tag{3.7 b}$$

Furthermore, it will only be necessary to evaluate $\widetilde{f}^{(m)}$ acting only on operands of the form

$$(\phi_1 \otimes \delta_1, \phi_2 \otimes \delta_2, \ldots, \phi_m \otimes \delta_m),$$

where

$$\phi_i \in \mathbb{R}^n, \quad \delta_i \in \mathbb{R}^N, \qquad i = 1, 2, \ldots, m.$$

The value of $\widetilde{f}^{(m)}$ is found to be

$$(\boldsymbol{A} \operatorname{dot}_m \phi_1\phi_2\cdots\phi_m) \otimes (\mathbf{f}^{(m)}\delta_1\delta_2\cdots\delta_m).$$

and this leads to

Theorem 3.7A The elementary differentials of $\widetilde{f}$, evaluated at $\widetilde{\mathbf{y}}_0$ are

$$\widetilde{\mathbf{F}}(\mathrm{t}) = \boldsymbol{\Phi}(\mathrm{t}) \otimes \mathbf{F}(\mathrm{t}),$$

where

$$\boldsymbol{\Phi}(\mathrm{t}) = \begin{cases} \boldsymbol{A}\mathbf{1}, & \mathrm{t} = \tau, \\ \boldsymbol{A} \operatorname{dot}_m \boldsymbol{\Phi}(\mathrm{t}_1)\boldsymbol{\Phi}(\mathrm{t}_2)\cdots\boldsymbol{\Phi}(\mathrm{t}_m), & \mathrm{t} = [\mathrm{t}_1\mathrm{t}_2\cdots\mathrm{t}_m]. \end{cases} \tag{3.7 c}$$

Proof. For $\mathrm{t} = \tau$, we have

$$\widetilde{\mathbf{F}}(\mathrm{t}) = (\boldsymbol{A} \otimes I)(\mathbf{1} \otimes \mathbf{f}) = \boldsymbol{A}\mathbf{1} \otimes \mathbf{f} = \boldsymbol{\Phi}(\tau) \otimes \mathbf{F}(\tau).$$

For $\mathrm{t} = [\mathrm{t}_1\mathrm{t}_2\cdots\mathrm{t}_m]$,

$$\widetilde{\mathbf{F}}(\mathrm{t}) = (\boldsymbol{A} \otimes I)(\mathbf{1} \otimes \mathbf{f}^{(m)})\big(\boldsymbol{\Phi}(\mathrm{t}_1) \otimes \mathbf{F}(\mathrm{t}_1), \boldsymbol{\Phi}(\mathrm{t}_2) \otimes \mathbf{F}(\mathrm{t}_2), \ldots, \boldsymbol{\Phi}(\mathrm{t}_m) \otimes \mathbf{F}(\mathrm{t}_m)\big),$$

with component number i equal to

$$\begin{aligned} \sum_{j=1}^{n} \boldsymbol{a}_{ij} \prod_{k=1}^{m} \boldsymbol{\Phi}(\mathrm{t}_j)_k \mathbf{f}^{(m)} \mathbf{F}(\mathrm{t}_1)\mathbf{F}(\mathrm{t}_2)\cdots\mathbf{F}(\mathrm{t}_m) &= e_i^{\mathsf{T}} \boldsymbol{A} \operatorname{dot}_m \boldsymbol{\Phi}(\mathrm{t}_1)\boldsymbol{\Phi}(\mathrm{t}_2)\cdots\boldsymbol{\Phi}(\mathrm{t}_m)\mathbf{F}(\mathrm{t}) \\ &= (e_i^{\mathsf{T}}\boldsymbol{\Phi}(\mathrm{t}))\mathbf{F}(\mathrm{t}). \end{aligned}$$ □

We are now in a position to give an alternative proof for Theorem 3.6B (p. 125).

Write $\boldsymbol{\Phi}_i(\mathrm{t}) = \varphi_i(\mathrm{t})$, $i = 1, 2, \ldots, s$, $\boldsymbol{\Phi}_{s+1}(\mathrm{t}) = \Phi_i(\mathrm{t})$, so that (3.7 c) becomes

$$\varphi(\mathrm{t}) = \begin{cases} A\mathbf{1}, & \mathrm{t} = \tau, \\ A\,\mathrm{dot}_m\, \varphi(\mathrm{t}_1)\varphi(\mathrm{t}_2)\cdots\varphi(\mathrm{t}_m), & \mathrm{t} = [\mathrm{t}_1\mathrm{t}_2\cdots\mathrm{t}_m]. \end{cases}$$

$$\Phi(\mathrm{t}) = \begin{cases} b^{\mathsf{T}}\mathbf{1}, & \mathrm{t} = \tau, \\ b^{\mathsf{T}}\,\mathrm{dot}_m\, \varphi(\mathrm{t}_1)\varphi(\mathrm{t}_2)\cdots\varphi(\mathrm{t}_m), & \mathrm{t} = [\mathrm{t}_1\mathrm{t}_2\cdots\mathrm{t}_m]. \end{cases}$$

These give the values of $\varphi(\mathrm{t})$ and $\Phi(\mathrm{t})$ in Definition 3.6A (p. 125) and the result follows.

3.8 Attainable values of elementary weights and differentials

Elementary weights

Given a finite set of trees, we first consider the question "What values can the elementary weights take on over this set of trees?" It is assumed that there is no restriction on the number of stages of the Runge–Kutta methods which generate the values of these elementary weights.

The answer to this question is that any values of the elementary weights can arise. An important consequence is that there is no bound on the attainable order of Runge–Kutta methods if there is no restriction on the number of stages.

As a preliminary step we introduce Lemma 3.8A on a class of functions on a finite set. This is a trivial case of the Stone–Weierstrass Theorem [81] (Rudin,1976).

Lemma 3.8A Let I be a finite set and let F denote the set of real-valued functions on I such that

1. F is a vector space.
2. F contains the unity function $\widehat{1}(i) = 1$.
3. F is closed under pointwise multiplication.
4. F distinguishes points.

Then F contains every function $I \to \mathbb{R}$.

Proof. Let $I = \{1, 2, \ldots, n\}$ and let f_i, $i = 1, 2, \ldots, n$, be a given vector. A function $f \in F$ will be constructed such that $f(i) = f_i$. For i, j distinct points in I, let $\varphi_{ij} \in F$ be chosen such that $\varphi_{ij}(x_i) \neq \varphi_{ij}(x_j)$. Define f by

$$f = \sum_{i \in I} f_i \prod_{j \neq i} \frac{\varphi_{ij} - \varphi_{ij}(x_j)\widehat{1}}{\varphi_{ij}(x_i) - \varphi_{ij}(x_j)},$$

where the product is pointwise. □

Now use this result to consider the possible values of the elementary differentials generated by some Runge–Kutta method. Let $\widehat{\mathrm{T}}$ denote a finite subset of the rooted trees and define F to be the set of vectors of values that can be taken by the elementary weights evaluated for $\mathrm{t} \in \widehat{\mathrm{T}}$.

Theorem 3.8B For any distinct set of trees $\{\mathrm{t}_1, \mathrm{t}_2, \ldots, \mathrm{t}_N\}$, and a corresponding set of real numbers, $\{\theta_1, \theta_2, \ldots, \theta_N\}$, there exists a Runge–Kutta method (A, b^{T}, c) such that

$$\Phi(\mathrm{t}_i) = \theta_i, \qquad i = 1, 2, \ldots, N.$$

Proof. Assume $|\mathrm{t}_i| \leq n$, $i = 1, 2, \ldots, N$, for some positive integer n. For $n = 1$, the result is clear. Proceed by induction for $n > 1$, and assume the result holds when n is replaced by a lower positive integer.

Let F denote the set of possible values of $\{\Phi(\mathrm{t}_1), \Phi(\mathrm{t}_2), \ldots, \Phi(\mathrm{t}_N)\}$. We will prove that a Runge–Kutta method exists satisfying the assumptions in Lemma 3.8A, where F denotes the set of possible values of $\{\Phi(\mathrm{t}_1), \Phi(\mathrm{t}_2), \ldots, \Phi(\mathrm{t}_N)\}$, over all Runge–Kutta methods. We will show that F satisfies the conditions of Lemma 3.8A.

1. F is a vector space
   Given tableaux

$$\begin{array}{c|c} \widetilde{c} & \widetilde{A} \\ \hline & \widetilde{b}^{\mathsf{T}} \end{array}, \qquad \begin{array}{c|c} \widehat{c} & \widehat{A} \\ \hline & \widehat{b}^{\mathsf{T}} \end{array} \tag{3.8 a}$$

   with corresponding elementary weight vectors

$$\widetilde{\boldsymbol{\Phi}} = \begin{bmatrix} \widetilde{\varphi} \\ \widetilde{\Phi} \end{bmatrix}, \qquad \widehat{\boldsymbol{\Phi}} = \begin{bmatrix} \widehat{\varphi} \\ \widehat{\Phi} \end{bmatrix},$$

   then the linear combination $\widetilde{C}\widetilde{\Phi} + \widehat{C}\widehat{\Phi}$ is generated by

$$\begin{array}{c|cc} \widetilde{c} & \widetilde{A} & 0 \\ \widehat{c} & 0 & \widehat{A} \\ \hline & \widetilde{C}\widetilde{b}^{\mathsf{T}} & \widehat{C}\widehat{b}^{\mathsf{T}} \end{array}.$$

2. F contains the unity function $\widehat{1}(i) = 1$.
   The unity function is generated by

$$\begin{array}{c|c} 1 & 1 \\ \hline & 1 \end{array}.$$

3. F is closed under pointwise multiplication.
   Let $\widetilde{I} = \{1, 2, \ldots, \widetilde{s}\}$, $\widehat{I} = \{1, 2, \ldots, \widehat{s}\}$ be the index sets for the tableaux in (3.8 a). Consider the tableau formed by Kronecker multiplication on the index set $\widetilde{I} \otimes \widehat{I}$,

$$\begin{array}{c|c} \widetilde{c} \otimes \widehat{c} & \widetilde{A} \otimes \widehat{A} \\ \hline & \widetilde{b}^{\mathsf{T}} \otimes \widehat{b}^{\mathsf{T}} \end{array}.$$

For this tableau,

$$
\begin{aligned}
\varphi(\tau) &= (\widetilde{A}\otimes\widehat{A})(\widetilde{\mathbf{1}}\otimes\widehat{\mathbf{1}})\\
&= \widetilde{\varphi}(\tau)\otimes\widehat{\varphi}(\tau),\\
\Phi(\tau) &= (\widetilde{b}^{\mathsf{T}}\otimes\widehat{b}^{\mathsf{T}})(\widetilde{\mathbf{1}}\otimes\widehat{\mathbf{1}})\\
&= \widetilde{\Phi}(\tau)\widehat{\Phi}(\tau),\\
\varphi(\mathrm{t}) &= (\widetilde{A}\otimes\widehat{A})\big((\mathrm{dot}_m\,\widetilde{\varphi}(\mathrm{t}_1)\widetilde{\varphi}(\mathrm{t}_2)\cdots\widetilde{\varphi}(\mathrm{t}_m)\otimes(\mathrm{dot}_m\,\widehat{\varphi}(\mathrm{t}_1)\widehat{\varphi}(\mathrm{t}_2)\cdots\widehat{\varphi}(\mathrm{t}_m)\big)\\
&= \widetilde{\varphi}(\mathrm{t})\otimes\widehat{\varphi}(\mathrm{t}),\\
\Phi(\mathrm{t}) &= (\widetilde{b}^{\mathsf{T}}\otimes\widehat{b}^{\mathsf{T}})\big((\mathrm{dot}_m\,\widetilde{\varphi}(\mathrm{t}_1)\widetilde{\varphi}(\mathrm{t}_2)\cdots\widetilde{\varphi}(\mathrm{t}_m)\otimes(\mathrm{dot}_m\,\widehat{\varphi}(\mathrm{t}_1)\widehat{\varphi}(\mathrm{t}_2)\cdots\widehat{\varphi}(\mathrm{t}_m)\big)\\
&= \widetilde{\Phi}(\mathrm{t})\widehat{\Phi}(\mathrm{t}),
\end{aligned}
$$

where $\mathrm{t}=[\mathrm{t}_1\mathrm{t}_2\cdots\mathrm{t}_m]$. Hence, for all $\mathrm{t}\in T$, $\Phi(\mathrm{t})=\widetilde{\Phi}(\mathrm{t})\otimes\widehat{\Phi}(\mathrm{t})=\widetilde{\Phi}(\mathrm{t})\widehat{\Phi}(\mathrm{t})$.

4. F distinguishes points
   In this part of the proof, it is sufficient to assume $N=2$. Let $\mathrm{t}_1=[\mathrm{t}_{11}^{k_1}\cdots]$, $\mathrm{t}_2=[\mathrm{t}_{21}{}^{m_1}\cdots]$, assumed to be distinct. For a given method (A,b^{T},c), construct the tableau

$$
\begin{array}{c|cc}
c & A & 0\\
b^{\mathsf{T}}\mathbf{1} & b^{\mathsf{T}} & 0\\
\hline
 & 0^{\mathsf{T}} & 1
\end{array}\,,
$$

   so that F contains an arbitrary linear combination of terms like

$$
\begin{bmatrix} \Phi(\mathrm{t}_{11})^{k_1}\cdots & \Phi(\mathrm{t}_{21})^{m_1}\cdots \end{bmatrix}.
$$

   By the induction hypothesis, $\Phi(\mathrm{t}_{11}),\ldots,\Phi(\mathrm{t}_{21}),\ldots$ can have any values. □

Elementary differentials

A similar result on the attainable values of elementary differentials is given in Theorem 3.8D. This is preceded by an example of the values of elementary differentials in a special case.

Preliminary lemma

Lemma 3.8C Let

$$f(y)=A\exp\big(\mathrm{diag}(y)\big)\mathbf{1},$$

then

$$f^{(n)}(y)=A\exp\big(\mathrm{diag}(y)\big).\mathrm{dot}_n\,. \tag{3.8 b}$$

Proof. For small $\|\delta\|$,

$$\begin{aligned} f(y+\delta) &= A\exp\big(\operatorname{diag}(y)\mathbf{1}+\operatorname{diag}(\delta)\mathbf{1}\big)\\ &= A\exp\big(\operatorname{diag}(y)\big)\mathbf{1}+A\exp\big(\operatorname{diag}(y)\big)\operatorname{diag}(\delta)\mathbf{1}+\mathrm{o}(\|\delta\|)\\ &= f(y)+A\operatorname{diag}(\delta)\exp\big(\operatorname{diag}(y)\big)\mathbf{1}+\mathrm{o}(\|\delta\|). \end{aligned}$$

Hence $f'(y)\delta = A\operatorname{diag}(\delta)\exp\big(\operatorname{diag}(y)\big)\mathbf{1}$. By applying this result n times, we have

$$\begin{aligned} f^{(n)}(y)\delta_1\delta_2\cdots\delta_n &= A\operatorname{diag}(\delta_1)\operatorname{diag}(\delta_2)\cdots\operatorname{diag}(\delta_n)\exp\big(\operatorname{diag}(y)\big)\mathbf{1}\\ &= A\exp\big(\operatorname{diag}(y)\big)\operatorname{diag}(\mathrm{dot}_n\,\delta_1\delta_2\cdots\delta_n)\mathbf{1}\\ &= A\exp\big(\operatorname{diag}(y)\big)\,\mathrm{dot}_n\,\delta_1\delta_2\cdots\delta_n, \end{aligned}$$

implying (3.8 b). □

Main result

Theorem 3.8D For any positive integer n, let T_n denote the set $\{\mathrm{t}_1,\mathrm{t}_2,\dots,\mathrm{t}_N\}$ of all trees satisfying $|\mathrm{t}_i|\le n$, $i=1,2,\dots,N$. Then for any real sequence θ_i, $i=1,2,\dots,N$, there exists a finite-dimensional vector space $X=\mathbb{R}^N$, $f: X\to X$ and a non-zero vector $x\in X$ such that for any $\mathrm{t}\in\mathrm{T}_n$,

$$x^{\mathsf{T}}\mathbf{F}(\mathrm{t}_i)=\theta_i.$$

Proof. Let (A,b^{T},c) denote a Runge–Kutta method satisfying the requirements of Theorem 3.8B. Let $N=s$, and define $\boldsymbol{A}$ from (3.7 a), $f(y)=\boldsymbol{A}\exp(\operatorname{diag}(y))\mathbf{1}$ and $y_0=0$ with $x^{\mathsf{T}}=e_{s+1}^{\mathsf{T}}$. From Lemma 3.8C, $f^{(n)}(y)=\boldsymbol{A}\exp(\operatorname{diag}(y)).\mathrm{dot}_n$, and

$$\mathbf{f}=\boldsymbol{A}\mathbf{1}=\boldsymbol{\Phi},\qquad \mathbf{f}^{(n)}=\boldsymbol{A}\,\mathrm{dot}_n\,.$$

Use B^+ induction Chapter 2, Section 2.2 (p. 45) to show that $\mathbf{F}(\mathrm{t})=\boldsymbol{\Phi}(\mathrm{t})$, where $\mathrm{t}=[\mathrm{t}_1\mathrm{t}_2\cdots\mathrm{t}_n]$ starting from $t=\tau$. We have

$$\begin{aligned} x^{\mathsf{T}}\mathbf{F}(\mathrm{t}) &= x^{\mathsf{T}}\mathbf{f}^{(n)}\mathbf{F}(\mathrm{t}_1)\mathbf{F}(\mathrm{t}_2)\cdots\mathbf{F}(\mathrm{t}_n)\\ &= x^{\mathsf{T}}\boldsymbol{A}\exp(y_0I)\,\mathrm{dot}_n\,\boldsymbol{\Phi}(\mathrm{t}_1)\boldsymbol{\Phi}(\mathrm{t}_2)\cdots\boldsymbol{\Phi}(\mathrm{t}_n)\\ &= x^{\mathsf{T}}\boldsymbol{A}\,\mathrm{dot}_n\,\boldsymbol{\Phi}(\mathrm{t}_1)\boldsymbol{\Phi}(\mathrm{t}_2)\cdots\boldsymbol{\Phi}(\mathrm{t}_n)\\ &= \Phi(t), \end{aligned}$$

and the result follows from Theorem 3.8B. □

3.9 Composition of B-series

Introducing the composition theorem

The composition of Runge–Kutta methods, or more generally as integration methods, interpreted as the composition of their B-series, was considered in [14] (Butcher, 1972). Because these B-series are dense in the sense of sequence density, this implies the composition formula for B-series in general. A direct proof of the composition theorem was provided in [52] (Hairer, Wanner, 1974).

The principal result in the current section, Theorem 3.9C (p. 139), is an alternative approach to the composition of B-series.

Compositions of mappings and B-series

Consider two mappings $\mathfrak{a}_h$, $\mathfrak{b}_h$, where $\mathfrak{a}_h$ is a central mapping; that is $y_0 \mapsto y_0 + \mathscr{O}(h)$. Several maps constructed from $\mathfrak{a}_h$ and $\mathfrak{b}_h$ are shown in the diagram (3.9 a). The map $y_3 = \mathfrak{b}_h y_1$ is the composition $\mathfrak{b}_h \circ \mathfrak{a}_h$ as in the diagram and represents the substitution of $y_1 = \mathfrak{a}_h y_0$ in place of y_0 in $y_2 = \mathfrak{b}_h y_0$

$$\begin{tikzcd}
y_1 \arrow[r, "\mathfrak{b}_h"] & y_3 \\
y_0 \arrow[u, "\mathfrak{a}_h"] \arrow[ur, dashed, "\mathfrak{b}_h \circ \mathfrak{a}_h"] \arrow[r, "\mathfrak{b}_h"] & y_2
\end{tikzcd} \tag{3.9 a}$$

The principal idea of B-series analysis is to represent $\mathfrak{a}_h$ by $\mathbf{B}_h a$ and $\mathfrak{b}_h$ by $\mathbf{B}_h b$ and to represent the composition by $\mathbf{B}_h ab$, where $ab \in \mathsf{B}^*$ can be written in terms of a and b. In the special case that $b \in \mathsf{B}$, or $b \in \mathsf{B}^0$, then ab is also a member of B or B^0, respectively. In diagram (3.9 b), the B-series counterpart of (3.9 a) is shown

$$\begin{tikzcd}
a \arrow[r, "b"] & ab \\
1 \arrow[u, "a"] \arrow[ur, dashed, "ab"] \arrow[r, "b"] & b
\end{tikzcd} \tag{3.9 b}$$

A common application of compositions is in the alteration of the base-point in a particular Taylor expansion by a substitution $y_0 \mapsto y_1 = (\mathbf{B}_h y_0)a$ with the result written in terms of a Taylor series about y_0. The relationship between the various mappings is shown on the left of (3.9 c) with the corresponding B-series representation shown on the right.

$$
\begin{tikzcd}
y_1 \arrow[r, "\mathfrak{b}_h"] & y_3 \\
y_0 \arrow[u, "\mathfrak{a}_h"] \arrow[r, "\mathfrak{b}_h"] \arrow[ur, dashed, "\mathfrak{b}_h \circ \mathfrak{a}_h"] & y_2 \arrow[u, "\ell(a)_h"']
\end{tikzcd}
\qquad
\begin{tikzcd}
a \arrow[r, "b"] & ab \\
1 \arrow[u, "a"] \arrow[r, "b"] \arrow[ur, dashed, "ab"] & b \arrow[u, "L(a)"']
\end{tikzcd}
\tag{3.9c}
$$

The mapping $\ell(a)_h$ in the left diagram is necessarily linear because if $\mathfrak{b}_h$ and $\mathfrak{c}_h$ are mappings, and θ scalar, then

$$
\begin{aligned}
\big((\mathfrak{b}_h + \mathfrak{c}_h) \circ \mathfrak{a}_h\big)(y_0) &= (\mathfrak{b}_h \circ \mathfrak{a}_h)(y_0) + (\mathfrak{c}_h \circ \mathfrak{a}_h)(y_0), \\
\big((\theta\,\mathfrak{b}_h) \circ \mathfrak{a}_h\big)(y_0) &= \theta\big(\mathfrak{b}_h \circ \mathfrak{a}_h\big)(y_0).
\end{aligned}
$$

The linear operator in the left diagram $L(a) : \mathsf{B}^* \to \mathsf{B}^*$ is defined by

$$L(a)b = ab,$$

where ab is defined by the extended group operation $(a,b) \in \mathsf{B} \times \mathsf{B}^* \mapsto ab \in \mathsf{B}^*$.

The groups of central mappings and central B-series

If $\mathfrak{a}_h, \mathfrak{b}_h, \mathfrak{c}_h$ are central mappings, represented by $a, b, c \in \mathsf{B}$ respectively, then we have a rich structure partly represented by the diagram

$$
\begin{tikzcd}[column sep=large]
y_0 \arrow[r, bend left, "\mathfrak{a}_h"] \arrow[rr, bend left=40, "\mathfrak{b}_h \circ \mathfrak{a}_h"] & y_1 \arrow[l, bend left, "\mathfrak{a}_h^{-1}"] \arrow[r, bend left, "\mathfrak{b}_h"] \arrow[rr, bend left=40, "\mathfrak{c}_h \circ \mathfrak{b}_h"] & y_2 \arrow[l, bend left, "\mathfrak{b}_h^{-1}"] \arrow[r, bend left, "\mathfrak{c}_h"] \arrow[ll, bend left=40, "\mathfrak{a}_h^{-1} \circ \mathfrak{b}_h^{-1}"] & y_3 \arrow[l, bend left, "\mathfrak{c}_h^{-1}"] \arrow[ll, bend left=40, "\mathfrak{b}_h^{-1} \circ \mathfrak{c}_h^{-1}"]
\end{tikzcd}
$$

The diagram representing the corresponding members of B is

$$
\begin{tikzcd}[column sep=large]
1 \arrow[r, bend left, "a"] \arrow[rr, bend left=40, "ab"] & a \arrow[l, bend left, "a^{-1}"] \arrow[r, bend left, "b"] \arrow[rr, bend left=40, "bc"] & ab \arrow[l, bend left, "b^{-1}"] \arrow[r, bend left, "c"] \arrow[ll, bend left=40, "b^{-1}a^{-1}"] & abc \arrow[l, bend left, "c^{-1}"] \arrow[ll, bend left=40, "c^{-1}b^{-1}"]
\end{tikzcd}
$$

This section is devoted to understanding the nature of the composition $(a,b) \to ab$ as a representation of $\mathfrak{b}_h \circ \mathfrak{a}_h$ and using this for practical analysis of numerical questions.

Analysis of substitutions

Our principal aim in this section is to show how $(ab)(\mathrm{t})$ can be written in terms of the mappings $a : \mathrm{T} \to \mathbb{R}$ and $b : \mathrm{T} \to \mathbb{R}$, where

$$(\mathbf{B}_h y_0)(ab) = \big(\mathbf{B}_h\big((\mathbf{B}_h y_0)a\big)\big)b.$$

A convenient first step is to write the terms in $\mathbf{B}_h\big((\mathbf{B}_h y_0)a\big)$ as linear combinations of the terms in $\mathbf{B}_h$ and a second step is to evaluate the coefficients of $\sigma(\mathrm{t})^{-1}h^{|\mathrm{t}|}\mathbf{F}(\mathrm{t})$, $(\mathrm{t}\in T^{\#})$ in $\mathbf{B}_h\big((\mathbf{B}_h y_0)a\big)b$.

Introductory analysis to order 4

We will illustrate the nature of composition by doing the detailed calculation up to order 4 terms. Starting with $\mathbf{B}_h(b)(y_0)$, we will then substitute $\mathbf{B}_h(a)(y_0)$ for y_0 and rearrange to find $\mathbf{B}_h(ab)(y_0)$. Note that the usual convention of omitting y_0 in $\mathbf{f}y_0, \mathbf{f}'y_0, \mathbf{f}''y_0, \dots$, will be followed.

Begin with

$$\begin{aligned}\mathbf{B}_h(b)y_0 = b_0y_0 + hb_1\mathbf{f} + h^2b_2\mathbf{f}'\mathbf{f} + h^3(\tfrac{1}{2}b_3\mathbf{f}''\mathbf{f}\mathbf{f} + b_4\mathbf{f}'\mathbf{f}'\mathbf{f})\\ + h^4(\tfrac{1}{6}b_5\mathbf{f}'''\mathbf{f}\mathbf{f}\mathbf{f} + b_6\mathbf{f}''\mathbf{f}\mathbf{f}'\mathbf{f} + \tfrac{1}{2}b_7\mathbf{f}'\mathbf{f}''\mathbf{f}\mathbf{f} + b_8\mathbf{f}'\mathbf{f}'\mathbf{f}'\mathbf{f}) + \mathscr{O}(h^5),\end{aligned} \tag{3.9 d}$$

and transform this into $\mathbf{B}_h(b)\big(\mathbf{B}_h(a)y_0\big)$ by replacing y_0 by $\mathbf{B}_h(a)y_0$ and every occurrence of $\mathbf{f}^{(n)} = \mathbf{f}^{(n)}(y_0)$ by $\mathbf{f}^{(n)}(\mathbf{B}_h(a)y_0)$, for $n = 1,2,3$. These transformations, taken to as many terms as required for our purpose, and with appropriate $\mathscr{O}(h^m)$ terms omitted for convenience, are

$$\begin{aligned}y_0 &\mapsto y_0 + ha_1\mathbf{f} + h^2a_2\mathbf{f}'\mathbf{f} + h^3\big(\tfrac{1}{2}a_3\mathbf{f}''\mathbf{f}\mathbf{f} + a_4\mathbf{f}'\mathbf{f}'\mathbf{f}\big)\\ &\qquad + h^4\big(\tfrac{1}{6}a_5\mathbf{f}'''\mathbf{f}\mathbf{f}\mathbf{f} + a_6\mathbf{f}''\mathbf{f}\mathbf{f}'\mathbf{f} + \tfrac{1}{2}a_7\mathbf{f}'\mathbf{f}''\mathbf{f}\mathbf{f} + a_8\mathbf{f}'\mathbf{f}'\mathbf{f}'\mathbf{f}\big),\\ h\mathbf{f} &\mapsto h\mathbf{f} + h^2a_1\mathbf{f}'\mathbf{f} + h^3\big(\tfrac{1}{2}a_1^2\mathbf{f}''\mathbf{f}\mathbf{f} + a_2\mathbf{f}'\mathbf{f}'\mathbf{f}\big)\\ &\qquad + h^4\big(\tfrac{1}{6}a_1^3\mathbf{f}'''\mathbf{f}\mathbf{f}\mathbf{f} + a_1a_2\mathbf{f}''\mathbf{f}\mathbf{f}'\mathbf{f} + \tfrac{1}{2}a_3\mathbf{f}'\mathbf{f}''\mathbf{f}\mathbf{f} + a_4\mathbf{f}'\mathbf{f}'\mathbf{f}'\mathbf{f}\big),\\ h\mathbf{f}' &\mapsto h\mathbf{f}' + h^2a_1\mathbf{f}''\mathbf{f} + h^3\big(\tfrac{1}{2}a_1^2\mathbf{f}'''\mathbf{f}\mathbf{f} + a_2\mathbf{f}''\mathbf{f}'\mathbf{f}\big),\\ h\mathbf{f}'' &\mapsto h\mathbf{f}'' + h^2a_1\mathbf{f}'''\mathbf{f},\\ h\mathbf{f}''' &\mapsto h\mathbf{f}'''.\end{aligned} \tag{3.9 e}$$

Evaluate the transformed coefficients of $b_0,\dots,b_8$ in (3.9 d), in turn, using (3.9 e). The transformations of y_0 and $h\mathbf{f}$ have already been given. The coefficient of b_2 becomes

$$\begin{aligned}&h^2\mathbf{f}'\mathbf{f} \mapsto\\ &\big(h\mathbf{f}' + h^2a_1\mathbf{f}''\mathbf{f} + \tfrac{1}{2}h^3a_1^2\mathbf{f}'''\mathbf{f}\mathbf{f} + h^3a_2\mathbf{f}''\mathbf{f}'\mathbf{f}\big)\big(h\mathbf{f} + h^2a_1\mathbf{f}'\mathbf{f} + \tfrac{1}{2}h^3a_1^2\mathbf{f}''\mathbf{f}\mathbf{f} + h^3a_2\mathbf{f}'\mathbf{f}'\mathbf{f}\big)\\ &= h^2\mathbf{f}'\mathbf{f} + h^3a_1\mathbf{f}''\mathbf{f}\mathbf{f} + \tfrac{1}{2}h^3a_1\mathbf{f}'\mathbf{f}'\mathbf{f}\\ &\quad + \tfrac{1}{2}h^4a_1^3\mathbf{f}'''\mathbf{f}\mathbf{f}\mathbf{f} + h^4(a_1^2 + a_2)\mathbf{f}''\mathbf{f}\mathbf{f}'\mathbf{f} + \tfrac{1}{2}h^4a_1^2\mathbf{f}'\mathbf{f}''\mathbf{f}\mathbf{f} + h^4a_2\mathbf{f}'\mathbf{f}'\mathbf{f}'\mathbf{f}.\end{aligned}$$

In a similar way we find the transformations

$$\begin{aligned}\tfrac{1}{2}h^3\mathbf{f}''\mathbf{f}\mathbf{f} &\mapsto \tfrac{1}{2}h^3\mathbf{f}''\mathbf{f}\mathbf{f} + \tfrac{1}{2}h^4a_1\mathbf{f}'''\mathbf{f}\mathbf{f}\mathbf{f} + h^4a_1\mathbf{f}''\mathbf{f}\mathbf{f}'\mathbf{f},\\ h^3\mathbf{f}'\mathbf{f}'\mathbf{f} &\mapsto h^3\mathbf{f}'\mathbf{f}'\mathbf{f} + h^4a_1\mathbf{f}''\mathbf{f}\mathbf{f}'\mathbf{f} + h^4a_1\mathbf{f}'\mathbf{f}'\mathbf{f}'\mathbf{f},\end{aligned}$$

$$\begin{aligned}
\tfrac{1}{6}h^4\mathbf{f}'''\mathbf{fff} &\mapsto \tfrac{1}{6}h^4\mathbf{f}'''\mathbf{fff},\\
h^4\mathbf{f}''\mathbf{ff}'\mathbf{f} &\mapsto h^4\mathbf{f}''\mathbf{ff}'\mathbf{f},\\
\tfrac{1}{2}h^4\mathbf{f}'\mathbf{f}''\mathbf{ff} &\mapsto \tfrac{1}{2}h^4\mathbf{f}'\mathbf{f}''\mathbf{ff},\\
h^4\mathbf{f}'\mathbf{f}'\mathbf{f}'\mathbf{f} &\mapsto h^4\mathbf{f}'\mathbf{f}'\mathbf{f}'\mathbf{f}.
\end{aligned}$$

Write $y_1 = (\mathbf{B}_h y_0)a$.

The second step, of finding the coefficients of $\sigma(\mathrm{t})^{-1}h^{|\mathrm{t}|}\mathbf{F}(\mathrm{t})$ in

$$b_0 y_1 + \sum_{i=1}^{8} b_i h^{|\mathbf{t}_i|}\mathbf{F}(\mathbf{t}_i)y_1,$$

gives the results

$$\begin{aligned}
(ab)_0 &:= (ab)(\varnothing) = b_0,\\
(ab)_1 &:= (ab)(\mathbf{t}_1) = b_0a_1 + b_1,\\
(ab)_2 &:= (ab)(\mathbf{t}_2) = b_0a_2 + b_1a_1 + b_2,\\
(ab)_3 &:= (ab)(\mathbf{t}_3) = b_0a_3 + b_1a_1^2 + 2b_2a_1 + b_3,\\
(ab)_4 &:= (ab)(\mathbf{t}_4) = b_0a_4 + b_1a_2 + b_2a_1 + b_4,\\
(ab)_5 &:= (ab)(\mathbf{t}_5) = b_0a_5 + b_1a_1^3 + 3b_2a_1^2 + 3b_3a_1 + b_5,\\
(ab)_6 &:= (ab)(\mathbf{t}_6) = b_0a_6 + b_1a_1a_2 + b_2(a_1^2 + a_2) + b_3a_1 + b_4a_1 + b_6,\\
(ab)_7 &:= (ab)(\mathbf{t}_7) = b_0a_7 + b_1a_3 + b_2a_1^2 + 2b_4a_1 + b_7,\\
(ab)_8 &:= (ab)(\mathbf{t}_8) = b_0a_8 + b_1a_4 + b_2a_2 + b_4a_1 + b_8,
\end{aligned}$$

which, written in matrix form, become

$$\begin{bmatrix} (ab)_0\\ (ab)_1\\ (ab)_2\\ (ab)_3\\ (ab)_4\\ (ab)_5\\ (ab)_6\\ (ab)_7\\ (ab)_8 \end{bmatrix} = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
a_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
a_2 & a_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0\\
a_3 & a_1^2 & 2a_1 & 1 & 0 & 0 & 0 & 0 & 0\\
a_4 & a_2 & a_1 & 0 & 1 & 0 & 0 & 0 & 0\\
a_5 & a_1^3 & 3a_1^2 & 3a_1 & 0 & 1 & 0 & 0 & 0\\
a_6 & a_1a_2 & a_1^2+a_2 & a_1 & a_1 & 0 & 1 & 0 & 0\\
a_7 & a_3 & a_1^2 & 0 & 2a_1 & 0 & 0 & 1 & 0\\
a_8 & a_4 & a_2 & 0 & a_1 & 0 & 0 & 0 & 1
\end{bmatrix} \begin{bmatrix} b_0\\ b_1\\ b_2\\ b_3\\ b_4\\ b_5\\ b_6\\ b_7\\ b_8 \end{bmatrix}. \tag{3.9 f}$$

The elements of the matrix appearing in (3.9 f), written with each a_i replaced by $\mathbf{t}_i$, are

	$\emptyset$	τ	$[\tau]$	$[\tau^2]$	$[[\tau]]$	$[\tau^3]$	$[\tau[\tau]]$	$[[\tau^2]]$	$[[[\tau]]]$
$\emptyset$	1	0	0	0	0	0	0	0	0
τ	τ	1	0	0	0	0	0	0	0
$[\tau]$	$[\tau]$	τ	1	0	0	0	0	0	0
$[\tau^2]$	$[\tau^2]$	τ^2	2τ	1	0	0	0	0	0
$[[\tau]]$	$[[\tau]]$	$[\tau]$	τ	0	1	0	0	0	0
$[\tau^3]$	$[\tau^3]$	τ^3	$3\tau^2$	3τ	0	1	0	0	0
$[\tau[\tau]]$	$[\tau[\tau]]$	$\tau\,[\tau]$	$\tau^2+[\tau]$	τ	τ	0	1	0	0
$[[\tau^2]]$	$[[\tau^2]]$	$[\tau^2]$	τ^2	0	2τ	0	0	1	0
$[[[\tau]]]$	$[[[\tau]]]$	$[[\tau]]$	$[\tau]$	0	τ	0	0	0	1

It will be observed that, up to trees of order 4, the elements of this matrix agree with those of $\boldsymbol{\Lambda}$, introduced in Chapter 2, Section 2.9 (p. 95). Our aim will be to prove this in general and thus establish Theorem 3.9B and Corollary 3.9F.

Perturbed elementary differentials

For $a \in \mathbf{B}$, we will obtain a formula for

$$h^{|t'|}\mathbf{F}(t')\big((\mathbf{B}_h a)y_0\big) = h^{|t'|}\mathbf{F}(t')\big(y_1\big), \quad \text{where} \quad y_1 = y_0 + \sum_{t\in T} h^{|t|}a(t)\mathbf{F}(t),$$

written in the form of a B-series.

Lemma 3.9A

$$h\mathbf{F}(\tau)\big((\mathbf{B}_h a)y_0\big) = (\mathbf{B}_h b)y_0$$

where b is defined by

$$\begin{aligned} b(\emptyset) &= 0, \\ b(\tau) &= 1, \\ b([t_1 t_2 \cdots t_m]) &= a(t_1)a(t_2)\cdots a(t_m), \quad t_1, t_2, \dots, t_m \in T. \end{aligned} \tag{3.9 g}$$

Proof. The result is equivalent to Theorem 3.4E (p. 116). □

Note that $b(t)$ in this result is equal to $a(t \setminus \tau)$. Now generalize Lemma 3.9A.

Theorem 3.9B For $\mathrm{t}' \in \mathrm{T}$,

$$\frac{h^{|\mathrm{t}'|}}{\sigma(\mathrm{t}')}\mathbf{F}(\mathrm{t}')\big((\mathbf{B}_h a)y_0\big) = \sum_{\mathrm{t}\geq\mathrm{t}'}\frac{h^{|\mathrm{t}|}}{\sigma(\mathrm{t})}a(\mathrm{t}\setminus\mathrm{t}')\mathbf{F}(\mathrm{t}). \tag{3.9 h}$$

Proof. The case $\mathrm{t}' = \tau$ is covered by Lemma 3.9A. The general case will be proved by beta-induction for $\mathrm{t}' = \mathrm{t}'_0 * \mathrm{t}'_1$, where $\mathrm{t}'_0 = [\mathrm{t}_1'^{\,n-1}\mathrm{f}']$, and t'_1 is not a factor of the forest f'. Write $\mathrm{t} = \mathrm{t}_0 * \mathrm{t}_1$, where $\mathrm{t}_0 = [\mathrm{t}_1{}^{n-1}\mathrm{f}]$, and t_1 is not a factor of the forest f. From the induction hypothesis,

$$\frac{h^{|\mathrm{t}'_0|}}{\sigma(\mathrm{t}'_0)}\mathbf{F}(\mathrm{t}'_0)\big((\mathbf{B}_h a)y_0\big) = \sum_{\mathrm{t}_0\geq\mathrm{t}'}\frac{h^{|\mathrm{t}_0|}}{\sigma(\mathrm{t})}a(\mathrm{t}_0\setminus\mathrm{t}'_0)\mathbf{F}(\mathrm{t}_0),$$
$$\frac{h^{|\mathrm{t}'_1|}}{\sigma(\mathrm{t}'_1)}\mathbf{F}(\mathrm{t}'_1)\big((\mathbf{B}_h a)y_0\big) = \sum_{\mathrm{t}_1\geq\mathrm{t}'_1}\frac{h^{|\mathrm{t}_1|}}{\sigma(\mathrm{t}_1)}a(\mathrm{t}_1\setminus\mathrm{t}'_1)\mathbf{F}(\mathrm{t}),$$

so that (3.9 h), after division by n', becomes

$$\frac{h^{|\mathrm{t}'|}C}{\sigma(\mathrm{t}')}\mathbf{F}(\mathrm{t}')\big((\mathbf{B}_h a)y_0\big) = \sum_{\mathrm{t}_1\geq\mathrm{t}'_1}\sum_{\mathrm{t}_0\geq\mathrm{t}'_0}\frac{h^{|\mathrm{t}|}}{\sigma(\mathrm{t})}\mathbf{F}(\mathrm{t}),$$

where

$$C = \tfrac{n}{n'}a(\mathrm{t}_0\setminus\mathrm{t}'_0)a(\mathrm{t}_1\setminus\mathrm{t}'_1)$$

and where we have used the facts that $\sigma(\mathrm{t}') = n'\sigma(\mathrm{t}'_0)\sigma(\mathrm{t}'_1)$, $\sigma(\mathrm{t}) = n\sigma(\mathrm{t}_0)\sigma(\mathrm{t}_1)$. Simplify C using the formulae

$$\mathrm{t}_0\setminus\mathrm{t}'_0 = (\mathrm{t}_1\setminus\mathrm{t}'_1)^{n'-1}\mathrm{t}_1^{n-n'}\binom{n-1}{n'-1}(\mathrm{f}\setminus\mathrm{f}),$$
$$(\mathrm{t}_0 * \mathrm{t}_1)\setminus(\mathrm{t}'_0 * \mathrm{t}'_1) = (\mathrm{t}_1\setminus\mathrm{t}'_1)^{n'}\mathrm{t}_1^{n-n'}\binom{n}{n'}(\mathrm{f}\setminus\mathrm{f})$$

to give

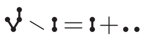
$$C = a\big((\mathrm{t}_0 * \mathrm{t}_1)\setminus(\mathrm{t}'_0 * \mathrm{t}'_1)\big).$$

By adding up C for all pairs for which $\mathrm{t}_0 * \mathrm{t}_1 = \mathrm{t}$, we arrive at (3.9 g). □

As an example, use $\mathrm{t}' = \mathbf{t}_2 = \tau * \tau$, $\mathrm{t} = \mathbf{t}_6$: We have two choices of the pair $(\mathrm{t}_1, \mathrm{t}_0)$

$(\mathrm{t}_1, \mathrm{t}_0) =$  and $(\mathrm{t}_1, \mathrm{t}_0) =$ ,

leading to



Formula for *ab*

For each tree, $(ab)(\mathrm{t})$ is a linear combination of terms of the form $b(\mathrm{t}')$ where $|\mathrm{t}'| \le |\mathrm{t}|$, with an additional term corresponding to t', replaced by $\varnothing$. We will find these linear combinations.

Theorem 3.9C Let $a \in \mathsf{B}$, $b \in \mathsf{B}^*$. Then

$$\begin{aligned}(ab)(\varnothing) &= b(\varnothing),\\ (ab)(\mathrm{t}) &= b(\varnothing)a(\mathrm{t}) + \sum_{\mathrm{t}' \le \mathrm{t}} b(\mathrm{t}')a(\mathrm{t} \smallsetminus \mathrm{t}'), \quad \mathrm{t} \in \mathrm{T}.\end{aligned}$$

Proof. Let $y_1 = (\mathbf{B}_h a)y_0$. Add $b(\varnothing)\big(y_1 = y_0 + \sum_{\mathrm{t}\in\mathrm{T}}(\sigma(\mathrm{t})^{-1}h^{|\mathrm{t}|}a(\mathrm{t})F(\mathrm{t})\big)$ to the sum over $\mathrm{t}' \in \mathrm{T}$ of $b(\mathrm{t}')$ times (3.9 h). The result is

$$\begin{aligned}b(\varnothing)y_1 &+ \sum_{\mathrm{t}\in\mathrm{T}} \frac{h^{|\mathrm{t}|}}{\sigma(\mathrm{t})}\mathbf{F}(\mathrm{t})(y_1)b(\mathrm{t})\\ &= b(\varnothing)y_0 + \sum_{\mathrm{t}\in\mathrm{T}} \frac{h^{|\mathrm{t}|}}{\sigma(\mathrm{t})}\mathbf{F}(\mathrm{t})\Big(a(\mathrm{t})b(\varnothing) + \sum_{\mathrm{t}'\le\mathrm{t}} a(\mathrm{t}\smallsetminus\mathrm{t})b(\mathrm{t}')\Big),\end{aligned}$$

which can be written

$$\begin{aligned}(ab)(\varnothing)y_0 &+ \sum_{\mathrm{t}\in\mathrm{T}} \frac{h^{|\mathrm{t}|}}{\sigma(\mathrm{t})}\mathbf{F}(\mathrm{t})(ab)(\mathrm{t})\\ &= b(\varnothing)y_0 + \sum_{\mathrm{t}\in\mathrm{T}} \frac{h^{|\mathrm{t}|}}{\sigma(\mathrm{t})}\mathbf{F}(\mathrm{t})\Big(a(\mathrm{t})b(\varnothing) + \sum_{\mathrm{t}'\le\mathrm{t}} a(\mathrm{t}\smallsetminus\mathrm{t}')b(\mathrm{t}')\Big).\end{aligned}$$

Compare the coefficients of y_0 and $\mathbf{F}(\mathrm{t})$ for all $\mathrm{t} \in \mathrm{T}$ and the result follows. □

A detailed example of Theorem 3.9C, for a particular order 6 tree t, will be given in Table 15. The specific tree chosen is

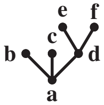
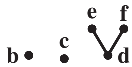
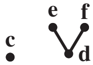
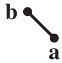
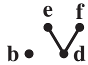
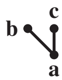
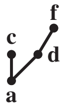
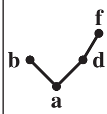
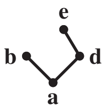
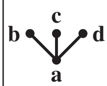
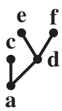
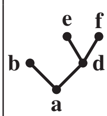
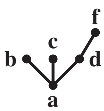
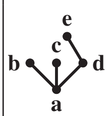
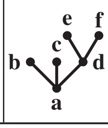
$$\mathrm{t} = \mathbf{t}_{20} = $$  . (3.9 i)

Sweedler notation

The results of Theorem 3.9C can be written in Sweedler form as

$$\begin{aligned}\Delta(\varnothing) &= 1 \otimes \varnothing,\\ \Delta(\mathrm{t}) &= a(\mathrm{t}) \otimes \varnothing + \sum_{\mathrm{t}'\le\mathrm{t}} a(\mathrm{t}\smallsetminus\mathrm{t}) \otimes \mathrm{t}'.\end{aligned}$$

Table 15 Details of terms in $(ab)(\mathrm{t})$, with t given by (3.9 i)

$\mathrm{t} \setminus \mathrm{t}'$	t'	Term	$\mathrm{t} \setminus \mathrm{t}'$	t'	Term
	$\varnothing$	$a_{20}b_0$			$a_1^2a_3b_1$
		$a_1a_3b_2$			$a_1a_3b_2$
		$a_1^4b_2$			$a_1^3b_4$
		$a_1^3b_4$			a_3b_3
		$a_1^2b_7$			$a_1^2b_6$
		$a_1^2b_6$			$a_1^2b_6$
		$a_1^2b_6$			$a_1^2b_5$
		a_1b_{11}			a_1b_{11}
		a_1b_{10}			a_1b_{10}
1		b_{20}			

Using the standard tree numbering and the notation $a_n := a(\mathbf{t}_n)$, these examples are, in Sweedler notation,

$$
\begin{aligned}
\Delta(\mathbf{t}_1) &= a_1 \otimes \varnothing + 1 \otimes \mathbf{t}_1,\\
\Delta(\mathbf{t}_2) &= a_2 \otimes \varnothing + a_1 \otimes \mathbf{t}_1 + 1 \otimes \mathbf{t}_2,\\
\Delta(\mathbf{t}_3) &= a_3 \otimes \varnothing + a_1^2 \otimes \mathbf{t}_1 + 2a_1 \otimes \mathbf{t}_2 + 1 \otimes \mathbf{t}_3,\\
\Delta(\mathbf{t}_4) &= a_4 \otimes \varnothing + a_2 \otimes \mathbf{t}_1 + a_1 \otimes \mathbf{t}_3 + 1 \otimes \mathbf{t}_4.
\end{aligned}
$$

Using $\lambda(a,\mathrm{t})$

Now recall a notation based on λ, first introduced in [14] (Butcher 1972).

Definition 3.9D For $a \in \mathsf{B}$, $\mathrm{t} \in T$,

$$\lambda(a,\mathrm{t}) = \sum_{\mathrm{t}' \le \mathrm{t}} a(\mathrm{t} \setminus \mathrm{t}')\mathrm{t}'.$$

In the following statement, the beta product notation will be generalized so that it becomes linear when its operands are applied to a formal linear combination of trees. That is

$$\Big(\sum_{\mathrm{t}\in T} x(\mathrm{t})\mathrm{t}\Big) * \Big(\sum_{\mathrm{t}'\in T} y(\mathrm{t}')\mathrm{t}'\Big) = \sum_{\mathrm{t}\in T}\sum_{\mathrm{t}'\in T} x(\mathrm{t})y(\mathrm{t}')(\mathrm{t}*\mathrm{t}').$$

Theorem 3.9E Let $a \in \mathsf{B}$, $\mathrm{t}_1, \mathrm{t}_2 \in T$. Then

$$
\begin{aligned}
\lambda(a,\tau) &= \tau,\\
\lambda(a,\mathrm{t}_1 * \mathrm{t}_2) &= a(\mathrm{t}_2)\lambda(a,\mathrm{t}_1) + \lambda(a,\mathrm{t}_1) * \lambda(a,\mathrm{t}_2), \qquad \mathrm{t}_1, \mathrm{t}_2 \in T.
\end{aligned}
\tag{3.9j}
$$

Proof. Let $\mathrm{t} = \mathrm{t}_1 * \mathrm{t}_2$ and write

$$\lambda(a,\mathrm{t}_1) = \sum_{\mathrm{t}_1' \le \mathrm{t}_1} a(\mathrm{t}_1 \setminus \mathrm{t}_1')\mathrm{t}_1', \qquad \lambda(a,\mathrm{t}_2) = \sum_{\mathrm{t}_2' \le \mathrm{t}_2} a(\mathrm{t}_2 \setminus \mathrm{t}_2')\mathrm{t}_2'.$$

The trees such that $\mathrm{t}' \le \mathrm{t}$ are of the form
(i) t_1', with $\mathrm{t} \setminus \mathrm{t}' = \mathrm{t}_2(\mathrm{t}_1 \setminus \mathrm{t}_1')$ or (ii) $\mathrm{t}_1' * \mathrm{t}_2'$, with $\mathrm{t} \setminus \mathrm{t}' = (\mathrm{t}_1 \setminus \mathrm{t}_1')(\mathrm{t}_2 \setminus \mathrm{t}_2)$.
Hence,

$$
\begin{aligned}
\lambda(a,\mathrm{t}) &= \sum_{\mathrm{t}_1' \le \mathrm{t}_1} a(\mathrm{t}_2)a(\mathrm{t}_1 \setminus \mathrm{t}_1')\mathrm{t}_1' + \sum_{\mathrm{t}_1' \le \mathrm{t}_1}\sum_{\mathrm{t}_2' \le \mathrm{t}_2} a(\mathrm{t}_1 \setminus \mathrm{t}_1')a(\mathrm{t}_2 \setminus \mathrm{t}_2)(\mathrm{t}_1' * \mathrm{t}_2')\\
&= a(\mathrm{t}_2)\sum_{\mathrm{t}_1' \le \mathrm{t}_1} a(\mathrm{t}_1 \setminus \mathrm{t}_1')\mathrm{t}_1' + \Big(\sum_{\mathrm{t}_1' \le \mathrm{t}_1} a(\mathrm{t}_1 \setminus \mathrm{t}_1')\mathrm{t}_1'\Big) * \Big(\sum_{\mathrm{t}_2' \le \mathrm{t}_2} a(\mathrm{t}_2 \setminus \mathrm{t}_2)\mathrm{t}_2'\Big)\\
&= a(\mathrm{t}_2)\lambda(a,\mathrm{t}_1) + \lambda(a,\mathrm{t}_1) * \lambda(a,\mathrm{t}_2).
\end{aligned}
$$

□

Algorithm 5 Evaluate λ

Input: *order*, *first*, *last*, *L*, *R*, *prod* from Algorithm 3 (p. 64); *p_top*, *a*
Output: λ

```
%
%   λ[1 : p_top, 1 : p_top] is the matrix of λ values
%   p_top ≤ p_max
%   a[1:last[p_top]] is the vector of a values
%
for i from 1 to last[p_top] do
   for j from 1 to last[p_top] do
      λ[i, j] ← 0
   end for
end for
λ[1,1] ← 1
for n from 2 to last[p_top] do
   i ← L[n]
   j ← R[n]
   for k from 1 to i do
      λ[n,k] ← a[j] ∗ λ[i,k]
   end for
   for k from 1 to i do
      for ℓ from 1 to j do
         λ[n, prod(k,ℓ)] ← λ[n, prod(k,ℓ)] + λ[i,k] ∗ λ[j,ℓ]
      end for
   end for
end for
```

Examples of $\lambda(a,\mathrm{t})$

Start with $\mathrm{t} = \tau = \mathbf{t}_1$ and calculate for all trees such that $|\mathrm{t}| \leq 4$.

$$\begin{aligned}
\lambda(a,\mathbf{t}_1) & &&= \mathbf{t}_1,\\
\lambda(a,\mathbf{t}_2) &= \lambda(a,\mathbf{t}_1 * \mathbf{t}_1) = a_1\mathbf{t}_1 + \mathbf{t}_1 * \mathbf{t}_1 &&= a_1\mathbf{t}_1 + \mathbf{t}_2,\\
\lambda(a,\mathbf{t}_3) &= \lambda(a,\mathbf{t}_2 * \mathbf{t}_1) = a_1(a_1\mathbf{t}_1 + \mathbf{t}_2) + (a_1\mathbf{t}_1 + \mathbf{t}_2) * \mathbf{t}_1 &&= a_1^2\mathbf{t}_1 + 2a_1\mathbf{t}_2 + \mathbf{t}_3,\\
\lambda(a,\mathbf{t}_4) &= \lambda(a,\mathbf{t}_1 * \mathbf{t}_2) = a_2\mathbf{t}_1 + \mathbf{t}_1 * (a_1\mathbf{t}_1 + \mathbf{t}_2) &&= a_2\mathbf{t}_1 + a_1\mathbf{t}_2 + \mathbf{t}_4.
\end{aligned}$$

Exercise 37 Evaluate $\lambda(a,\mathbf{t}_6)$ based on $\mathbf{t}_6 = \mathbf{t}_4 * \mathbf{t}_1$.

Exercise 38 Evaluate $\lambda(a,\mathbf{t}_6)$ based on $\mathbf{t}_6 = \mathbf{t}_2 * \mathbf{t}_2$.

Algorithm for λ

An algorithm to evaluate λ is presented in Algorithm 5. As a partial explanation of how Algorithm 5 works, consider the evaluation of $\lambda(a,\mathbf{t}_6)$. From earlier steps it will already be known that

$$\begin{aligned}\lambda(a,\mathbf{t}_1) &= \mathbf{t}_1,\\ \lambda(a,\mathbf{t}_2) &= a_1\mathbf{t}_1+\mathbf{t}_2,\\ \lambda(a,\mathbf{t}_4) &= a_2\mathbf{t}_1+a_1\mathbf{t}_2+\mathbf{t}_3.\end{aligned}$$

These are represented by vectors in $\mathbb{R}^8$ (to allow for all trees to order 4):

$$\begin{aligned}\lambda(a,\mathbf{t}_1) &\mapsto [1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0],\\ \lambda(a,\mathbf{t}_2) &\mapsto [a_1\ \ 1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0],\\ \lambda(a,\mathbf{t}_4) &\mapsto [a_2\ \ a_1\ \ 0\ \ 1\ \ 0\ \ 0\ \ 0\ \ 0].\end{aligned}$$

Two alternative calculations to evaluate the representation of $\lambda(a,\mathbf{t}_6)$ are based on $\mathbf{t}_6=\mathbf{t}_2*\mathbf{t}_2$ and $\mathbf{t}_6=\mathbf{t}_4*\mathbf{t}_1$, respectively. These give the results

$$\begin{aligned}&[0\ \ a_1^2\ \ a_1\ \ a_1\ \ 0\ \ 1\ \ 0\ \ 0]+[a_1a_2\ \ a_2\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0]\\ &\qquad\qquad\qquad\qquad = [a_1a_2\ \ a_1^2+a_2\ \ a_1\ \ a_1\ \ 0\ \ 1\ \ 0\ \ 0],\\ &[0\ \ a_2\ \ a_1\ \ 0\ \ 0\ \ 1\ \ 0\ \ 0]+[a_1a_2\ \ a_1^2\ \ 0\ \ a_1\ \ 0\ \ 0\ \ 0\ \ 0]\\ &\qquad\qquad\qquad\qquad = [a_1a_2\ \ a_1^2+a_2\ \ a_1\ \ a_1\ \ 0\ \ 1\ \ 0\ \ 0].\end{aligned}$$

Efficient representation of λ coefficients

In applications of Algorithm 5, it will be convenient to work with a modified version of λ, in which the diagonal elements are omitted. That is, we find it convenient to work with

$$\widehat{\lambda}(a,t) := \lambda(a,t)-t,$$

rather than $\lambda(a,t)$ itself.

Corresponding to the result of Theorem 3.9B, we have

> Corollary 3.9F
>
> $$\begin{aligned}\widehat{\lambda}(a,\tau) &= 0,\\ \widehat{\lambda}(a,t_1*t_2) &= a(t_2)\widehat{\lambda}(a,t_1)+\widehat{\lambda}(a,t_1)*\widehat{\lambda}(a,t_2)\\ &\quad + a(t_2)t_1+\widehat{\lambda}(a,t_1)*t_2+t_1*\widehat{\lambda}(a,t_2), \qquad t_1,t_2\in T.\end{aligned}$$

Proof. Substitute into (3.9 j). □

Algorithm 6 Evaluate $\widehat{\lambda}$

Input: $order, first, last, L, R, prod$ from Algorithm 3 (p. 64); p_top, a
Output: $\widehat{\lambda}$

%
% $\widehat{\lambda}[n,1:last[order[n]-1], n=1,2,\dots,last[p\_top]$ is the sequence of $\widehat{\lambda}$ values
% $p\_top \le p\_max$
% $a[1:last[p\_top]$ is the vector of a values
%
$\widehat{\lambda}[1,1] \leftarrow 0$
for n **from** 2 **to** $last[p\_top]$ **do**
  for i **from** 1 **to** $last[order[i]-1]]$ **do**
    $\widehat{\lambda}[n,i] \leftarrow 0$
  end for
  $i \leftarrow L[n]$
  $j \leftarrow R[n]$
  $\widehat{\lambda}[n,i] \leftarrow a[j]$
  for k **from** 1 **to** $last[order[i]-1]$ **do**
    $\widehat{\lambda}[n,k] \leftarrow a[j] * \widehat{\lambda}[i,k]$
  end for
  for ℓ **from** 1 **to** $last[order[j]-1]$ **do**
    $\widehat{\lambda}[n, prod[i,\ell]] \leftarrow \widehat{\lambda}[n, prod[i,\ell]] + \widehat{\lambda}[j,\ell]$
  end for
  for k **from** 1 **to** $last[order[i]-1]$ **do**
    $\widehat{\lambda}[n, prod[k,j]] \leftarrow \widehat{\lambda}[n, prod(k,j]] + \widehat{\lambda}[i,k]$
  end for
  for k **from** 1 **to** $last[order[i]-1]$ **do**
    for ℓ **from** 1 **to** $last[order[j]-1]$ **do**
      $\widehat{\lambda}[n, prod(k,\ell)] \leftarrow \widehat{\lambda}[n, prod(k,\ell)] + \widehat{\lambda}[i,k] * \widehat{\lambda}[j,\ell]$
    end for
  end for
end for

Implementation of algorithms

A realization of Algorithm 5 (p. 142) and Algorithm 6 gave consistent results. A slightly edited sample of results from Algorithm 6 are, as a list of lists,

$$\begin{aligned}
&\{\{\}, \{a_1\}, \{a_1^2, 2a_1\}, \{a_2, a_1\}, \{a_1^3, 3a_1^2, 3a_1, 0\}, \{a_1a_2, a_2+a_1^2, a_1, a_1\},\\
&\{a_3, a_1^2, 0, 2a_1\}, \{a_4, a_2, 0, a_1\}, \{a_1^4, 4a_1^3, 6a_1^2, 0, 4a_1, 0, 0, 0\},\\
&\{a_2a_1^2, 2a_1a_2+a_1^3, a_2+2a_1^2, a_1^2, a_1, 2a_1, 0, 0\}, \{a_1a_3, a_3+a_1^3, a_1^2, 2a_1^2, 0, 2a_1, a_1, 0\},\\
&\{a_1a_4, a_4+a_1a_2, a_2, a_1^2, 0, a_1, 0, a_1\}, \{a_2^2, 2a_1a_2, a_1^2, 2a_2, 0, 2a_1, 0, 0\},\\
&\{a_5, a_1^3, 0, 3a_1^2, 0, 0, 3a_1, 0\}, \{a_6, a_1a_2, 0, a_2+a_1^2, 0, 0, a_1, a_1\},\\
&\{a_7, a_3, 0, a_1^2, 0, 0, 0, 2a_1\}, \{a_8, a_4, 0, a_2, 0, 0, 0, a_1\}\}
\end{aligned}$$

Inverse of B-series

Given an invertible map $\mathcal{A}_h = \mathbf{B}_h(a)$ we want to find the coefficients a^{-1} to represent $\mathbf{B}_h(a^{-1})$ corresponding to $\mathcal{A}_h^{-1}$. This can be done very simply from the composition Theorem 3.9C by writing (3.9 h) in one of the forms

$$b(t) = -a(t) - \sum_{\varnothing < \mathrm{t}' < \mathrm{t}} a(\mathrm{t} \setminus \mathrm{t}')b(\mathrm{t}'),$$
$$b(t) = -a(t) - \sum_{\varnothing < \mathrm{t}' < \mathrm{t}} b(\mathrm{t} \setminus \mathrm{t}')a(\mathrm{t}'),$$

where $(ab)(\mathrm{t}) = 0$, so that $b = a^{-1}$.

Antipode and Newton iteration

A second formulation of a B-series inverse is to use the antipode.

Finally, the inverse may be found by a Newton-like iteration

$$\begin{aligned} x^{[0]} &= 1, \\ x^{[k]} &= 2x^{[k-1]} - x^{[k-1]}ax^{[k-1]}, \qquad k = 1, 2, \dots . \end{aligned} \tag{3.9 k}$$

This iteration scheme is quadratically convergent in the sense that the order of the trees, for which it is exact, doubles in each iteration.

Theorem 3.9G Define $x^{[k]}$ by the iteration scheme (3.9 k). Then

$$(ax^{[k]})(\mathrm{t}) = 0, \qquad |\mathrm{t}| \le 2^{k-1}.$$

Proof. Let $y = ax^{[k-1[}$. We will show that $y(\tau) = 0$ when $k = 1$ and that, if $y(\mathrm{t}) = 0$ for $|\mathrm{t}| \le n$, then

$$(2y - y^2)(\mathrm{t}) = 0, \qquad |\mathrm{t}| \le 2n.$$

From the composition rule,

$$2y(\mathrm{t}) - (y^2)(\mathrm{t}) = \sum_{\varnothing < \mathrm{t}' < \mathrm{t}} y(\mathrm{t} \setminus \mathrm{t}')y(\mathrm{t}')$$

and all terms on the right-hand side are zero. □

Compositions using $\boldsymbol{\Lambda}$

Recall the introduction of $\boldsymbol{\Lambda}$ in Chapter 2, Section 2.9 (p. 95). $\boldsymbol{\Lambda}$ is the lower triangular infinite matrix indexed by the $\varnothing$, followed by the sequence of all trees in the standard order. The diagonal elements are given by $\boldsymbol{\Lambda}(\varnothing, \varnothing) = 1$, $\boldsymbol{\Lambda}(\mathrm{t}, \mathrm{t}) = 1$ for all $\mathrm{t} \in T$ and the lower diagonals are given by $\boldsymbol{\Lambda}(\mathrm{t}, \varnothing) = \mathrm{t}$ and $\boldsymbol{\Lambda}(\mathrm{t}, \mathrm{t}') = \mathrm{t} \setminus \mathrm{t}'$. We also write v as an infinite vector with the same index set, where $\mathsf{v}(\varnothing) = \varnothing$ and $\mathsf{v}(\mathrm{t}) = \mathrm{t}$. Thus,

$$\boldsymbol{\Lambda} = \begin{bmatrix} 1 & & & & & \\ \mathrm{t}_1 & 1 & & & & \\ \mathrm{t}_2 & \mathrm{t}_1 & 1 & & & \\ \mathrm{t}_3 & \mathrm{t}_1^2 & 2\mathrm{t}_1 & 1 & & \\ \mathrm{t}_4 & \mathrm{t}_2 & \mathrm{t}_1 & 0 & 1 & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \qquad \mathsf{v} = \begin{bmatrix} \varnothing \\ \mathrm{t}_1 \\ \mathrm{t}_2 \\ \mathrm{t}_3 \\ \mathrm{t}_4 \\ \vdots \end{bmatrix}.$$

Expressions like $\boldsymbol{\Lambda}(x)$ will denote the matrix $\boldsymbol{\Lambda}$ with every element replaced by a real number formed by evaluating $x(\mathrm{t})$ for each t appearing in this element. Similarly, $\mathsf{v}(u)$ is the result of evaluating u in the case of each element. This means

$$\boldsymbol{\Lambda}(x) = \begin{bmatrix} 1 & & & & & \\ x(\mathrm{t}_1) & 1 & & & & \\ x(\mathrm{t}_2) & x(\mathrm{t}_1) & 1 & & & \\ x(\mathrm{t}_3) & x(\mathrm{t}_1)^2 & 2x(\mathrm{t}_1) & 1 & & \\ x(\mathrm{t}_4) & x(\mathrm{t}_2) & x(\mathrm{t}_1) & 0 & 1 & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \qquad \mathsf{v}(u) = \begin{bmatrix} u(\varnothing) \\ u(\mathrm{t}_1) \\ u(\mathrm{t}_2) \\ u(\mathrm{t}_3) \\ u(\mathrm{t}_4) \\ \vdots \end{bmatrix}.$$

It is convenient to rewrite Theorem 3.9C (p. 139) in the form

$$\mathsf{v}(ab) = \boldsymbol{\Lambda}(a)\mathsf{v}(b). \tag{3.91}$$

By regarding the infinite representation of a member of B^* in terms of v as the standard representaion, we can write (3.91) in the form

$$ab = \boldsymbol{\Lambda}(a)b,$$

where ab on the left is an algebraic expression and $\boldsymbol{\Lambda}(a)b$ on the right is a matrix–vector product.

Properties of $\boldsymbol{\Lambda}$ applied to functions of trees

The following identities hold

$$\begin{aligned} \boldsymbol{\Lambda}(x)\boldsymbol{\Lambda}(y) &= \boldsymbol{\Lambda}(xy), \\ \boldsymbol{\Lambda}(x)\mathsf{v}(u) &= \mathsf{v}(xu), \\ x(yz) &= (xy)z, \\ x(u+v) &= xu + xv, \\ x(cu) &= c(xu) \qquad (c \text{ scalar}). \end{aligned}$$

Fractional powers

The integer powers of $a \in \mathsf{B}$ can be calculated as $\Lambda(a)^{n-1}a$ or as the first column of $\Lambda(a)^n$. To extend this to fractional powers, write $\Lambda(a) = I + L$, with I the infinite dimensional identity matrix. We then have the binomial theorem formula

$$\Lambda(a)^x = I + \sum_{i=1}^{\infty} \frac{x(x-1)\cdots(x-i+1)}{i!} L^i. \tag{3.9 m}$$

The following result has an application in Chapter 6, Section 6.6 (p. 241).

Lemma 3.9H

$$a^x(\tau) = xa(\tau).$$

Proof. $a^x(\tau)$ is the $(\tau, \varnothing)$ element of the $i = 1$ term in (3.9 m). □

Summary of Chapter 3 and the way forward

Summary

One of the key questions in the analysis of numerical approximations to differential equations, and related problems, is in the comparison of two mappings. One mapping would be the exact flow through a specified time step and the other would be a numerical scheme, such as a Runge–Kutta method. The B-series approach to questions like this is to write the Taylor expansions of the two mappings in a special way, in terms of “elementary differentials”, and to then compare the coefficients in corresponding terms. The terms themselves can be indexed in terms of the objects known as “rooted trees” and the theory of B-series hinges on this indexing.

The chapter includes a discussion of multi-dimensional Taylor series and it is shown how this leads to the formulation of elementary differentials and B-series. Some sample problems, which are both easy and fundamental, are solved, first in a low order introduction and then in full generality. Special attention is given to the flow through a unit step, followed by an implicit variant of the Euler method. It is remarkable that the latter simple example is a direct path to the large family of Runge–Kutta methods. The composition rule for B-series is introduced and some of its wider ramifications are explored.

The way forward

Group and algebraic structures

In Chapter 4 we will look at the algebra associated with B^* as a mathematical system in its own right. It will be seen that the elements of B form a group with the same group operation as required for the B-series composition rule to hold.

We will continue to use $\mathrm{D} \in \mathsf{B}^0$ and $\mathrm{E} \in \mathsf{B}$ but with the algebraic meaning as well as their meaning as representing B-series coefficients. The meaning of $a\mathrm{D}$ given in Section 3.4 can now be compared with $a\mathrm{D}$ evaluated by the composition rule.

Linear operators on B^0

Corresponding to unit valency stumps introduced in Section 2.7, and their algebra, we can construct an algebra of linear operators on B^0. The multiplicative semigroup of linear operators consists of the identity operator and arbitrary products of

$$h\mathbf{f}', \quad h^2\mathbf{f}''\mathbf{f}, \quad h^2\mathbf{f}'\mathbf{f}',$$

corresponding to the valency one stumps

$$\tau_1, \quad \tau_2\tau_1, \quad \tau_1^2.$$

Runge–Kutta methods

The order conditions, established in Theorem 3.6C (p. 127), are applied in Chapter 5 to the derivation of specific practical methods.

Teaching and study notes

This is the central chapter of this volume and it should be studied in detail before later chapters are attempted. The key ideas are the multi-dimensional Taylor series and the Taylor expansion of the flow and related problems in terms of "elementary differentials". The concept of B-series is introduced through the formula

$$\mathscr{a}_h y_0 = (\mathbf{B}_h y_0)a = a(\varnothing)y_0 + \sum_{\mathrm{t}\in\mathrm{T}} \frac{h^{|\mathrm{t}|}a(\mathrm{t})}{\sigma(\mathrm{t})}(\mathbf{F}(\mathrm{t})y_0).$$

If $a(\varnothing) = 1$, then $\mathscr{b}_h \circ \mathscr{a}_h$ is defined by the mapping $y_0 \mapsto \mathscr{b}_h(\mathscr{a}_h y_0)$ and is given by the B-series composition rule

$$\mathbf{B}_h(ab)(y_0) = \mathbf{B}_h(b)(\mathbf{B}_h(a)(y_0)).$$

An important consideration is the nature of the product

$$(a,b) \mapsto ab.$$

Possible supplementary reading includes

Butcher, J.C. *An algebraic theory of integration methods* (1972) [14]

Butcher, J.C. *Numerical Methods for Ordinary Differential Equations* (2016) [20]

Hairer, E., Nørsett, S.P. and Wanner, G. *Solving Ordinary Differential Equations I: Nonstiff Problems* (1993) [50]

Hairer, E. and Wanner, G. *On the Butcher group and general multi-value method.* (1974) [52]

Projects

Project 9 Carry out a detailed study of the analytical theory of Fréchet differentiation. A possible starting point is [81].

Project 10 Study Taylor series in $\mathbb{R}^N$, including error estimates.

Project 11 Learn about a modified B-series theory suitable for the problem $y''(x) = f\big(y(x)\big)$, with y and y' given at an initial point.