Formal Power Series Methods in Nonlinear Control Theory

W. Steven Gray

Formal Power Series Methods
in Nonlinear Control Theory
Edition 1.2

W. Steven Gray

Book Website: Updates and other information related to this book are available at:

/http://www.ece.odu.edu/∼sgray/fps-book

Citation Information:

W. S. Gray, *Formal Power Series Methods in Nonlinear Control Theory*, Ed. 1.2, 2022, available at http://www.ece.odu.edu /∼ sgray/fps-book.

Edition 1.2 - March 19, 2022

Preface

These lecture notes introduce a collection of techniques for the analysis of nonlinear control systems rooted in the theory of formal power series and their associated combinatorial algebras. The origins of the formal power series methods in control theory begin with the work R. E. Kalman in the 1960s, primarily in connection with the partial realization problem for linear systems. Parallel developments by M. P. Schützenberger in automata theory and the remarkable discoveries of K.-T. Chen in the 1950s concerning the algebra of iterated path integrals subsequently led M. Fliess in the 1970s to introduce what is now called a Chen-Fliess series or Fliess operator. The underlying formal power series which generates such an operator provides an elegant and compact way to represent the input-output map of a control affine nonlinear system. As these generating series are indexed by words over a noncommutative alphabet, there is a natural link between nonlinear control theory and the combinatorics of words, a mature and beautiful field going back to seminal papers of A. Thue at the start of the twentieth century.

Following the initial work of Fliess, the area grew rapidly with important contributions by P. E. Crouch, G. Duchamp, A. Ferfera, R. L. Grossman, C. Hespel, V. Hoang Ngoc Minh, A. Isidori, G. Jacob, B. Jakubczyk, M. Kawski, D. Krob, M. Lamnabhi, F. Lamnabhi-Lagarrigue, R. G. Larson, P. Lemux, N. E. Oussous, M. Petitot, C. Reutenauer, F. Rotella, W. J. Rugh, E. D. Sontag, H. J. Sussmann, X. G. Viennot, and Y. Wang. The work of A. Ferfera was especially valuable for understanding how to describe interconnected nonlinear systems using Chen-Fliess series. More recently, the author, in collaboration with L. A. Duffaut Espinosa and K. Ebrahimi-Fard, has built on the work of Ferfera to show that there are combinatorial Hopf algebras underlying the feedback structures appearing in control theory. They are useful for doing explicit calculations. This approach was largely inspired by analogous combinatorial algebras appearing in work of A. Connes, D. Kreimer, and H. Moscovici in quantum field theory and in J. C. Butcher's approach to numerical integration. One of the goals of these notes is to introduce the reader to some of these more recent developments. Finally, it should be stated that these notes are simply an introduction to the subject from one researcher's perspective. A more encyclopedic treatment of this topic is well beyond the scope of this project.

It is generally assumed that the reader has had an introduction to linear system theory, say at the level of Kailath's *Linear Systems*, and some exposure to elementary topics in real analysis, abstract algebra, and differential geometry. But otherwise, the treatment of the subject is from first principles and is as self-contained as possible. The material is organized as follows. The first chapter is an overview of the central topics that appear in later chapters. It is written in a more casual style and meant to motivate the formal power series approach to system theory while staying mainly in a linear system setting. It is designed to be largely independent of the other chapters so that a more experienced reader can start directly with Chapter 2 with little loss of continuity. The rest of the book follows a more systematic theorem-proof format. Chapter 2 introduces some elementary background and tools concerning formal power series. Chapter 3 then addresses the analysis of nonlinear input-output systems and their interconnections from a formal power series perspective. Chapters 4 and 5 introduce the notions of rational series and Lie series, respectively, which are then applied in Chapter 6 to develop the theory for state space realizations of nonlinear input-output systems. The last chapter introduces a collection of special topics and applications reflecting, of course, the author's interests and experience with the subject.

W. S. Gray
June 2021

Acknowledgments

Many people have helped me along the way with this project. I first want to acknowledge my primary research collaborators in this area: Luis A. Duffaut Espinosa, Kurusch Ebrahimi-Fard, and Yuan Wang. Many of the ideas which appear in this book were formed and shaped during our work together. In addition, I want to thank my former graduate students who worked with me on this topic: Lance Berlin, Luis A. Duffaut Espinosa, Md. Aminul Haq, Yaqin Li, Natalie Pham, Makhin Thitsa, G. S. Ventkatesh, and Irina M. Winter-Arboleda. They provided fresh insights and contributed significantly to the progress in this field. I also want to acknowledge my colleagues: Rafael Dahmen, Héctor Figueroa, Loïc Foissy, Bronisław Jakubczyk, Dmitry S. Kaliuzhnyi-Verbovetskyi, Matthias Kawski, Terry Lyons, Simon J. A. Malham, Mathias Palmstrøm, Mihály Petreczky, Witold Respondek, Eva Riccomagno, Wilson (Jack) Rugh, and Alexander Schmeding for the many interesting technical discussions related to the topic. I am especially indebted to the organizers of the Trimester in Combinatorics and Control held in Madrid, Spain in 2010, namely, Kurusch Ebrahimi-Fard, Matthias Kawski, and David Martín de Diego, for the invitation to participate in this unique event. I learned a great deal through my collaborations there and the many subsequent workshops and conferences. I am also indebted to the Instituto de Ciencias Matemáticas in Madrid and my host Kurusch Ebrahimi-Fard for supporting my sabbatical leave during 2014. I truly enjoyed every minute of my visit there. It was a great privilege to be among such fine scholars. Finally, I want to thank my wife, Thu Khanh, and my children Michelle and Philippe for their generous love and support during the many years that I worked on this topic and the manuscript.

Table of Contents

1. Introduction

This chapter introduces some elementary concepts concerning real analytic functions, formal functions, linear integral operators, and linear state space realizations. The primary goal is to motivate the more general treatment of these topics in subsequent chapters. From a system theory point of view, real analytic functions provide a convenient class of input and output signals. They can also be used to describe various types of analytic *systems*. A formal function is a type of generalized function which is sometimes more convenient for algebraic analysis than a traditional function. Integral operators, linear or otherwise, describe a category of input-output systems frequently encountered in applications. Starting with the linear case provides a familiar setting in which to get oriented. Operators which have a finite dimensional linear time-invariant state space realization are of particular importance as they are computationally convenient and ubiquitous in systems and control theory.

1.1 Real Analytic Functions

A function $u : U \subseteq \mathbb{R} \to \mathbb{R}$ is said to be *real analytic* at a point $t_0 \in U$, if it can be represented in terms of a convergent power series

$$u(t) = \sum_{n=0}^{\infty} c(n) \frac{(t - t_0)^n}{n!} \tag{1.1}$$

on an interval $(t_0 - T, t_0 + T) \subset U$, where c is a sequence of real numbers, and T is either a positive real number or $T = +\infty$. The largest such T for which the series (1.1) converges is referred to as the *radius of convergence* of u at t_0. When $T = +\infty$, u is said to be *entire*. The function u is real analytic on an interval $(a, b) \subset U$ if it is real analytic at every point $t_0 \in (a, b)$. In this case, the radius of convergence may vary as a function of t_0.

One can extend the definition of u to a mapping on the complex plane, $\mathbb{C}$, by letting

$$u(z) = \sum_{n=0}^{\infty} c(n) \frac{(z - t_0)^n}{n!} \tag{1.2}$$

on the largest open disk $\mathcal{D}_0 = \{z \in \mathbb{C} : |z - t_0| < T\}$ for which the series converges. The function is then analytic at t_0 in the sense of complex variables, that is, the derivative of u exists not only at t_0 but at every point in some neighborhood of t_0 in the complex plane. In fact, all the higher order derivatives of u are well defined in such a neighborhood. The Cauchy integral formula says in this case that the n-th derivative of u at t_0 can be computed as

$$u^{(n)}(t_0) = \frac{n!}{2\pi i} \oint_C \frac{u(z)}{(z - t_0)^{n+1}} \, dz,$$

where C is a closed contour lying within $\mathcal{D}_0$ and encircling the point t_0, for example, all z satisfying $|z - t_0| = T' < T$. Since u is analytic on and within the region $\mathcal{D}_0' = \{z \in \mathbb{C} : |z - t_0| \leq T'\}$, the real-valued function $u(z)$ is continuous on this closed and bounded region. Hence, there exists a nonnegative real number K satisfying

$$K = \max_{z \in \mathcal{D}_0'} |u(z)| \,.$$

Setting $z(t) = t_0 + T'\mathrm{e}^{it}$ on $[0, 2\pi]$ and applying the identity

$$\oint_C f(z) \, dz = \int_0^{2\pi} f(z(t)) z'(t) \, dt$$

gives

$$\begin{aligned} |c(n)| &= \left| u^{(n)}(t_0) \right| \\ &= \frac{n!}{2\pi} \left| \oint_C \frac{u(z)}{(z - t_0)^{n+1}} \, dz \right| \\ &= \frac{n!}{2\pi} \left| \int_0^{2\pi} \frac{u(t_0 + T'\mathrm{e}^{it})}{(T'\mathrm{e}^{it})^{n+1}} \, iT'\mathrm{e}^{it} \, dt \right| \\ &\leq \frac{n!}{2\pi} \cdot \frac{K}{(T')^n} \int_0^{2\pi} dt \\ &= KM^n n!, \quad \forall n \geq 0, \end{aligned} \tag{1.3}$$

where $M := 1/T'$. This means then that the coefficients of an analytic function can only grow in modulus at a certain maximum rate as n increases. A sequence like $c(n) = (n!)^2$, for example, can never define an analytic function. The inequality (1.3) is called a *Cauchy growth condition* and the real numbers K and M are *growth constants* for the sequence c. They are not necessarily unique unless one considers the smallest possible growth constants. Normally, M will be referred to as a *geometric* growth constant for c. If some estimate M' is available a priori for a given sequence c then a lower bound on the radius of convergence for series (1.2) can be computed using the fact that

$$\begin{aligned} |u(z)| &\leq \sum_{n=0}^{\infty} |c(n)| \frac{|(z-t_0)^n|}{n!} \\ &\leq \sum_{n=0}^{\infty} K(M'|z-t_0|)^n. \end{aligned}$$

That is, the series converges at least for all $z \in \mathbb{C}$ such that $M'|z - t_0| < 1$, or equivalently, when

$$|z - t_0| < 1/M' < T.$$

Conversely, if u is analytic at the origin with the nearest singularity being $z' \neq 0$, then for any $\epsilon > 0$ there exists $N \in \mathbb{N}_0 := \{0, 1, 2, \ldots\}$ such that

$$|c(n)| \leq \left(\frac{1}{|z'|} + \epsilon\right) n!, \quad n \geq N.$$

In which case, one can always find a constant $K > 1$ such that

$$|c(n)| \leq K \left(\frac{1}{|z'|} + \epsilon\right) n!, \quad n \geq 0.$$

It is not difficult to show that $1/|z'|$ is in fact the *smallest* possible geometric growth constant for c. Finally, if it is known that c satisfies the more restrictive growth condition

$$|c(n)| \leq KM^n, \quad \forall n \geq 0 \tag{1.4}$$

then clearly

$$|u(z)| \leq K\mathrm{e}^{M|z-t_0|}, \quad \forall z \in \mathbb{C},$$

implying that u is entire. Analogous statements can be made for the real analytic series (1.1) by restricting z to the real number line.

A power series representation of a real analytic function is unique in a local sense. Suppose there exists two series representations of u at t_0, namely,

$$u(t) = \sum_{n=0}^{\infty} c(n) \frac{(t-t_0)^n}{n!}$$

and

$$u(t) = \sum_{n=0}^{\infty} d(n) \frac{(t-t_0)^n}{n!}$$

with radii of convergence T_c and T_d, respectively. Since u is continuous at t_0 (see Problem 1.1.3), it follows that

$$\lim_{t\to t_0} u(t) = \lim_{t\to t_0} \sum_{n=0}^{\infty} c(n) \frac{(t-t_0)^n}{n!} = c(0)$$
$$\lim_{t\to t_0} u(t) = \lim_{t\to t_0} \sum_{n=0}^{\infty} d(n) \frac{(t-t_0)^n}{n!} = d(0),$$

and thus, $c(0) = d(0)$. A similar argument can be made for the function

$$\tilde{u}(t) := \frac{u(t) - c(0)}{t - t_0} = \frac{u(t) - d(0)}{t - t_0}$$

to show that $c(1) = d(1)$, and so on. In addition, since $c(n) = d(n)$ for all $n \geq 0$, it follows immediately that $T_c = T_d =: T$. Thus, if there exists a power series representation of u at t_0, it is unique. It is also easily shown that if $t_1 \in \mathbb{R}$ such that $|t_1 - t_0| < T$ then the series representation of u can be *re-centered* about t_1. That is, u can be written as

$$u(t) = \sum_{n=0}^{\infty} b(n) \frac{(t-t_1)^n}{n!}, \quad |t - t_1| < T - |t_1 - t_0|,$$

where each $b(n)$ is given by an absolutely convergent series

$$b(n) = \sum_{k=0}^{\infty} c(n+k) \frac{(t_1 - t_0)^k}{k!}, \quad n \geq 0 \tag{1.5}$$

(see Problem 1.1.4). So once a power series representation of u is identified at one point, any other power series representation at another point within the radius of convergence can be explicitly computed using equation (1.5). Modulo this type of transformation, it makes sense to speak of *the* coefficients of u in a neighborhood of t_0.

A fundamental idea throughout this book is that within the radius of convergence, the series coefficients of u completely characterize the function u. Therefore, operating on the coefficients in some manner produces a corresponding transformation of the function and vice versa. Consider the following examples.

Example 1.1 As discussed earlier, if u is real analytic at t_0, then it is differentiable on a neighborhood of t_0. Observe that

$$\begin{aligned} u'(t) &= \sum_{n=1}^{\infty} c(n)\frac{(t-t_0)^{n-1}}{(n-1)!} \\ &= \sum_{n=0}^{\infty} c(n+1)\frac{(t-t_0)^{n}}{n!} \\ &=: \sum_{n=0}^{\infty} d(n)\frac{(t-t_0)^{n}}{n!}, \quad |t-t_0|<T \end{aligned} \tag{1.6}$$

(see Problem 1.1.5). Therefore, the left-shift mapping between two sequences

$$x^{-1} : c \mapsto d, \tag{1.7}$$

where $d(n) = c(n+1)$, $n \geq 0$ corresponds to mapping u to u'. □

Example 1.2 Consider taking the square of u. Clearly,

$$\begin{aligned} u^2(t) &= \sum_{n=0}^{\infty} c(n)\frac{(t-t_0)^{n}}{n!}\sum_{k=0}^{\infty} c(k)\frac{(t-t_0)^{k}}{k!} \\ &= \sum_{n=0}^{\infty}\left(\frac{c(0)}{0!}\frac{c(n)}{n!} + \frac{c(1)}{1!}\frac{c(n-1)}{(n-1)!} + \cdots + \right. \\ &\qquad\quad \left.\frac{c(n)}{n!}\frac{c(0)}{0!}\right)(t-t_0)^n \\ &=: \sum_{n=0}^{\infty} d(n)\frac{(t-t_0)^{n}}{n!}, \quad |t-t_0|<T. \end{aligned}$$

Thus, the mapping between sequences

$$(\cdot)^2 : c \mapsto d,$$

where $d(n) := \sum_{k=0}^{n} \binom{n}{k} c(k)c(n-k)$, $n \geq 0$ corresponds to mapping u to u^2.

□

When u has coefficients which satisfy (1.4), u is entire, and its right-sided Laplace transform is well defined (cf. Problem 1.1.6). Specifically,

$$\begin{aligned}\mathscr{L}[u](s) &:= \int_0^\infty u(t)e^{-st}\,dt \\ &= \sum_{n=0}^{\infty} c(n) \int_0^\infty \frac{t^n}{n!}\, e^{-st}\,dt \\ &= s^{-1} \sum_{n=0}^{\infty} c(n)(s^{-1})^n, \end{aligned} \tag{1.8}$$

provided $\mathrm{Re}(s) > 0$. In which case,

$$\begin{aligned}|\mathscr{L}[u](s)| &\leq K \left|s^{-1}\right| \sum_{n=0}^{\infty} (M \left|s^{-1}\right|)^n \\ &= \frac{K\left|s^{-1}\right|}{1 - M\left|s^{-1}\right|}\end{aligned}$$

whenever $\left|s^{-1}\right| < 1/M =: S$. The Laplace transform of u, when written as a power series in s^{-1}, is described by the same sequence, c, as is its counterpart u in (1.1) modulo the factors $1/n!$ and the extra factor of s^{-1} corresponding to a right-shift of c. It is often convenient to introduce an abstract symbol, x, and to write the sequence c as a formal power series

$$c = \sum_{n=0}^{\infty} c(n)\, x^n.$$

The word *formal* in this case refers to the fact no actual summation of the series terms is considered, so convergence is not an issue. In this context, no notational distinction is usually made between the *sequence* c and the *series* c. The mapping

$$\mathscr{L}_f\ : u \mapsto c,$$

assuming $t_0 = 0$ in equation (1.1), is called the *formal Laplace transform* (see Table 1.1 and Problem 1.1.7). It is well defined whether or

Table 1.1. The formal Laplace transform of some common functions.

$u(t),\ t \geq 0$	$\mathscr{L}_f[u]$
$t^k/k!,\ k \geq 0$	x^k
$t^k e^{at}/k!,\ k \geq 0,\ a \neq 0$	$\sum_{n=0}^{\infty} \binom{n+k}{k} a^n x^{n+k}$
$\sin(bt)$	$\sum_{n=0}^{\infty} (-1)^n b^{2n+1} x^{2n+1}$
$\cos(bt)$	$\sum_{n=0}^{\infty} (-1)^n b^{2n} x^{2n}$
$\sinh(bt)$	$\sum_{n=0}^{\infty} b^{2n+1} x^{2n+1}$
$\cosh(bt)$	$\sum_{n=0}^{\infty} b^{2n} x^{2n}$

not the series (1.8) converges for any value of s^{-1}. But when it does, then clearly

$$\mathscr{L}[u](s) = x\, \mathscr{L}_f[u]\big|_{x \to s^{-1}}$$

everywhere in the region of convergence for $\mathscr{L}[u]$ where the series converges and where $\mathrm{Re}(s) > 0$.

One approach to characterizing the radius of convergence of (1.8) involves forming the *Hankel matrix* of c, namely,

$$\mathcal{H}_c = \begin{bmatrix} c(0) & c(1) & c(2) & \cdots \\ c(1) & c(2) & c(3) & \cdots \\ c(2) & c(3) & c(4) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

In the event that $\mathcal{H}_c$ has finite rank $n > 0$, it follows that the first $n+1$ columns of the matrix must be linearly dependent. That is, there exists a polynomial $\tilde{q} = \sum_{\ell=0}^{n} \tilde{q}(\ell) x^\ell$ with at least one coefficient $\tilde{q}(\ell) \neq 0$ such that

$$\begin{bmatrix} c(0) & c(1) & \cdots & c(n) & \cdots \\ c(1) & c(2) & \cdots & c(n+1) & \cdots \\ c(2) & c(3) & \cdots & c(n+2) & \cdots \\ \vdots & \vdots & & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \tilde{q}(0) \\ \tilde{q}(1) \\ \vdots \\ \tilde{q}(n) \\ 0 \\ 0 \\ \vdots \end{bmatrix} = 0,$$

or equivalently,

$$\sum_{\ell=0}^{n} c(k+\ell)\tilde{q}(\ell) = 0, \quad k \geq 0. \tag{1.9}$$

Exploiting the Hankel structure, it is possible to show more specifically that the first n columns of $\mathcal{H}_c$ are linearly independent, and thus, $\tilde{q}(n) \neq 0$ (see Problem 1.1.8). Now let q be the polynomial derived from $\tilde{q}$ by simply reversing the order of its coefficients, i.e., $q(i) = \tilde{q}(n-i)$, $i = 0, 1, \ldots, n$. Observe that

$$\begin{aligned} cq &= \sum_{i=0}^{\infty} \sum_{j=0}^{n} c(i)q(j)x^{i+j} \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{n} c(i-j)q(j)x^{i} \\ &= \sum_{i=0}^{\infty} p(i)x^{i} \\ &= p, \end{aligned} \tag{1.10}$$

assuming that $c(i) = 0$ when $i < 0$ and defining

$$\begin{aligned} p(i) &= \sum_{j=0}^{n} c(i-j)q(j) \\ &= \sum_{j=0}^{n} c(i-j)\tilde{q}(n-j). \end{aligned}$$

Note that equation (1.9) implies for $k \geq 0$ that

$$\begin{aligned} p(n+k) &= \sum_{j=0}^{n} c(n+k-j)\tilde{q}(n-j) \\ &= \sum_{\ell=0}^{n} c(k+\ell)\tilde{q}(\ell) \\ &= 0. \end{aligned}$$

Thus, p can be viewed as the image of $\tilde{q}$ under an augmented Hankel matrix, namely,

$$\left[\begin{array}{ccccc} 0 & 0 & \cdots & c(0) & \cdots \\ 0 & 0 & \cdots & c(1) & \cdots \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & c(0) & \cdots & c(n-1) & \cdots \\ \hline c(0) & c(1) & \cdots & c(n) & \cdots \\ c(1) & c(2) & \cdots & c(n+1) & \cdots \\ c(2) & c(3) & \cdots & c(n+2) & \cdots \\ \vdots & \vdots & & \vdots & \ddots \end{array}\right] \left[\begin{array}{c} \tilde{q}(0) \\ \tilde{q}(1) \\ \vdots \\ \tilde{q}(n) \\ 0 \\ 0 \\ \vdots \end{array}\right] = \left[\begin{array}{c} p(0) \\ p(1) \\ \vdots \\ p(n-1) \\ \hline 0 \\ 0 \\ \vdots \end{array}\right]. \tag{1.11}$$

In which case, $p = cq$ is a polynomial in x of at most degree $n-1$. One could formally write then that $c = pq^{-1}$, where q^{-1} is understood to be a formal power series with the property $q^{-1}q = qq^{-1} = 1$. In this situation, c is called a *rational* series. Readers familiar with linear system theory will recognize this as the formal counterpart to the relationship between a rational transfer function and the system Hankel matrix constructed from the series coefficients (*Markov parameters*) of the transfer function when written as a power series in the variable s^{-1}. But in the present context, the corresponding Laplace transform of u is a rational function in s^{-1}. Specifically, $\mathscr{L}[u](s) = s^{-1}N(s^{-1})/D(s^{-1})$, where

$$N(s^{-1}) = p|_{x\rightarrow s^{-1}}, \quad \deg(N) \leq n-1 \tag{1.12}$$

$$D(s^{-1}) = q|_{x\rightarrow s^{-1}}, \quad \deg(D) \leq n. \tag{1.13}$$

It can be shown that N and D have no common roots as polynomials in s^{-1}, that is, the rational function N/D is *irreducible* (see Problem 1.1.9). Therefore, the radius of convergence of the series representation of $\mathscr{L}[u]$ at the origin is $S = \min_i |\lambda_i|$, where λ_i is the i-th root of the polynomial D. When this analysis is combined with the requirement that $\mathrm{Re}(s) > 0$, which holds if and only if $\mathrm{Re}(s^{-1}) > 0$, the resulting region of convergence is shown in Figure 1.1.

Example 1.3 If $c = \sum_{i\geq 1} 2^{i-1}x^i$ then the corresponding Hankel matrix is

$$\mathcal{H}_c = \left[\begin{array}{ccccc} 0 & 1 & 2 & 4 & \cdots \\ 1 & 2 & 4 & 8 & \cdots \\ 2 & 4 & 8 & 16 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array}\right],$$

which clearly has rank two. Thus, there exists polynomials p and q such that $c = pq^{-1}$ with $\deg(p) \leq 1$ and $\deg(q) \leq 2$. From (1.11),

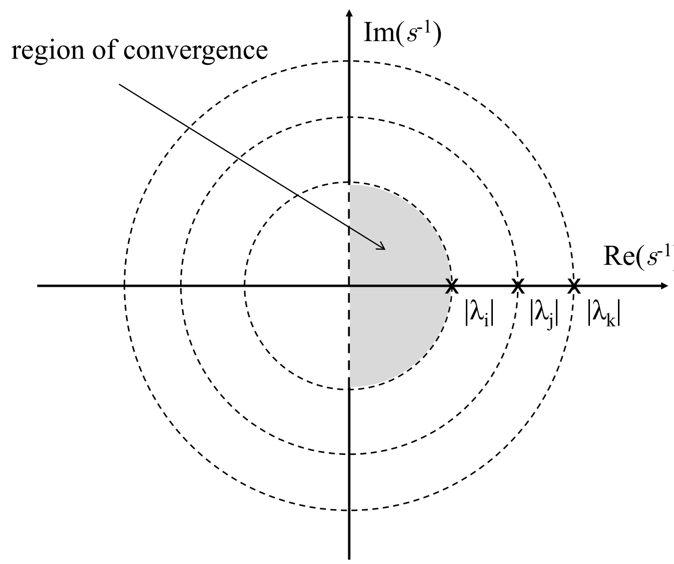


Fig. 1.1. The region of convergence for the series representation of $\mathscr{L}[u]$.

observe that

$$\left[\begin{array}{ccccc} 0 & 0 & 0 & 1 & \cdots \\ 0 & 0 & 1 & 2 & \cdots \\ \hline 0 & 1 & 2 & 4 & \cdots \\ 1 & 2 & 4 & 8 & \cdots \\ 2 & 4 & 8 & 16 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array}\right] \left[\begin{array}{c} 0 \\ -2 \\ 1 \\ 0 \\ 0 \\ \vdots \end{array}\right] = \left[\begin{array}{c} 0 \\ 1 \\ \hline 0 \\ 0 \\ \vdots \end{array}\right].$$

Therefore, $\tilde{q} = x^2 - 2x$ or $q = 1 - 2x$ and $p = x$. To confirm this calculation, note that

$$c = pq^{-1} = \frac{x}{1-2x} = x\sum_{i=0}^{\infty} 2^i x^i = \sum_{i=1}^{\infty} 2^{i-1} x^i.$$

The Laplace transform of the corresponding input

$$u(t) = \sum_{i=1}^{\infty} 2^{i-1}\frac{t^i}{i!} = \frac{1}{2}\left(\mathrm{e}^{2t} - 1\right), \quad t \geq 0$$

is

$$\mathscr{L}[u](s) = s^{-1}\sum_{i=1}^{\infty} 2^{i-1}s^{-i} = \frac{s^{-2}}{1-2s^{-1}}.$$

The radius of convergence for the infinite sum is clearly defined by $\left|2s^{-1}\right| < 1$ or $\left|s^{-1}\right| < 1/2$. This corresponds to the location of the pole of $D(s^{-1}) = 1 - 2s^{-1}$ at $s^{-1} = 1/2$. Of course in system theory it is more customary to write rational functions in terms of s rather than s^{-1}, but the latter is actually more natural when doing series analysis.

□

The notion of real analyticity can be extended to multivariable functions in the following manner. A function $f : U \subset \mathbb{R}^k \to \mathbb{R}^\ell$ is said to be *real analytic* at a point $z_0 = (z_{1,0}, \ldots, z_{k,0}) \in \mathbb{R}^k$ if it has a convergent power series representation

$$f(z_1,\ldots,z_k) = \sum_{n_1,\ldots,n_k=0}^{\infty} c(n_1,\ldots,n_k)\frac{(z_1 - z_{1,0})^{n_1}}{n_1!}\cdots\frac{(z_k - z_{k,0})^{n_k}}{n_k!}$$

on some open neighborhood $V \subset U$ of z_0, where each $c(n_1,\ldots,n_k) \in \mathbb{R}^\ell$. Extending the setup to the complex variable setting and applying the corresponding version of the Cauchy integral formula gives an expression for the coefficients

$$\begin{aligned} &c(n_1,\ldots,n_k) \\ &= \frac{n_1!\cdots n_k!}{(2\pi i)^n}\oint_{C_1}\frac{1}{(z_1 - z_{1,0})^{n_1+1}}\cdots\oint_{C_k}\frac{f(z_1,\ldots,z_k)}{(z_k - z_{k,0})^{n_k+1}}\,dz_k\cdots dz_1, \end{aligned}$$

where each integral is defined componentwise, $n := n_1 + n_2 \cdots + n_k$, and $|z_i - z_{i,0}| < R_i$ for $i = 1, 2, \ldots, k$. Define the growth constants

$$\begin{aligned} K &= \max_i \max_{|z_i - z_{i,0}|<R_i} |f(z_1,\ldots,z_k)| \\ M_i &= \frac{1}{R_i} \end{aligned}$$

assuming $|y| := \max\{|y_1|,\ldots,|y_\ell|\}$. Then it follows directly that

$$\begin{aligned} |c(n_1,\ldots,n_k)| &\leq KM_1^{n_1}\cdots M_k^{n_k}n_1!\cdots n_k! &(1.14)\\ &\leq KM^n n_1!\cdots n_k! &(1.15)\\ &\leq KM^n n! &(1.16) \end{aligned}$$

with $M := \max_{1\leq i\leq k} M_i$ and using the multinomial coefficients property

$$(n_1, n_2, \ldots, n_k)! := \frac{n!}{n_1!\, n_2! \cdots n_k!} \geq 1.$$

Clearly, the inequality (1.14) is the multivariable extension of the Cauchy growth condition (1.3). Sometimes in calculations, however, it is easier to use the more generous upper bounds (1.15) or (1.16).

1.2 Formal Functions

A function $f : U \subset \mathbb{R}^m \to \mathbb{R}^\ell$ is said to be *smooth* at a point $z_0 \in U$ if its partial derivatives

$$\left.\frac{\partial^k f(z)}{\partial z_{i_1} \partial z_{i_2} \cdots \partial z_{i_k}}\right|_{z=z_0}$$

are well defined for every $i_j \in \{1, \ldots, m\}$, $k = 1, 2, \ldots$. The set of all such functions will be denoted by $C^\infty(z_0)$. Two functions $f, g \in C^\infty(z_0)$ will be called $\sim_g$ *equivalent* if there exists an open neighborhood U_0 of z_0 in the domain of each function such that $f = g$ on U_0.[1] It is easily verified that this is an equivalence relation on $C^\infty(z_0)$. The quotient set, that is, the set of all equivalence classes, is represented by $C^\infty(z_0)/\!\sim_g$.

Definition 1.1 *A* ***germ*** *at z_0 is an equivalence class in $C^\infty(z_0)/\sim_g$.*

Two functions in $f, g \in C^\infty(z_0)$ will be called $\sim_j$ *equivalent* when

$$\left.\frac{\partial^k f(z)}{\partial z_{i_1} \partial z_{i_2} \cdots \partial z_{i_k}}\right|_{z=z_0} = \left.\frac{\partial^k g(z)}{\partial z_{i_1} \partial z_{i_2} \cdots \partial z_{i_k}}\right|_{z=z_0}$$

for all $i_j \in \{1, \ldots, m\}$, $k = 0, 1, \ldots$. It is equally straightforward to show that $\sim_j$ also defines an equivalence relation on $C^\infty(z_0)$, and the corresponding quotient set is denoted by $C^\infty(z_0)/\!\sim_j$.

Definition 1.2 *An* ***infinite jet*** *at z_0 is an equivalence class in $C^\infty(z_0)/\!\sim_j$.*

Clearly $f \sim_g g$ implies that $f \sim_j g$, but the following example illustrates that the converse of this implication is false.

[1] The reader may want to review the material in Appendix Section A.2

Example 1.4 Consider the functions $f(z) = 0$ for all $z \in \mathbb{R}$ and

$$g(z) = \begin{cases} e^{-1/z^2} & : z \neq 0 \\ 0 & : z = 0. \end{cases}$$

Both f and g are in $C^\infty(0)$, and it can be directly checked that $f \sim_j g$ while $f \not\sim_g g$. That is, an infinite jet is not a *faithful* representation of a germ. Put another way, not all functions are equal to their Taylor series in a region of convergence with nonzero radius unless they are real analytic at the point in question. In this case, f is real analytic at $z_0 = 0$, while g is not (see Problem 1.2.1). Note in particular that $\left|g^{(n)}(0)\right| = 0 < KM^n n!$, $n = 0, 1, \ldots$ for any $K, M > 0$. So satisfying the Cauchy growth condition alone is *not* sufficient for equating a function with its Taylor series, that is,

$$g(z) \neq \sum_{n=0}^{\infty} g^{(n)}(0)\frac{z^n}{n!}$$

at every point except $z = 0$. □

In the previous section, it was shown that given a function $f : U \subset \mathbb{R} \to \mathbb{R}$ which is real analytic at a point, say $z_0 = 0$, one can uniquely identify a formal power series $c = \mathscr{L}_f[f]$ which satisfies a Cauchy growth condition. The above example, however, illustrates that the reverse process of mapping a series to a smooth function is not as simple as one might first surmise. Without the additional analyticity assumption, such a series can only represent a *class* of functions. This motivates the following definition.

Definition 1.3 *A* ***formal function*** *at z_0 is a class of functions described by an infinite jet at z_0.*

A convenient way to describe the set of all formal functions at a point, namely $C^\infty(z_0)/\sim_j$, is to identify every infinite jet with a formal power series defined over a set of symbols $X = \{x_1, x_2, \ldots, x_m\}$, usually called an *alphabet*, where each *letter* x_i represents an argument z_i of f. Specifically, define for any *word* $x_{i_1}x_{i_2}\cdots x_{i_k}$ the corresponding ℓ-vector

$$\left.\frac{\partial^k f(z)}{\partial z_{i_1}\partial z_{i_2}\cdots\partial z_{i_k}}\right|_{z=z_0}. \tag{1.17}$$

Since the order of partial differentiation is not important, it is natural to allow the letters of X to commute, i.e., $x_i x_j = x_j x_i$. In which case, an ordering $x_1 < x_2 < \cdots < x_m$ can be introduced and a given infinite jet can be uniquely identified with the formal power series

$$c = \sum_{n_1,n_2,\ldots,n_m \geq 0} c(n_1,n_2,\ldots,n_m)\, x_1^{n_1} x_2^{n_2} \cdots x_m^{n_m},$$

where $x_i^{n_i} := x_i x_i \cdots x_i$ (x_i appears n_i times). If $\mathbb{R}^\ell\,[[X]]$ denotes the set of all possible c, then clearly $C^\infty(z_0)/\sim_j$ can be identified with at least a subset of $\mathbb{R}^\ell\,[[X]]$. But it turns out that one can say more than this. Consider, for example, the special case where $m = 1$ so that $f : U \subset \mathbb{R} \to \mathbb{R}^\ell$. and $X = \{x\}$. It can be shown that the formal Laplace transform

$$\begin{aligned}\mathscr{L}_f : C^\infty(z_0) &\to \mathbb{R}^\ell\,[[X]] \\ : f &\mapsto c,\end{aligned}$$

where the coefficients of c correspond to the partial derivatives of f, is a surjective mapping.[2] Hence, for the series $c = \sum_{n\geq 0}(n!)^2 x^n$ there exists *at least one* function in $C^\infty(z_0)$ which is well defined on a neighborhood of z_0 and whose derivatives grow at this non Cauchy rate. If $\mathscr{L}_f$ is restricted to real analytic functions then the coefficients of c, as discussed in the previous section, must satisfy a Cauchy growth rate. But in general, the image of $\mathscr{L}_f$ is all of $\mathbb{R}^\ell\,[[X]]$. In this context, a *formal Borel transform* is any mapping of the form

$$\mathscr{B}_f : \mathbb{R}^\ell\,[[X]] \to C^\infty(z_0)$$

such that $\mathscr{L}_f\,\mathscr{B}_f(c) = c$ for all $c \in \mathbb{R}^\ell\,[[X]]$. That is, $\mathscr{B}_f$ is any right inverse of $\mathscr{L}_f$. It constitutes a left inverse, namely, $\mathscr{B}_f\,\mathscr{L}_f(f) = f$, only in special cases, for example, when f is real analytic, and $\mathscr{B}_f$ maps all such series to the function defined by its convergent Taylor series. It is more common to define the formal Laplace-Borel transform pair as

$$\begin{aligned}\mathscr{L}_f : \sum_{n=0}^{\infty} c(n)\frac{z^n}{n!} &\mapsto \sum_{n=0}^{\infty} c(n)x^n \\ \mathscr{B}_f : \sum_{n=0}^{\infty} c(n)x^n &\mapsto \sum_{n=0}^{\infty} c(n)\frac{z^n}{n!}.\end{aligned}$$

Their interpretation is strictly in the formal sense when the series involved do not converge, otherwise, they have the usual interpretation

[2] This is called Borel's Lemma.

as mappings between functions. In either case, $\sum_{n\geq 0} c(n)x^n$ is called the *generating function* associated with c, while $\sum_{n\geq 0} c(n)z^n/n!$ is referred to as its *exponential generating function*.

1.3 Linear Integral Operators

Since real analytic functions are always absolutely integrable within their region of convergence, they provide a convenient class of kernel functions for integral operators. Consider a causal linear integral operator

$$y(t) = \sum_{i=1}^{m} \int_{t_0}^{t} H_i(t,\tau)u_i(\tau)\, d\tau, \tag{1.18}$$

where $y(t) \in \mathbb{R}^\ell$, $t \geq t_0$ and every u_i is piecewise continuous. If each kernel function $H_i : \mathbb{R}^2 \to \mathbb{R}^\ell$ is real analytic at (t_0, t_0), then there exists at least a finite $T > 0$ such that on the set $\mathcal{D} = \{(t,\tau) \in \mathbb{R}^2 : t_0 + T \geq t \geq \tau \geq t_0\}$, $H_i(t,\tau)$ can be expressed as a uniformly convergent series

$$H_i(t,\tau) = \sum_{n_0,n_1=0}^{\infty} c(n_1, i, n_0)\, \frac{(t-\tau)^{n_1}}{n_1!} \frac{(\tau - t_0)^{n_0}}{n_0!}. \tag{1.19}$$

Here the coefficients have been indexed in manner that is more consistent with the way in which a formal power series will be used to describe an integral operator. Causality mandates that the kernel functions be identically zero when $t < \tau$, so these series representations are only used on $\mathcal{D}$. Substituting equation (1.19) into equation (1.18) and integrating term-by-term, it follows that

$$y(t) = \sum_{n_0,n_1=0}^{\infty} \sum_{i=1}^{m} c(n_1, i, n_0) \int_{t_0}^{t} \frac{(t-\tau)^{n_1}}{n_1!} u_i(\tau) \frac{(\tau - t_0)^{n_0}}{n_0!}\, d\tau. \tag{1.20}$$

Now introduce a letter x_i for each input u_i, and define also a fictitious input function $u_0 \equiv 1$ and corresponding letter x_0. For each x_i define an associated integral

$$E_{x_i}[u](t, t_0) = \int_{t_0}^{t} u_i(\tau)\, d\tau.$$

In which case,

$$E_{x_0}[u](t,t_0) = t - t_0,$$

and if the integration is repeated

$$E_{x_0^2}[u](t,t_0) := \int_{t_0}^{t} u_0(\tau) E_{x_0}[u](\tau,t_0)\, d\tau = \frac{(t-t_0)^2}{2!}.$$

After n_0 successive integrations,

$$E_{x_0^{n_0}}[u](t,t_0) = \frac{(t-t_0)^{n_0}}{n_0!}.$$

Similarly, for any $i = 1,\dots,m$, let

$$\begin{aligned} E_{x_i x_0^{n_0}}[u](t,t_0) &= \int_{t_0}^{t} u_i(\tau) E_{x_0^{n_0}}[u](\tau,t_0)\, d\tau \\ &= \int_{t_0}^{t} u_i(\tau) \frac{(\tau-t_0)^{n_0}}{n_0!}\, d\tau. \end{aligned} \tag{1.21}$$

This last integral is close in form to those that appear in the series (1.20) except for the terms involving n_1. So integrate the expression (1.21) once more and apply the integration by parts formula:

$$\begin{aligned} E_{x_0 x_i x_0^{n_0}}[u](t,t_0) &= \int_{t_0}^{t}\int_{t_0}^{\tau} u_i(\xi)\frac{(\xi-t_0)^{n_0}}{n_0!}\, d\xi\, d\tau \\ &= \int_{t_0}^{\tau} u_i(\xi)\frac{(\xi-t_0)^{n_0}}{n_0!}\, d\xi\, (\tau-t_0)\Big|_{t_0}^{t} - \\ &\qquad \int_{t_0}^{t} (\tau-t_0)\, u_i(\tau)\frac{(\tau-t_0)^{n_0}}{n_0!}\, d\tau \\ &= \int_{t_0}^{t} (t-\tau)\, u_i(\tau)\frac{(\tau-t_0)^{n_0}}{n_0!}\, d\tau. \end{aligned} \tag{1.22}$$

A straightforward induction gives

$$E_{x_0^{n_1} x_i x_0^{n_0}}[u](t,t_0) = \int_{t_0}^{t} \frac{(t-\tau)^{n_1}}{n_1!}\, u_i(\tau)\frac{(\tau-t_0)^{n_0}}{n_0!}\, d\tau \tag{1.23}$$

(see Problem 1.3.1). Thus, series (1.20) has the alternative expression in terms of *iterated integrals*

$$y(t) = \sum_{n_0,n_1=0}^{\infty}\sum_{i=1}^{m} c(n_1,i,n_0)\, E_{x_0^{n_1} x_i x_0^{n_0}}[u](t,t_0). \tag{1.24}$$

The form of this series suggests that indexing the summations in terms of words over the alphabet $X = \{x_0, x_1, \ldots, x_m\}$ would be more natural. So the following notation is introduced:

$$(c, \eta) = \begin{cases} c(n_1, i, n_0) : \eta = x_0^{n_1} x_i x_0^{n_0} \\ \quad 0 \quad\quad\quad : \text{otherwise} \end{cases}$$

for all $n_0, n_1 \geq 0$ and $i = 1, \ldots, m$. In which case, the series (1.24) has the concise representation

$$y(t) = \sum_{\eta \in X^*} (c, \eta) E_\eta[u](t, t_0), \tag{1.25}$$

where X^* is the set of all words over X (including the empty word $\emptyset$) and $E_\emptyset := 1$. A key observation in this context is that the letters of X do *not* commute, since, for example, the integrals $E_{x_0 x_1}$ and $E_{x_1 x_0}$ are not equivalent. Thus, $c = \sum_{\eta \in X^*} (c, \eta)\, \eta$ must be viewed as a noncommutative formal power series. It will be referred to as the *generating series* for this integral operator. The symbol $\mathbb{R}^\ell\langle\langle X \rangle\rangle$ will denote the set of all possible noncommutative formal power series over X taking coefficients from $\mathbb{R}^\ell$. Note that any series $c \in \mathbb{R}^\ell\langle\langle X \rangle\rangle$ can be equivalently described as a mapping of the form

$$\begin{aligned} c : X^* &\to \mathbb{R}^\ell \\ : \eta &\mapsto (c, \eta). \end{aligned}$$

In principle, one could define an input-output operator, $F_c : u \mapsto y$, using the expression (1.25) for any $c \in \mathbb{R}^\ell\langle\langle X \rangle\rangle$ provided that the series converges for every u in the set of admissible inputs. Such operators are called *Fliess operators*, and the series is known as the *Chen-Fliess functional expansion*. From an historical point of view, Fliess operators can be viewed as a special class of *Volterra operators*. An operator V is called a Volterra operator if it can be represented by a convergent series

$$\begin{aligned} y(t) &= V[u](t) \\ &= w_0(t) + \sum_{k=1}^{\infty} \sum_{i_1,\ldots,i_k=1}^{m} \int_{t_0}^{t} \int_{t_0}^{\tau_k} \cdots \int_{t_0}^{\tau_2} w_{i_k \cdots i_1}(t, \tau_k, \ldots, \tau_1) \cdot \\ &\qquad u_{i_k}(\tau_k) \cdots u_{i_1}(\tau_1)\, d\tau_1 \cdots d\tau_k, \end{aligned}$$

where each kernel function, $w_{i_k\cdots i_1}$, is an ℓ vector-valued function defined on a set

$$\mathcal{D}_k = \{(t, \tau_k, \ldots, \tau_1) \in \mathbb{R}^{k+1} : t_0 + T \geq t \geq \tau_k \geq \cdots \geq \tau_1 \geq t_0\}.$$

Each integral in this series can be viewed as a generalized convolution integral, and a *finite* Volterra operator refers to the case where only a finite number of the kernel functions are nonzero. Volterra operators date back to the 1880s and are arguably the most widely encountered type of nonlinear operators encountered in physics and engineering. Observe that the linear operator in (1.18) is among the simplest examples of a Volterra operator, and it was rewritten in (1.25) as a Fliess operator. It will be shown in Chapter 3 that any Volterra operator having real analytic kernels has a Fliess operator representation over some admissible set of inputs.

In the theory of linear systems, the series coefficients of a real analytic kernel function can be used to determine the nature of the operator. For example, in the linear time-invariant case, each kernel function $H_i(t, \tau)$ in (1.19) reduces to the form

$$H_i(t-\tau) = \sum_{k=0}^{\infty} (c, x_0^k x_i) \frac{(t-\tau)^k}{k!}$$

with the corresponding Fliess operator F_c given by

$$\begin{aligned} y(t) &= \sum_{i=1}^{m} \sum_{k=0}^{\infty} (c, x_0^k x_i) \int_{t_0}^{t} \frac{(t-\tau)^k}{k!} u_i(\tau)\, d\tau \\ &= \sum_{i=1}^{m} \sum_{k=0}^{\infty} (c, x_0^k x_i) E_{x_0^k x_i}[u](t, t_0). \end{aligned} \tag{1.26}$$

Suppose the output is scalar-valued. For each $i = 1, 2, \ldots, m$ define the formal power series $c_i = \sum_{k=0}^{\infty} (c, x_0^k x_i) x_0^k x_i$. If every Hankel matrix

$$\mathcal{H}_{c_i} = \begin{bmatrix} (c, x_i) & (c, x_0 x_i) & (c, x_0^2 x_i) & \cdots \\ (c, x_0 x_i) & (c, x_0^2 x_i) & (c, x_0^3 x_i) & \cdots \\ (c, x_0^2 x_i) & (c, x_0^3 x_i) & (c, x_0^4 x_i) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{1.27}$$

has finite rank then every c_i is rational in the sense that there exist polynomials in x_0, say a_i and b_i, such that $c_i = (b_i a_i^{-1}) x_i$ (see equation (1.10)). In which case, the power series c can be written as

$$c=\sum_{i=1}^{m} c_i=\sum_{i=1}^{m}(b_i a_i^{-1})x_i. \tag{1.28}$$

In general, any formal power series over X is said to be *rational* if it can be written in terms of a finite number of polynomials using a finite number of sums, products and inversions. Series (1.28) is a special case where the polynomials in the letters $\{x_1,x_2,\ldots,x_m\}$ are all homogeneous with degree one. In Chapter 4, this idea will be more fully developed. It will be shown in the next section that systems having rational generating series of the form (1.28) always have finite dimensional state space realizations.

The formal Laplace transform of the linear Fliess operator $F_c: u\mapsto y$ is defined to be

$$\mathscr{L}_f : F_c \mapsto c,$$

assuming that c can be uniquely determined given F_c. In the linear time-invariant case observe that the formal Laplace transforms of the input and output functions, $c_u=\sum_{k\geq 0}(c_u,x_0^k)x_0^k$ and $c_y=\sum_{k\geq 0}(c_y,x_0^k)x_0^k$, respectively, are related by

$$\begin{aligned} c_y &= c\circ c_u \\ &:= c|_{x_i\to x_0 c_{u_i}} \\ &= \sum_{i=1}^{m}\sum_{j=0}^{\infty}(c,x_0^j x_i)x_0^{j+1}c_{u_i}, \end{aligned}$$

or equivalently,

$$(c\circ c_u,x_0^k):=\sum_{i=1}^{m}\sum_{j=0}^{k-1}(c,x_0^{k-j-1}x_i)(c_{u_i},x_0^j),\quad k\geq 1 \tag{1.29}$$

(see Problem 1.3.3). The summation on the right-hand side with respect to j is a convolution sum, as is expected from linear system theory. The impulse response of the operator relative to the input u_i corresponds to

$$h_i:=c_i|_{x_i\to 1}=\sum_{j=0}^{\infty}(c,x_0^j x_i)x_0^j,\quad i=1,2,\ldots,m. \tag{1.30}$$

The formal power series product '$\circ$' is a special case of the *composition product*. In general, it gives the generating series for the composition of two Fliess operators, that is, for arbitrary series $c \in \mathbb{R}^\ell\langle\langle X\rangle\rangle$ and $d \in \mathbb{R}^m\langle\langle X\rangle\rangle$ $F_c \circ F_d = F_{c\circ d}$. So from a systems point of view, it describes the series interconnection of two input-output systems. This product will be first introduced in Chapter 2 and then further developed in Chapter 3. In particular, it will be shown that the composition product is an example of what is called *Hopf convolution*, a well known product in the theory of Hopf algebras. This machinery will give not only deeper insight into the algebra, but also provide convenient computational tools for solving real problems.

Up to this point, all the discussion has been for analytic operators. But a *formal* Fliess operator can also be defined in the event that the underlying series describing F_c does not converge, i.e., when generating series c does not satisfy a Cauchy growth condition. In this case, the mapping $c_u \mapsto c_y = c \circ c_u$ takes formal inputs to formal outputs with no underlying assumption of convergence. This mapping is viewed as the formal counterpart of the mapping $u \mapsto y = F_c[u]$ in the real analytic case. In some analysis encountered in later chapters, it will be easier to first establish algebraic results via the use of formal functions, formal operators, or formal differential equations, after which convergence issues can be determined independently. In other instances, such as in this section, it is more intuitive to start with the analytic case, and then extract out the algebraic structures on which the formal counterparts are based.

1.4 State Space Realizations of Rational Operators

An operator F_c is said to be *rational* whenever its generating series c is rational. As an example, consider the case where c is given by equation (1.28). It is straightforward to show that the generating series for the output function,

$$c_y = \sum_{i=1}^{m} (b_i a_i^{-1}) x_0 c_{u_i},$$

is rational whenever c_u is rational, i.e., when each c_{u_i} is rational (see Problem 1.3.3). A pillar of linear system theory is that the rational operator F_c always has a finite dimensional state space *realization* of the form

$$\dot{z}(t) = Az(t) + \sum_{i=1}^{m} B_i u_i(t), \quad z(t_0) = 0 \tag{1.31}$$

$$y(t) = Cz(t), \tag{1.32}$$

where $A \in \mathbb{R}^{n\times n}$, $B_i \in \mathbb{R}^{n\times 1}$ and $C \in \mathbb{R}^{1\times n}$, and whose solution $\phi(t,t_0,0,u)$ satisfies

$$y(t) = F_c[u](t) = C\phi(t,t_0,0,u)$$

for every integrable input u. Setting $m=1$ and integrating both sides of the state equation (1.31) gives

$$z(t) = \int_{t_0}^{t} Az(\tau)\,d\tau + \int_{t_0}^{t} B_1u_1(\tau)\,d\tau.$$

Substituting for $z(\tau)$ on the right-hand side with the entire expression gives

$$\begin{aligned} z(t) &= \int_{t_0}^{t} A\left[\int_{t_0}^{\tau_2} Az(\tau_1)\,d\tau_1 + \int_{t_0}^{\tau_2} B_1u_1(\tau_1)\,d\tau_1\right]d\tau_2 + \int_{t_0}^{t} B_1u_1(\tau)\,d\tau \\ &= A^2\int_{t_0}^{t}\int_{t_0}^{\tau_2} z(\tau_1)\,d\tau_1 d\tau_2 + AB_1\int_{t_0}^{t}\int_{t_0}^{\tau_2} u(\tau_1)\,d\tau_1 d\tau_2 + \\ &\qquad B_1\int_{t_0}^{t} u_1(\tau)\,d\tau. \end{aligned}$$

Continuing in this way gives what is called the *Peano-Baker series* representation of the solution to the state equation

$$z(t) = \sum_{k=0}^{\infty} A^kB_1\int_{t_0}^{t}\int_{t_0}^{\tau_{k+1}}\cdots\int_{t_0}^{\tau_2} u_1(\tau_1)\,d\tau_1 d\tau_2\cdots d\tau_{k+1}.$$

From the output equation (1.32) it then follows that

$$y(t) = \sum_{k=0}^{\infty} CA^kB_1E_{x_0^kx_1}[u](t),$$

or equivalently,

$$y(t) = \sum_{\eta\in X^*} (c,\eta)E_\eta[u](t,t_0),$$

where

$$(c,\eta) = \begin{cases} CA^kB_1 : \eta = x_0^kx_1, \quad k\geq 0 \\ \quad 0 \quad\;\; : \text{otherwise.}\end{cases}$$

The triple (A, B_1, C) is said to be a *representation* of the generating series c. The distinction between realizing F_c and representing c is minor in the present context. It is straightforward to show that (A, B, C) realizes F_c if and only if it represents c (see Problem 1.4.1). But for more general types of rational series, this type of connection is not so transparent. It will be shown in Chapter 6, using the tools developed in Chapter 4, that the problem of realizing a rational operator goes well beyond the traditional boundaries of linear system theory and involves more fundamentally the class of *bilinear* systems, that is, state space systems of the form

$$\begin{aligned} \dot{z}(t) &= Az(t) + \sum_{i=1}^{m} N_i z(t) u_i(t), \quad z(t_0) = z_0 \\ y(t) &= Cz(t). \end{aligned}$$

In particular, note the product between the state $z(t)$ and the input u_i. Despite the name, this system is truly nonlinear (see Problem 1.4.2).

Another observation about rational operators is that $c_y = c \circ c_u$ may not be a polynomial even when each c_{u_i} is a polynomial. This fact is central to defining an abstract notion of *system state*. Suppose, for example, that $c = (ba^{-1})x_1$, where the polynomials a and b have no common roots. Assume that given any polynomial p, $\tilde{p}$ denotes the polynomial whose coefficients appear in the reverse order of those defining p. In light of the Hankel matrix discussion in Section 1.1, it is clear that $\deg(\tilde{a}) = \mathrm{rank}(\mathcal{H}_c) = n$. It can be assumed without loss of generality that $\tilde{a}$ is monic, i.e., the coefficient of its highest order term is one so that

$$\tilde{a} = (\tilde{a}, \emptyset) + (\tilde{a}, x_0)x_0 + \cdots + (\tilde{a}, x_0^{n-1})x_0^{n-1} + x_0^n.$$

Then any polynomial in x_0, say $\tilde{c}_u$, can be uniquely decomposed via Euclidian division into the form $\tilde{c}_u = \tilde{a}\tilde{p} + \tilde{r}$, where $\tilde{p}$ and $\tilde{r}$ are polynomials with the remainder polynomial having the form

$$\tilde{r} = (\tilde{r}, \emptyset) + (\tilde{r}, x_0)x_0 + \cdots + (\tilde{r}, x_0^{n-1})x_0^{n-1}.$$

In which case,

$$c_y = c \circ c_u = (ba^{-1}x_1) \circ (ap + r) = bx_0 p + (ba^{-1})x_0 r$$

is polynomial if and only if $r = 0$. In this setting, two input polynomials $\tilde{c}_u$ and $\tilde{c}_{u'}$ are said to be equivalent when their corresponding remainder polynomials, $\tilde{r}$ and $\tilde{r}'$, are identical. Their respective output series, c_y and $c_{y'}$, can therefore only differ at most by a polynomial. A collection of equivalent inputs forms an equivalence class which can be uniquely identified by the n coefficients of the remainder polynomial which they have in common. Thus, the quotient set can be put in one-to-one correspondence with the n dimensional vector space $\mathbb{R}^n$. It is this representation of the quotient set that provides the familiar notion of a state space in linear systems theory. To see this more clearly, consider how the state $\tilde{r}_-$ is transformed by introducing a new constant term $u_+ \in \mathbb{R}$ according to the mapping

$$\Phi : (\tilde{r}_-, u_+) \mapsto \tilde{r}_+ = [x_0\tilde{r}_- + u_+]_{\tilde{a}},$$

where $[\cdot]_{\tilde{a}}$ denotes the operator which extracts the remainder polynomial from a given polynomial after division by $\tilde{a}$. Using the fact that $x_0^n = \tilde{a} - (\tilde{a}, \emptyset) - (\tilde{a}, x_0)x_0 - \cdots - (\tilde{a}, x_0^{n-1})x_0^{n-1}$, observe that

$$\begin{aligned}
x_0\tilde{r}_- + u_+ &= (\tilde{r}_-, \emptyset)x_0 + (\tilde{r}_-, x_0)x_0^2 + \cdots + (\tilde{r}_-, x_0^{n-1})x_0^n + u_+ \\
&= \tilde{a}(\tilde{r}_-, x_0^{n-1}) + [-(\tilde{a}, \emptyset)(\tilde{r}_-, x_0^{n-1}) + u_+] + \\
&\quad\ [(\tilde{r}_-, \emptyset) - (\tilde{a}, x_0)(\tilde{r}_-, x_0^{n-1})]x_0 + \cdots + \\
&\quad\ [(\tilde{r}_-, x_0^{n-2}) - (\tilde{a}, x_0^{n-1})(\tilde{r}_-, x_0^{n-1})]x_0^{n-1} \\
&= \tilde{a}(\tilde{r}_-, x_0^{n-1}) + \tilde{r}_+.
\end{aligned}$$

Therefore, $\tilde{r}_+ = \Phi(\tilde{r}_-, u_+)$ can be written in component form as

$$\begin{bmatrix} (\tilde{r}_+, \emptyset) \\ (\tilde{r}_+, x_0) \\ \vdots \\ (\tilde{r}_+, x_0^{n-1}) \end{bmatrix}
= \begin{bmatrix} 0 & \cdots & 0 & -(\tilde{a}, \emptyset) \\ 1 & \cdots & 0 & -(\tilde{a}, x_0) \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & -(\tilde{a}, x_0^{n-1}) \end{bmatrix}
\begin{bmatrix} (\tilde{r}_-, \emptyset) \\ (\tilde{r}_-, x_0) \\ \vdots \\ (\tilde{r}_-, x_0^{n-1}) \end{bmatrix}
+ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} u_+.$$

More conventionally, this is equivalent to writing the state equation (1.31) (with $m = 1$) in the integral form

$$z(t_+) - z(t_-) = A\int_{t_-}^{t^+} z(\tau)\, d\tau + B_1 \int_{t_-}^{t^+} u_1(\tau)\, d\tau,$$

where $z(t_+)$ denotes the n dimensional state vector at time instant $t = t_+$ corresponding to $\tilde{r}_+$ and $u_+ = \int_{t_-}^{t^+} u_1(\tau)\,d\tau$. In this setting, the system output $y(t_+)$ is generally understood to only be a function of the current system state, $z(t_+)$. If the input happens to remain in the equivalence class corresponding to $\tilde{r}_+$ for all $t > t_+$ then the respective output would be *completely* determined by $\tilde{r}_+$. In which case, there must exist a series c_{y_+} depending only on $\tilde{r}_+$ such that

$$y(t) = \sum_{k=0}^{\infty} (c_{y_+}, x_0^k) \frac{(t-t_+)^k}{k!}, \quad t \geq t_+. \tag{1.33}$$

It is perhaps easiest to see from classical arguments that $c_{y_+} = \mathcal{H}_c(\tilde{r}_+)$, where $\mathcal{H}_c$ is the Hankel mapping corresponding to c. That is, $\mathcal{H}_c$ is the linear mapping on the real vector space of polynomials whose matrix representation is the Hankel matrix $\mathcal{H}_c$. (No notational distinction will be made between the mapping and its matrix representation.) Observe that if $\tilde{r}_- = 0$ and $u_+ = 1$ then clearly $\tilde{r}_+ = 1$ and $y|_{\tilde{r}_+=1}$ corresponds to the impulse response of the system (see (1.30)). So immediately, one can conclude that $c_{y_+}|_{\tilde{r}_+=1} = h_1 = \mathcal{H}_c(1)$. Now suppose that $\tilde{r}_- = 1$ and $u_+ = 0$. Then $\tilde{r}_+ = x_0$ and $y|_{\tilde{r}_+=x_0}$ is the derivative of the impulse response, that is, $c_{y_+}|_{\tilde{r}_+=x_0} = x_0^{-1}(h_1) = \mathcal{H}_c(x_0)$, where $x_0^{-1}(\cdot)$ denotes the left-shift operator (see (1.7)). Proceeding in this fashion, it becomes evident that

$$c_{y_+}|_{\tilde{r}_+=x_0^j} = x_0^{-j}(h_1) = \mathcal{H}_c(x_0^j), \quad j = 0, 1, \ldots, n-1,$$

where $x_0^{-j}(\cdot)$ is the left-shift operator applied j times. Using superposition for an arbitrary $\tilde{r}_+$, one must conclude that $c_{y_+} = \mathcal{H}_c(r_+)$. In light of (1.33),

$$\begin{aligned}
y(t_+) &= (c_{y_+}, \emptyset) \\
&= (\mathcal{H}_c(r_+), \emptyset) \\
&= \begin{bmatrix} (c, x_1) & (c, x_0x_1) & \cdots & (c, x_0^{n-1}x_1) \end{bmatrix} \begin{bmatrix} (\tilde{r}_+, \emptyset) \\ (\tilde{r}_+, x_0) \\ \vdots \\ (\tilde{r}_+, x_0^{n-1}) \end{bmatrix} \\
&= Cz(t_+).
\end{aligned}$$

The realization (A, B_1, C) is the well known *controllability canonical form* of F_c.

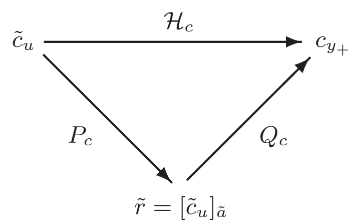


Fig. 1.2. The canonical factorization of the Hankel mapping $\mathcal{H}_c : \tilde{c}_u \mapsto c_{y_+}$.

It is also possible to characterize a state space realization for F_c using exclusively the Hankel mapping $\mathcal{H}_c : \tilde{c}_u \mapsto c_{y_+}$. Suppose the rank of its Hankel matrix representation is n. When $\tilde{c}_u$ and $\tilde{c}_{u'}$ are equivalent polynomials, their difference must be in the null space of $\mathcal{H}_c$. That is,

$$\mathcal{H}_c(\tilde{c}_u - \tilde{c}_{u'}) = \mathcal{H}_c(\tilde{a}(\tilde{p} - \tilde{p}')) = 0,$$

since $\mathcal{H}_c(\tilde{a}) = 0$ (see Problem 1.1.9). Conversely, observe that given any two polynomials $\tilde{c}_u$ and $\tilde{c}_{u'}$ with $\tilde{c}_u - \tilde{c}_{u'}$ in the null space of $\mathcal{H}_c$, it follows that they must be equivalent. Specifically, assume

$$\mathcal{H}_c(\tilde{c}_u - \tilde{c}_{u'}) = \mathcal{H}_c(\tilde{r} - \tilde{r}') = 0,$$

where the first n columns of the Hankel matrix are independent. Since $\tilde{r} - \tilde{r}'$ is a polynomial of at most degree $n-1$, one must conclude that $\tilde{r} - \tilde{r}' = 0$. In summary then,

$$\tilde{c}_u \sim \tilde{c}_{u'} \Longleftrightarrow \mathcal{H}_c(\tilde{c}_u) = \mathcal{H}_c(\tilde{c}_{u'}) \Longleftrightarrow \tilde{c}_u - \tilde{c}_{u'} \in \mathrm{null}(\mathcal{H}_c).$$

A standard result concerning such equivalence relations states that $\mathcal{H}_c$ has a unique decomposition of the form $\mathcal{H}_c(\tilde{c}_u) = Q_c P_c(\tilde{c}_u)$, where $P_c : \tilde{c}_u \mapsto \tilde{r} = [\tilde{c}_u]_{\tilde{a}}$ and $Q_c : \tilde{r} \mapsto c_{y_+}$ are linear mappings, and Q_c is an isomorphism onto $\mathbb{R}^n$. This is referred to as the *canonical factorization* of $\mathcal{H}_c$ (see Figure 1.2). Those familiar with linear systems theory will recognize this factorization as equivalent to a controllability/observability factorization of the Hankel matrix, $\mathcal{H}_c = \mathcal{O}(A, C)\mathcal{C}(A, B_1)$, where

$$\mathcal{O}(A, C) = \left[\begin{array}{cccc} C^T & A^T C^T & (A^T)^2 C^T & \cdots \end{array} \right]^T \tag{1.34}$$
$$\mathcal{C}(A, B_1) = \left[\begin{array}{cccc} B_1 & AB_1 & A^2 B_1 & \cdots \end{array} \right], \tag{1.35}$$

and (A, B_1, C) is the n dimensional controllability canonical realization of F_c (see Problem 1.4.4). As in the previous discussion, the set of equivalence classes forms a state space for F_c which can be identified with the vector space $\mathbb{R}^n$. Here n is the *minimal* state space dimension possible for the rational mapping under consideration. It will be shown in Chapter 4 that for arbitrary rational series, a generalized notion of the Hankel matrix places a central role in characterizing the existence and minimality of bilinear state space realizations.

In the event that F_c is not rational, it is natural to ask whether concepts concerning linear and bilinear state space realizations can be generalized in some natural way. The mathematical framework introduced in Chapter 5 to address this question is the theory of free Lie algebras. Roughly speaking, a free Lie algebras is a noncommutative algebra generated by an alphabet X. In general, an algebra has both a vector space structure and a vector product.[3] In the case of a free Lie algebra, the product is the Lie bracket $[x_i, x_j] = x_i x_j - x_j x_i$, where $x_i, x_j \in X$. It will be shown in Chapter 6 that an *input-affine* state space realization of F_c, namely,

$$\begin{aligned} \dot{z}(t) &= f(z(t)) + \sum_{i=1}^{m} g_i(z(t))u_i(t), \quad z(t_0) = z_0 \\ y(t) &= h(z(t)), \end{aligned}$$

where (f, g, h) are real analytic functions of the state, is guaranteed to exist if and only if the generalized Hankel mapping has a finite *rank* which is defined in a certain Lie algebra sense. The rational case in this context is the special situation where all the functions are linear, i.e., $f(z) = Az$, $g_i(z) = N_i z_i$ and $g(z) = Cz$.

Problems

Section 1.1

Problem 1.1.1 Consider the function $u(t) = 1/(1+t^2)$ defined on $\mathbb{R}$.

(a) Show that u is real analytic on $\mathbb{R}$ but is not entire.
(b) Plot u on the interval $[-1.2, 1.2]$.

[3] See Section A.1 in Appendix A.

(c) Define the truncated Taylor series of u at $t = 0$ to be

$$u_N(t) = \sum_{n=0}^{N} c(n)\frac{t^n}{n!}.$$

On the *same* figure generated in part (b), plot $u_N(t)$ for $N = 10, 20, 30$ over $(-1, 1)$.

(d) Explain what happens as N continues to increase.

Remark: In complex analysis, the following statements are equivalent for a function $u : \mathbb{C} \to \mathbb{C}$:

1. u is entire.
2. u is analytic on $\mathbb{C}$.
3. u can be represented by a single Taylor series.

The example above illustrates that statement 2 does *not* imply statement 3 if u is only real analytic on $\mathbb{R}$.

Problem 1.1.2 Consider a function $u : \mathbb{R} \to \mathbb{R}$ whose Taylor series at $t = 0$ has coefficients $c(n)$, $n \geq 0$ satisfing the *Gevrey* growth condition

$$|c(n)| \leq KM^n(n!)^s, \ \ \forall n \geq 0$$

with $0 \leq s < 1$.

(a) Determine the radius of convergence for this series representation.

(b) When $s = 0$ show that $|u(t)|$ can be bounded by an exponential function.

(c) Show how part (b) can be generalized for any $0 \leq s < 1$ using the *Mittag-Leffler function*.

Remark: For any fixed $0 \leq \alpha < 1$, the gamma function satisfies the inequality $\Gamma(\alpha n + 1) \leq K_\alpha M_\alpha^n (n!)^\alpha$ for some $K_\alpha, M_\alpha > 0$ provided $\alpha n \gg 1$.

Problem 1.1.3 Let $u : U \subset \mathbb{R} \to \mathbb{R}$ be real analytic at t_0. Show that u is continuous at t_0.

Problem 1.1.4 Verify equation (1.5) for the coefficients of a re-centered power series.

Problem 1.1.5 Show that if $u : U \subset \mathbb{R} \to \mathbb{R}$ is real analytic at t_0 then the derivative of u is well defined within the radius of convergence and can be computed by equation (1.6). Also show that in this situation u must be smooth within the radius of convergence.

Problem 1.1.6 Show that the function $u(t) = e^{t^2}$ is entire but does not have a well defined right-sided Laplace transform.

Problem 1.1.7 Verify all the entries in Table 1.1.

Problem 1.1.8 Let c be a series with corresponding Hankel matrix $\mathcal{H}_c$. Assume that the rank of $\mathcal{H}_c$ is $n < \infty$.

(a) Using the Hankel structure of $\mathcal{H}_c$, show that the $(n+1)$-st column of $\mathcal{H}_c$ is in the span of the first n columns of $\mathcal{H}_c$.
(b) Show by induction that every column of $\mathcal{H}_c$ beyond the first n columns must be in the span of the first n columns. Therefore, the first n columns of $\mathcal{H}_c$ must be linearly independent.
(c) Suppose $\tilde{q}$ is a nonzero polynomial satisfying equation (1.9). Using the result from part (b), show that $\tilde{q}(n) \neq 0$.

Problem 1.1.9 Suppose $c = pq^{-1}$, where p and q are polynomials in x with $\deg(p) < \deg(\tilde{q}) = \mathrm{rank}(\mathcal{H}_c)$.

(a) Show that $\mathcal{H}_c(\tilde{q}\tilde{r}) = 0$ for any polynomial $\tilde{r}$ in x.
(b) Prove that p and q can share no common roots.

Problem 1.1.10 For each series c below, determine, if possible, two polynomials p and q such that $c = pq^{-1}$. Also compute the Laplace transform for the corresponding input u and its region of convergence.

(a) $c = x^2$
(b) $c = x + x^3 + x^5 + \cdots$
(c) $c = 1 - \frac{1}{2}x + \frac{1}{3}x^2 - \cdots$

Section 1.2

Problem 1.2.1 Determine whether each function below is real analytic at $t = 0$. If so, determine the radius of convergence; if not, comment on whether the function is at least smooth at $t = 0$.

(a) $u(t) = e^t$

(b) $u(t) = 1/(t^2 + 1)$

(c) $u(t) = \begin{cases} e^{-1/t^2} & : t \neq 0 \\ 0 & : t = 0 \end{cases}$

(d) $u(t) = \begin{cases} t \log |t| & : t \neq 0 \\ 0 & : t = 0 \end{cases}$

Section 1.3

Problem 1.3.1 Provide the following details regarding the Chen-Fliess series for a linear time-invariant system:

(a) The integration by parts calculation that gives equation (1.22).
(b) The inductive proof that leads to equation (1.23).

Problem 1.3.2 A function $u : U \subset \mathbb{R} \to \mathbb{R}^m$ is said to be *absolutely integrable* on an interval $[t_0, t_1] \subset U$ if

$$\int_{t_0}^{t_1} |u_i(t)|\, dt < \infty, \quad i = 1, \ldots, m.$$

Assume that the interval $[t_0, t_1]$ is finite. Show that if u is piecewise continuous (meaning that each u_i has at most a finite number of *jump discontinuities*) then

(a) u is absolutely integrable;
(b) the inequality

$$\int_{t_0}^{t_1} |u_i(t)|^{\mathfrak{p}}\, dt < \infty, \quad i = 1, \ldots, m$$

is satisfied for every integer $\mathfrak{p} \geq 1$;
(c) the iterated integral $E_\eta[u](t, t_0)$ is finite for all $t \in [t_0, t_1]$ and every $\eta \in X^*$.

Problem 1.3.3 Show that the formal Laplace transforms of u and y, where $y = F_c[u]$ and c is given by equation (1.26), are related by $c_y = c \circ c_u$ as described in equation (1.29). Also verify that c_y is rational whenever c and c_u are rational, that is, when $c_i = b_i a_i^{-1}$ and $c_{u_i} = p_i q_i^{-1}$, where a_i, b_i, p_i and q_i are polynomials in x_0 for $i = 1, 2, \ldots, m$.

Section 1.4

Problem 1.4.1 Suppose that (A, B, C) is known to represent a formal power series $c : X^* \to \mathbb{R}$. Show that (A, B, C) also realizes the linear input-output operator

$$F_c : u \mapsto y(t) = \sum_{i=1}^{m} \int_{t_0}^{t} H_i(t-\tau)u_i(\tau)\, d\tau, \quad t \geq t_0,$$

where the series $H_i(t) = \sum_{k=0}^{\infty}(c, x_0^k x_i)\, t^k/k!$, $i = 1, \ldots, m$. Show that the converse statement is also true.

Problem 1.4.2 Consider a bilinear state space system

$$\begin{aligned}\dot{z}(t) &= N_0 z(t) + N_1 z(t)u(t), \quad z(0) = z_0\\ y(t) &= Cz(t).\end{aligned}$$

(a) Write the solution $z(t)$ of the state equation in terms of a Peano-Baker series.
(b) Give a series expression for the output $y(t)$.
(c) Determine $y(t)$ for the case:

$$N_0 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad N_1 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad z_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Problem 1.4.3 Let c be the generating series for the integral operator in equation (1.26) with $m = 1$.

(a) Show that $\mathcal{H}_c(\tilde{c}_u) = x_0^{-(n_u+1)}(c \circ c_u)$, where $\tilde{c}_u$ is a polynomial of degree n_u.
(b) Suppose c is a rational series. Let $\tilde{c}_u$ and $\tilde{c}_{u'}$ be equivalent polynomials with corresponding output series $c_{y_+} = \mathcal{H}_c(\tilde{c}_u)$ and $c_{y'_+} = \mathcal{H}_c(\tilde{c}_{u'})$. Using *only* the result from part (a), show that $c_{y_+} = c_{y'_+}$, or equivalently, that $\tilde{c}_u - \tilde{c}_{u'}$ is in the null space of $\mathcal{H}_c$.

Problem 1.4.4 Show that the observability and controllability matrices described by equations (1.34)-(1.35) produce the canonical factorization of the Hankel mapping $\mathcal{H}_c$.

Bibliographic Notes

More detailed bibliographic notes are deferred to the later chapters, where the topics in this chapter reappear more fully developed. Here some citations are provided of a more general nature for those readers wanting to expand their background in various directions.

Section 1.1 The study of real analytic functions in one variable is a standard topic in real analysis. Basic treatments of the subject appear in the introductory textbooks by Bartle [4], Bromwich [21], and Knopp [132]. More advanced topics can be found in the books by Balser [3] and Ruiz [167]. Power series in one variable are also treated in these same introductory texts. For the multivariable versions of these topics, the book by Gröbner [98] is very complete. Any basic text in complex analysis will address the topic of analytic functions defined on the complex plane, see, for example, [2, 23, 178]. For a thorough treatment of this topic in the multivariable setting see the book by Hörmander [109].

Section 1.2 An accessible treatment of formal functions appears in the book by Castrigiano and Hayes [29, Chapter 4]. The text by Wilf [204] provides a general introduction to the topic of generating functions. A number of books are available concerning formal power series, most do so in the context of formal languages or theoretical computer science. This includes the books by Berstel [7], Berstel and Reutenauer [8], Conway [46], Gross and Lentin [99], Harrison [106], Kuich and Salomaa [134], Reutenauer [158], Rèvèsz [159], Rozenberg and Salomaa [163], Salomaa [168], and Salomaa and Soittola [169]. A survey of the subject as it applies to systems theory appears in the tutorial paper by Fliess [74] and the textbook by Isidori [113, Chapter 3].

Section 1.3 Many texts are available treating linear integral operators. A few from the mathematical point of view include those by Kreyszig [133], and Naylor and Sell [152]. Those from a systems point of view include the texts by Callier and Desoer [25] and Kailath [125]. A number of references are available addressing nonlinear integral operators. The books by Isidori [113], Rugh [166], and Schetzen [172], as well as the papers by Brockett [20], Crouch [47], and Wong [207] provide good introductions.

Section 1.4 Realization theory for linear systems is treated very comprehensively in the textbook by Kailath [125]. Chapter 5, in particular, gives a nice overview of the algebraic approach to the subject, which underlies many of the concepts appearing in this book. See the early work of Kalman for some historical perspective on this approach [126, 127]. The textbooks by Callier and Desoer [25] and Chen [30] are also useful references. Concerning nonlinear state space realizations, the textbooks by Isidori [113], Khalil [130], Nijmeijer and van der Schaft [153], Sontag [181], and Vidyasagar [196] all give comprehensive introductions to the subject.

2. Formal Power Series

Many of the mathematical objects that appear in this book arise naturally in the theory of formal languages. So the starting point for this chapter is an introduction to formal power series in this setting. As these ideas are developed, it will become apparent that formal power series often have a combinatorial nature, that this, their manipulation involves partitions, permutations, etc. Considerable time will be spent on the shuffle product, as it is ubiquitous when combinatorics is applied in system theory. In this regard, combinatorial Hopf algebras are also useful, especially when explicit computations are needed. So the notion of a Hopf algebra is introduced along with some important examples which will be useful henceforth. Finally, a general notion of composition for formal power series is presented. This device will be used in Chapter 3 for describing the series interconnection of two input-output systems and to define a formal input-output map.

2.1 Formal Languages

A finite nonempty set of arbitrary symbols $X = \{x_0, x_1, \ldots, x_m\}$ is called an *alphabet*. Each element of X is called a *letter*, and any finite sequence of letters from X, $\eta = x_{i_k} \cdots x_{i_1}$, is called a *word* over X. Two words η and ξ are equivalent, i.e., $\eta = \xi$, if one word is letter by letter equivalent to the other. The *length* of a word η is equivalent to the number of letters in η and will be denoted by $|\eta|$. In addition, $|\eta|_{x_i}$ is equivalent to the number of times the letter x_i appears in η. The empty word, $\emptyset$, has length zero. The set of words with length k will be denoted by X^k. The set of all words is represented by X^*, while X^+ is the set of all words with positive length, i.e., the nonempty words. A *language* is any subset of X^*.

Consider the following binary operation on X^*.

Definition 2.1 *The **catenation product** on X^* is the associative mapping*

$$\begin{aligned}\mathcal{C} &: X^* \times X^* \to X^* \\ &: (\eta, \xi) \mapsto \eta\xi.\end{aligned}$$

That is, for any $\eta, \xi, \nu \in X^*$ it follows that

$$(\eta\xi)\nu = \eta(\xi\nu).$$

Furthermore, the empty word $\emptyset$ is an identity element for $\mathcal{C}$ since

$$\emptyset\eta = \eta\emptyset = \eta, \ \ \forall \eta \in X^*.$$

For any positive integer i and $\eta \in X^*$, the i-th *iterate* of η is $\eta^i = \eta \cdots \eta$, where η appears i times. Normally, $\eta^0 := \emptyset$. The triple $(X^*, \mathcal{C}, \emptyset)$ (or simply X^* when the rest is understood) is referred to in algebraic parlance as a *free monoid* on X. The adjective *free* is referring to the assumption that there are no relationships between the letters. For example, the letter x_1 can not be used to represent the word x_2x_3. In some situations involving *groups*, it is useful to have relationships like $x_ix_j = x_jx_i = \emptyset$ so that x_j can be thought of as the multiplicative inverse of x_i and vice versa.

Given two arbitrary monoids $(M, \square, e)$ and $(M', \square', e')$, a mapping $\rho : M \to M'$ is called a *monoid homomorphism* if

$$\rho(\eta \square \xi) = \rho(\eta) \square' \rho(\xi), \ \ \forall \eta, \xi \in M \tag{2.1}$$

and $\rho(e) = e'$. When ρ is bijective it is called an *isomorphism*. Given an arbitrary alphabet $X = \{x_0, x_1, \ldots, x_m\}$, any mapping $\rho : X \to M'$ can be uniquely extended to a homomorphism $\rho : X^* \to M'$ by letting

$$\rho(x_{i_k} x_{i_{k-1}} \cdots x_{i_1}) = \rho(x_{i_k}) \square' \rho(x_{i_{k-1}}) \square' \cdots \square' \rho(x_{i_1})$$

(see Problem 2.1.3). If $y_i := \rho(x_i)$ for each $x_i \in X$, then $\rho(X^*)$ can be viewed as a free *submonoid* of M' corresponding to the alphabet $Y = \{y_0, y_1, \ldots, y_m\}$. If ρ is injective, i.e., if $\rho(\eta) = \rho(\xi)$ always implies that $\eta = \xi$, $\forall \eta, \xi \in X^*$, then ρ is called a *coding* of X^*.

Example 2.1 Suppose $X = \{0, 1, \ldots, 9\}$ and $Y = \{0, 1\}$. The binary coded decimals

$$\rho(0) = 0000, \ \ \rho(1) = 0001, \ \ldots \ \rho(9) = 1001$$

define a coding of X^*, but not an isomorphism since, for example, the word $1010 \in Y^*$ is not in the range of ρ. □

Example 2.2 Consider the usual multiplicative monoid $\{\mathbb{R}^+, \cdot, 1\}$ on the set of positive real numbers, $\mathbb{R}^+$, and the additive monoid $\{\mathbb{R}, +, 0\}$ on the set of real numbers $\mathbb{R}$. Select any $a \in \mathbb{R}^+$ then the map

$$\begin{aligned} \rho &: \mathbb{R}^+ \to \mathbb{R} \\ &: y \mapsto y' = \log_a(y) \end{aligned}$$

defines an isomorphism since ρ is bijective,

$$\rho(y_1 y_2) = \rho(y_1) + \rho(y_2), \quad y_1, y_2 \in \mathbb{R}^+,$$

and $\rho(1)$=0. □

Example 2.3 The natural numbers $\mathbb{N}_0 := \{0, 1, 2, \ldots\}$ provide the submonoid $\{\mathbb{N}_0, +, 0\}$ of the monoid $\{\mathbb{R}, +, 0\}$. With $X = \{x\}$ and $X^* = \{\emptyset, x, x^2, \ldots\}$, the mapping $|x^i| = i$, $i \geq 0$ defines an isomorphism between X^* and $\mathbb{N}_0$. It is clearly a restriction of the isomorphism in the previous example, where $x \in \mathbb{R}^+$ is left unspecified, and $\emptyset$ is identified with $1 \in \mathbb{R}^+$. □

This last example suggests an alternative way to express a power series in one variable,

$$c = \sum_{i=0}^{\infty} c(i)\, x^i.$$

Namely, defining $(c, \eta) = c(|\eta|)$ for every $\eta \in X^*$, where $X = \{x\}$, the series can be written as the summation over X^*

$$c = \sum_{\eta \in X^*} (c, \eta)\, \eta.$$

Example 2.4 Any $\mathbb{R}$-linear mapping on the vector space $\mathbb{R}^n$ can be represented by a matrix in $\mathbb{R}^{n \times n}$. The collection of matrices $\mathbb{R}^{n \times n}$ clearly forms a monoid under the usual definition of matrix multiplication, where the identity matrix, I_n, is the multiplicative identity element. Given an alphabet $X = \{x_0, x_1, \ldots, x_m\}$, let μ denote a mapping which assigns a specific matrix to each letter, namely, $\mu(x_i) = N_i$

for $i = 0, 1, \dots, m$. Then there exists a unique free submonoid in $\mathbb{R}^{n\times n}$ generated by $N := \{N_0, N_1, \dots, N_m\}$ (the N_i's are not necessarily linearly independent). This type of matrix monoid plays a central role in the theory of linear representations for formal power series presented in Section 4.2.

□

2.2 Formal Power Series

Given an alphabet $X = \{x_0, x_1, \dots, x_m\}$, a *formal power series* c is any function of the form

$$c : X^* \to \mathbb{R}^\ell.$$

The *image* of a word $\eta \in X^*$ under c is denoted by (c, η) and is called the *coefficient* of η in c. It is customary to write c as the formal summation

$$c = \sum_{\eta\in X^*} (c, \eta)\, \eta.$$

The coefficient $(c, \emptyset)$ is referred to as the *constant term*, and c is called *proper* when this coefficient is zero. The *support* of c is the language

$$\mathrm{supp}(c) := \{\eta \in X^* : (c, \eta) \neq 0\}.$$

The *order* of c is defined as

$$\mathrm{ord}(c) = \begin{cases} \min\{|\eta| : \eta \in \mathrm{supp}(c)\} & : c \neq 0 \\ \infty & : c = 0. \end{cases}$$

So when c is proper, it follows that $\mathrm{ord}(c) > 0$. The set of all formal power series will be denoted by $\mathbb{R}^\ell\langle\langle X\rangle\rangle$. In addition, the set of all formal power series with finite support, i.e., the set of all polynomials, will be represented by $\mathbb{R}^\ell\langle X\rangle$. The *degree* of a polynomial p is

$$\deg(p) = \begin{cases} \max\{|\eta| : \eta \in \mathrm{supp}(p)\} & : p \neq 0 \\ -\infty & : p = 0. \end{cases}$$

As a matter of notation, $\eta = \emptyset$ denotes the empty word in X^*, while the polynomials $p = 1\emptyset$ and $p = 0\emptyset$ will usually be written simply as

$p = 1$ and $p = 0$, respectively.[1] When $\ell \geq 1$, the i-th *component series* of $c \in \mathbb{R}^\ell\langle\langle X\rangle\rangle$ is

$$c_i = \sum_{\eta \in X^*} (c, \eta)_i \eta,$$

where $(c, \eta)_i$ is the i-th component of the vector $(c, \eta) \in \mathbb{R}^\ell$. In which case, there is a natural bijection between $\mathbb{R}^\ell\langle\langle X\rangle\rangle$ and $(\mathbb{R}\langle\langle X\rangle\rangle)^\ell$.

The sets $\mathbb{R}^\ell\langle\langle X\rangle\rangle$ and $\mathbb{R}^\ell\langle X\rangle$ exhibit considerable algebraic structure. For example, each admits a vector space structure over $\mathbb{R}$ when addition $c + d$ is defined by the coefficients

$$(c + d, \eta) = (c, \eta) + (d, \eta), \quad \forall \eta \in X^*,$$

and scalar multiplication αc is given by

$$(\alpha c, \eta) = \alpha(c, \eta), \quad \forall \eta \in X^*, \quad \forall \alpha \in \mathbb{R}.$$

It is straightforward to show that

$$\begin{aligned} \operatorname{ord}(c + d) &\geq \min\{\operatorname{ord}(c), \operatorname{ord}(d)\}, \quad c, d \in \mathbb{R}\langle\langle X\rangle\rangle \\ \deg(p + q) &\leq \max\{\deg(p), \deg(q)\}, \quad p, q \in \mathbb{R}\langle X\rangle. \end{aligned}$$

When $\ell = 1$, each set forms a ring, an associative $\mathbb{R}$-algebra, and a module over the ring $\mathbb{R}\langle X\rangle$ using the following product.[2]

Definition 2.2 *The* ***catenation product*** *or* ***Cauchy product*** *of two series (or polynomials)* $c, d \in \mathbb{R}\langle\langle X\rangle\rangle$ *is* $cd = \sum_{\eta \in X^*} (cd, \eta)\, \eta$, *where*

$$(cd, \eta) = \sum_{\substack{\xi, \nu \in X^* \\ \eta = \xi\nu}} (c, \xi)(d, \nu), \quad \forall \eta \in X^*,$$

or more succinctly,

$$(cd, \eta) = \sum_{\eta = \xi\nu} (c, \xi)(d, \nu), \quad \forall \eta \in X^*.$$

In this case, the polynomial $p = 1$ acts as the multiplicative identity element (see Problem 2.2.1).

---

[1] When it is necessary to distinguish between the scalar 1 and the polynomial $1\emptyset$, the latter will be written as **1**.

[2] The reader may wish to consult Appendix A for a brief review of these algebraic concepts.

Example 2.5 Suppose $X = \{x\}$ and consider two polynomials:

$$\begin{aligned} c &= (c,\emptyset) + (c,x)\,x + (c,x^2)\,x^2 \\ d &= (d,\emptyset) + (d,x)\,x + (d,x^2)\,x^2. \end{aligned}$$

The familiar polynomial product of c and d is computed as

$$\begin{aligned} cd &= [(c,\emptyset) + (c,x)\,x + (c,x^2)\,x^2]\,[(d,\emptyset) + (d,x)\,x + (d,x^2)\,x^2] \\ &= (c,\emptyset)(d,\emptyset) + [(c,\emptyset)(d,x) + (c,x)(d,\emptyset)]\,x + [(c,\emptyset)(d,x^2) + \\ &\qquad (c,x)(d,x) + (c,x^2)(d,\emptyset)]\,x^2 + [(c,x)(d,x^2) + \\ &\qquad (c,x^2)(d,x)]\,x^3 + (c,x^2)(d,x^2)\,x^4 \\ &= \sum_{i=0}^{4}\left[\sum_{\substack{j,k\geq 0\\ j+k=i}} (c,x^j)(d,x^k)\right] x^i. \end{aligned}$$

Thus, it follows that

$$\begin{aligned} (cd,x^i) &= \sum_{\substack{j,k\geq 0\\ j+k=i}} (c,x^j)(d,x^k) \\ &= \sum_{\eta=\xi\nu} (c,\xi)(d,\nu). \end{aligned}$$

So the catenation product of polynomials (or series) over a single letter alphabet reduces to the usual polynomial product. It is easy to see in this case that the catenation product is commutative. Also observe that

$$(cd,x^i) = \sum_{j=0}^{i} (c,x^j)(d,x^{i-j}).$$

Hence, for single letter alphabets, the catenation product is exactly equivalent to the notion of series *convolution*. □

Example 2.6 Suppose $X = \{x_0, x_1\}$, $c = 2x_0x_1$ and $d = x_0 + x_1$. Then it follows that

$$\begin{aligned} cd &= 2x_0x_1(x_0 + x_1) \\ &= 2x_0x_1x_0 + 2x_0x_1^2, \end{aligned}$$

while on the other hand

$$\begin{aligned} dc &= (x_0 + x_1)\, 2x_0x_1 \\ &= 2x_0^2x_1 + 2x_1x_0x_1. \end{aligned}$$

Thus, the catenation product is not commutative when X contains more than one letter. □

Example 2.7 Consider the special case where the letters of $X = \{x_1, x_2, \ldots, x_m\}$ commute, that is, $x_ix_j = x_jx_i$ for all $x_i, x_j \in X^*$. Let $\mathbb{R}[[X]]$ denote the set of all formal power series on this commuting alphabet. In this case, an alternative definition of the catenation product is often useful, namely,

$$cd = \sum_{\eta\in X^*} (cd, \eta)\, \frac{\eta}{\eta!},$$

where

$$(cd, \eta) = \sum_{\eta=\xi\nu} (c, \xi)(d, \nu)\, \frac{\eta!}{\xi!\nu!}$$

and $\eta! := |\eta|_{x_1}!\, |\eta|_{x_2}! \cdots |\eta|_{x_m}!$. For a single letter alphabet, this yields what is commonly called *binomial convolution*

$$(cd, x^i) = \sum_{j=0}^{i} \binom{i}{j} (c, x^j)(d, x^{i-j}). \tag{2.2}$$

The catenation product in the general commutative case will be referred to as the *multinomial catenation product.* Binomial convolution naturally arises when describing the pointwise product of real analytic functions. That is, if

$$f_c(z) = \sum_{i=0}^{\infty} (c, x^i) \frac{z^i}{i!}$$

and likewise for f_d, then $f_cf_d = f_{cd}$ (see Problem 2.2.2). It will usually be clear from context which catenation product is at play. For noncommutative alphabets, it will always be the Cauchy product. For the commutative case where the pointwise product of functions is involved, it will always be the multinomial catenation product. □

One of the most common operations performed on series is to shift its coefficients in some manner to other words in the series. For example, given the single letter alphabet $X = \{x\}$, the familiar left-shift operator is

$$\begin{aligned} x^{-1}(c) &= x^{-1}((c,\emptyset) + (c,x)x + (c,x^2)x^2 + (c,x^3)x^3 + \cdots) \\ &= (c,x) + (c,x^2)x + (c,x^3)x^2 + (c,x^4)x^3 + \cdots . \end{aligned}$$

When applied twice, the operator $x^{-1}(x^{-1}(\cdot))$ could be thought of as $(x^2)^{-1}(\cdot)$. These ideas are generalized for an arbitrary alphabet in the following definition.

Definition 2.3 *Given any $\xi \in X^*$, the corresponding* ***left-shift operator*** *on X^* is defined as*

$$\begin{aligned} \xi^{-1} &: X^* \to \mathbb{R}\langle X\rangle \\ &: \eta \mapsto \begin{cases} \eta' : \eta = \xi\eta' \\ 0 : \text{otherwise.} \end{cases} \end{aligned}$$

Note that in the second half of this definition, η is being mapped to the zero polynomial, i.e., $p = 0$, as opposed to the empty word $\emptyset$. So this operator is a mapping into $\mathbb{R}\langle X\rangle$ and not into X^*. For any $c \in \mathbb{R}^\ell\langle\langle X\rangle\rangle$, this definition is extended linearly as

$$\begin{aligned} \xi^{-1}(c) &= \sum_{\eta \in X^*} (c,\eta)\, \xi^{-1}(\eta) \\ &= \sum_{\eta \in X^*} (c,\xi\eta)\, \eta. \end{aligned}$$

In which case, $\xi^{-1}(\cdot)$ acts linearly on the $\mathbb{R}$-vector space $\mathbb{R}^\ell\langle\langle X\rangle\rangle$, that is,

$$\xi^{-1}(\alpha_1 c_1 + \alpha_2 c_2) = \alpha_1 \xi^{-1}(c_1) + \alpha_2 \xi^{-1}(c_2)$$

for all $\alpha_i \in \mathbb{R}$ and $c_i \in \mathbb{R}^\ell\langle\langle X\rangle\rangle$ (see Problem 2.2.3). Two key properties of left-shift operators are given in the following lemma.

Lemma 2.1 *Let $x_i \in X$ and $\xi, \nu \in X^*$ be fixed. For any $c, d \in \mathbb{R}\langle\langle X\rangle\rangle$ it follows that:*

1. $(\xi\nu)^{-1}(c) = \nu^{-1}(\xi^{-1}(c))$

2. $x_i^{-1}(cd) = x_i^{-1}(c)\, d + (c,\emptyset)\, x_i^{-1}(d)$.

Proof:

1. This property follows directly from the definition,

$$\begin{aligned}
(\xi\nu)^{-1}(c) &= \sum_{\eta\in X^*} (c,\eta)\,(\xi\nu)^{-1}(\eta) \\
&= \sum_{\eta\in X^*} (c,\eta)\,\nu^{-1}(\xi^{-1}(\eta)) \\
&= \nu^{-1}\xi^{-1}\left(\sum_{\eta\in X^*} (c,\eta)\,\eta\right) \\
&= \nu^{-1}\xi^{-1}(c).
\end{aligned}$$

2. For any $x_i \in X$ observe that

$$\begin{aligned}
x_i^{-1}(cd) &= \sum_{\eta\in X^*} (cd, x_i\eta)\,\eta \\
&= \sum_{\eta\in X^*}\left(\sum_{x_i\eta=\xi\nu} (c,\xi)(d,\nu)\right)\eta \\
&= \sum_{\eta\in X^*}\left(\sum_{\eta=\xi\nu} (c,x_i\xi)(d,\nu) + (c,\emptyset)(d,x_i\eta)\right)\eta \\
&= \sum_{\eta\in X^*}\left(\sum_{\eta=\xi\nu} (x_i^{-1}(c),\xi)(d,\nu)\right)\eta + (c,\emptyset)\sum_{\eta\in X^*}(x_i^{-1}(d),\eta)\,\eta \\
&= x_i^{-1}(c)\, d + (c,\emptyset)\, x_i^{-1}(d).
\end{aligned}$$

■

2.3 The Ultrametric Space $\mathbb{R}^\ell\langle\langle X\rangle\rangle$

It will be useful in a number of situations to provide the vector space $\mathbb{R}^\ell\langle\langle X\rangle\rangle$ with a topological structure so that concepts like convergence are available. Convergence in this case does not mean convergence of a power series in $\mathbb{R}^\ell\langle\langle X\rangle\rangle$, but rather convergence of a *sequence* of power series in $\mathbb{R}^\ell\langle\langle X\rangle\rangle$. The approach taken here employs the following notion of *distance*.

Definition 2.4 *Given a set S, a function $\delta : S \times S \to \mathbb{R}$ is called an* ***ultrametric*** *if it satisfies the following properties for all $s, s', s'' \in S$:*

i. $\delta(s, s') \geq 0$

ii. $\delta(s, s') = 0$ *if and only if* $s = s'$

iii. $\delta(s, s') = \delta(s', s)$

iv. $\delta(s, s') \leq \max\{\delta(s, s''), \delta(s'', s')\}$.

The pair (S, δ) is referred to as an ***ultrametric space****.*

In the event that property *iv* is replaced with the triangle inequality,

$$\delta(s, s') \leq \delta(s, s'') + \delta(s'', s'), \tag{2.3}$$

(S, δ) is called a *metric space*. Clearly, *iv* implies (2.3) but not conversely. Thus, every ultrametric space is a metric space. Now for any fixed real number σ such that $0 < \sigma < 1$, consider the mapping

$$\begin{aligned} \mathrm{dist} &: \mathbb{R}^\ell\langle\langle X\rangle\rangle \times \mathbb{R}^\ell\langle\langle X\rangle\rangle \to \mathbb{R} \\ &: (c, d) \mapsto \sigma^{\mathrm{ord}(c-d)}. \end{aligned}$$

The following theorem is essential.

Theorem 2.1 *The $\mathbb{R}$-vector space $\mathbb{R}^\ell\langle\langle X\rangle\rangle$ with mapping* dist *is an ultrametric space.*

Proof: The proof is left as an exercise (see Problem 2.3.1). ■

Example 2.8 Suppose $X = \{x\}$ and $\sigma = 1/2$. If

$$\begin{aligned} c &= 1 + x + x^2 + \cdots \\ d &= 1 + x + x^2 \end{aligned}$$

then

$$c - d = x^3 + x^4 + x^5 + \cdots$$

so that $\mathrm{ord}(c - d) = 3$ and $\mathrm{dist}(c, d) = 1/8$. □

A sequence $\{s_1, s_2, \ldots\}$ in a metric space S is said to *converge* to $s \in S$, i.e., $\lim_{i\to\infty} s_i = s$, if $\lim_{i\to\infty} \delta(s_i, s) = 0$. This means precisely that for every $\epsilon > 0$ there exists a natural number N_ϵ such that $\delta(s_i, s) < \epsilon$ when $i \geq N_\epsilon$. Any such limit point s will always be unique (see Problem 2.3.2). As the ultrametric dist never exceeds one, there is no loss of generality in assuming that $0 < \epsilon \leq 1$ when applying the definition to sequences in $\mathbb{R}^\ell\langle\langle X\rangle\rangle$.

Example 2.9 Suppose $X = \{x\}$ and let

$$c = 1 + x + x^2 + \cdots$$
$$c_i = 1 + x + \cdots + x^i, \;\; i \geq 0.$$

Then $\mathrm{dist}(c_i, c) = \sigma^{i+1}$. For any $0 < \epsilon \leq 1$, set $N_\epsilon = \lceil \log(\epsilon)/\log(\sigma)\rceil$, where $\lceil\cdot\rceil$ denotes the ceiling function. Then if $i \geq N_\epsilon$ it follows that $\sigma^{i+1} < \epsilon$. Therefore, $\lim_{i\to\infty} c_i = c$. □

Example 2.10 Consider a proper series c and the sequence $\{1, c, c^2, \ldots\}$, where c^i denotes the catenation power, that is, $c^i := cc\cdots c$, where c appears i times and $c^0 := 1$. For any $i \geq 1$ observe

$$\mathrm{dist}(c^i, 0) = \sigma^{\mathrm{ord}(c^i - 0)} = \sigma^{i\,\mathrm{ord}(c)}.$$

The properness of c implies that $\mathrm{ord}(c) > 0$. For any $0 < \epsilon \leq 1$, set

$$N_\epsilon = \left\lceil \frac{\log(\epsilon)}{\log\left(\sigma^{\mathrm{ord}(c)}\right)} \right\rceil + 1,$$

Then it follows that $i \geq N_\epsilon$ gives $\mathrm{dist}(c^i, 0) < \epsilon$, and thus, $\lim_{i\to\infty} c^i = 0$. □

Using the definition to prove that a given sequence in $\mathbb{R}^\ell\langle\langle X\rangle\rangle$ converges requires one to identify a priori a limit $c \in \mathbb{R}^\ell\langle\langle X\rangle\rangle$. The classical way around this problem is to use the notion of a *Cauchy sequence*. A sequence $\{s_1, s_2, \ldots\}$ in any metric space (S, δ) is said to be a Cauchy sequence if for every $\epsilon > 0$ there exists a natural number N_ϵ such that $\delta(s_i, s_j) < \epsilon$ whenever $i, j \geq N_\epsilon$. It is easily verified that every convergent sequence is a Cauchy sequence (see Problem 2.3.2). A metric space is said to be *complete* if every Cauchy sequence is convergent.

Theorem 2.2 *The ultrametric space* $(\mathbb{R}^\ell\langle\langle X\rangle\rangle, \mathrm{dist})$ *is complete.*

Proof: Let $\{c_1, c_2, \ldots\}$ be a Cauchy sequence in $\mathbb{R}^\ell\langle\langle X\rangle\rangle$. Then for any $k \geq 0$ there exists a natural number N_k such that $\mathrm{dist}(c_i, c_j) < \sigma^k$ whenever $i, j \geq N_k$. Therefore, $\mathrm{ord}(c_i - c_j) > k$, or equivalently, $(c_i, \eta) = (c_j, \eta)$ when $|\eta| \leq k$. Now define a new series $c \in \mathbb{R}^\ell\langle\langle X\rangle\rangle$ by setting

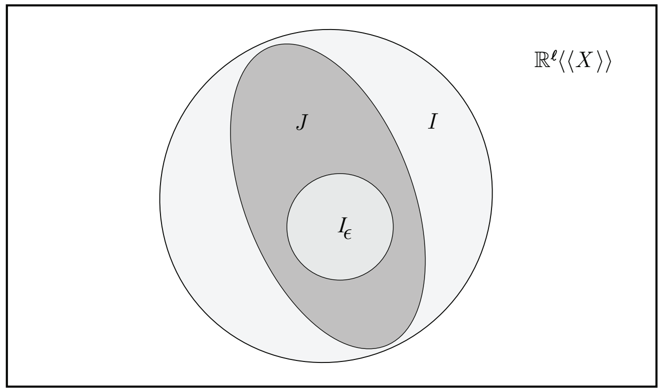


Fig. 2.1. The subsets in the definition of summable.

$$(c,\eta) = (c_{N_k},\eta), \quad \forall \eta \in X^k, \quad k \geq 0.$$

The claim is that $\lim_{i\to\infty} c_i = c$. Choose any $\epsilon > 0$. Select integer $k \geq 0$ such that $\sigma^k < \epsilon$. Then for all $i \geq N_k$

$$\mathrm{dist}(c_i, c) < \sigma^k < \epsilon.$$

This completes the proof. ■

Frequently, one encounters the situation where a family of formal power series needs to be added together to form a new series. If this family is infinite, it is not so obvious at first glance how such a sum should be defined. Clearly, the topological structure has to be involved. The following definition addresses this issue.

Definition 2.5 *Let $\{c_i\}_{i\in I}$ be a family of series in $\mathbb{R}^\ell\langle\langle X\rangle\rangle$. The family is said to be **summable** if there exists another series c in $\mathbb{R}^\ell\langle\langle X\rangle\rangle$ with the property that for any $\epsilon > 0$, there exists a finite subset $I_\epsilon \subset I$ such that for any other finite subset $J \subset I$ containing I_ϵ it follows that*

$$\mathrm{dist}\left(\sum_{j\in J} c_j, c\right) < \epsilon$$

(see Figure 2.1).

The set I_ϵ denotes the smallest subset of series one must combine in order to get within ϵ of c. Including more terms, such as those in J, can get one closer to c but no further away than distance ϵ. When $\{c_i\}_{i\in I}$ is summable, its sum will be written as $c = \sum_{i\in I} c_i$, and, in particular,

$$(c,\eta) = \sum_{i\in I(\eta)} (c_i,\eta), \quad \eta \in X^*,$$

where $I(\eta) := \{i \in I : (c_i,\eta) \neq 0\}$. It is left to the reader to show that when the sum c exists, it is unique. The following definition and theorem provide a convenient test for summability.

Definition 2.6 *A family of series $\{c_i\}_{i\in I}$ in $\mathbb{R}^\ell\langle\langle X\rangle\rangle$ is called* ***locally finite*** *if the set $I(\eta)$ is finite for every $\eta \in X^*$.*

Theorem 2.3 *If a family of series $\{c_i\}_{i\in I}$ in $\mathbb{R}^\ell\langle\langle X\rangle\rangle$ is locally finite then it is summable.*

Proof: For a locally finite family $\{c_i\}_{i\in I}$, define the series c whose coefficients are given by the finite summations

$$(c,\eta) = \sum_{i\in I(\eta)} (c_i,\eta), \ \forall \eta \in X^*.$$

To see that c is indeed the sum of $\{c_i\}_{i\in I}$, choose any $\epsilon > 0$, and let $k > 0$ be an integer such that $\sigma^k < \epsilon$. Define the corresponding finite subset of I

$$I_\epsilon = \bigcup_{|\xi|<k} I(\xi).$$

Now let J be any finite subset of I containing I_ϵ. Assuming $c \neq \sum_{j\in J} c_j$ (i.e., the nontrivial case) then it follows that

$$\begin{aligned}
\mathrm{ord}\left(\sum_{j\in J} c_j - c\right) &= \min\left\{|\eta| : \left(\sum_{j\in J} c_j - c, \eta\right) \neq 0, \ \eta \in X^*\right\} \\
&= \min\left\{|\eta| : \sum_{j\in J}(c_j,\eta) \neq (c,\eta), \ \eta \in X^*\right\} \\
&\geq k,
\end{aligned}$$

since for any word $\eta \in X^*$ with $|\eta| < k$

$$\sum_{j\in J}(c_j,\eta) = \sum_{j\in I(\eta)} (c_j,\eta) = (c,\eta).$$

In which case,

$$\operatorname{dist}\left(\sum_{j\in J} c_j, c\right) = \sigma^{\operatorname{ord}\left(c-\sum_{j\in J} c_j\right)} \leq \sigma^k < \epsilon.$$

Thus, the family $\{c_i\}_{i\in I}$ is summable, and c is the corresponding sum.

■

Example 2.11 For any fixed series $c \in \mathbb{R}^\ell\langle\langle X\rangle\rangle$, where X is an arbitrary alphabet, consider the family of series $\{c_\eta\}_{\eta\in X^*}$, where $c_\eta = (c,\eta)\,\eta$. Since the support of each series c_η contains at most one word, the family is locally finite and the sum is obviously c. Furthermore, since these supports are pairwise disjoint, one can unambiguously represent the mapping $c : X^* \to \mathbb{R}^\ell$ using the series notation

$$c = \sum_{\eta\in X^*} c_\eta = \sum_{\eta\in X^*} (c,\eta)\,\eta.$$

□

Example 2.12 Suppose $X = \{x\}$ and consider the family of monomials $\{c_i\}_{i\in\mathbb{N}_0}$, where $c_i = \alpha_i x$ with $\alpha_i \neq 0$ for all $i \in \mathbb{N}_0$. Clearly $I(x) = \mathbb{N}_0$ is not finite. Therefore, the family $\{c_i\}_{i\in\mathbb{N}_0}$ is *not* locally finite. The question of whether the family is summable depends entirely on the coefficients. For example, if $\alpha_i = 1/i$, $i \in \mathbb{N}_0$, then it is not summable as the coefficient of x in the corresponding sum is not finite. On the other hand, if $\alpha_i = 1/i!$, $i \in \mathbb{N}_0$, then this family has the sum $c = \sum_{i\in\mathbb{N}_0} c_i = \mathrm{e}x$, where $\mathrm{e} = 2.7182\ldots$. This demonstrates that the converse of Theorem 2.3 is not true in general. □

Example 2.13 Suppose X is an arbitrary alphabet. Let $\{\alpha_i\}_{i\in\mathbb{N}_0}$ be any sequence of real numbers and $c \in \mathbb{R}\langle\langle X\rangle\rangle$ any proper series. Consider the family of series $\{\alpha_i c^i\}_{i\in\mathbb{N}_0}$. For any $i \in \mathbb{N}_0$ and $\eta \in X^*$ such that $i > |\eta|$, it is immediate that $(c^i,\eta) = 0$ since $\operatorname{ord}(c^i) \geq i$. Therefore, $I(\eta) \subseteq \{0, 1, \ldots, |\eta|\}$. Thus, this family is locally finite, and hence summable. This fact allows one to extend the domain of definition for various real analytic functions from $\mathbb{R}$ to $\mathbb{R}\langle\langle X\rangle\rangle$. For example, given any proper $c \in \mathbb{R}\langle\langle X\rangle\rangle$, one can define:

$$c^* = (1-c)^{-1} = \sum_{i=0}^{\infty} c^i$$
$$\mathrm{e}^c = \sum_{i=0}^{\infty} c^i \, \frac{1}{i!}$$
$$\log(1+c) = \sum_{i=1}^{\infty} c^i \, \frac{(-1)^{i-1}}{i}.$$

The first function above arises naturally in the study of rational series (Chapter 4), while the latter two functions are important when the free Lie algebra on $\mathbb{R}\langle X\rangle$ is considered (Chapter 5). □

The final topic of this section is *contractive mappings*. This is a classic subject in the theory of metric spaces. Contractive mappings often arise in the study of differential equations, operator theory and functional analysis. Contractive mappings will be used in Chapter 3 to determine when the feedback interconnection of two input-output systems is well defined.

Definition 2.7 *Let (S,δ) be a metric space. A mapping $\mathcal{T}: S\to S$ is called a* ***contractive mapping*** *if there exists a real number $0<\alpha<1$ such that*

$$\delta(\mathcal{T}(s),\mathcal{T}(s')) \leq \alpha\,\delta(s,s'), \ \ \forall s,s'\in S.$$

Given any mapping, $\mathcal{T}$, a point $s^*\in S$ is said to be a *fixed point* if $\mathcal{T}(s^*)=s^*$. The following theorem gives a condition under which a fixed point exists and is unique.

Theorem 2.4 *Let (S,δ) be a complete nonempty metric space. Then every contractive mapping $\mathcal{T}: S\to S$ has precisely one fixed point in S.*

Proof: Select any $s_1\in S$ and generate a sequence in S by the iteration $s_{i+1}=\mathcal{T}(s_i)$, $i\geq 1$. It is first shown that if $\mathcal{T}$ is contractive then the sequence $\{s_1,s_2,\ldots\}$ is a Cauchy sequence. Observe that for any $i\geq 2$,

$$\begin{aligned}\delta(s_i,s_{i+1}) &= \delta(\mathcal{T}(s_{i-1}),\mathcal{T}(s_i))\\ &\leq \alpha\,\delta(s_{i-1},s_i)\\ &\leq \alpha^2\delta(s_{i-2},s_{i-1})\\ &\ \ \vdots\end{aligned}$$

$$\leq \alpha^{i-1}\,\delta(s_1, s_2).$$

Applying the triangle inequality and using the assumption that $0 < \alpha < 1$ gives for any $j > i \geq 1$

$$\begin{aligned}
\delta(s_i, s_j) &\leq \delta(s_i, s_{i+1}) + \delta(s_{i+1}, s_{i+2}) + \cdots + \delta(s_{j-1}, s_j) \\
&\leq (\alpha^{i-1} + \alpha^i + \cdots + \alpha^{j-2})\delta(s_1, s_2) \\
&= \alpha^{i-1}\frac{1-\alpha^{j-i}}{1-\alpha}\,\delta(s_1, s_2) \\
&< \frac{\alpha^{i-1}}{1-\alpha}\,\delta(s_1, s_2).
\end{aligned}$$

Clearly, the right-hand side of the last inequality above can be made arbitrarily small by choosing a sufficiently large i. This proves that $\{s_1, s_2, \ldots\}$ is a Cauchy sequence. Since S is assumed to be a complete metric space, there must exist a unique element $s^* \in S$ such that $\lim_{i\to\infty} s_i = s^*$.

It is next shown that s^* is the unique fixed point of $\mathcal{T}$. Observe that

$$\begin{aligned}
\delta(s^*, \mathcal{T}(s^*)) &\leq \delta(s^*, s_i) + \delta(s_i, \mathcal{T}(s^*)) \\
&\leq \delta(s^*, s_i) + \alpha\,\delta(s_{i-1}, s^*).
\end{aligned}$$

However, it was just shown that $\lim_{i\to\infty} s_i = s^*$. So after taking the same limit above, one must conclude that $\delta(s^*, \mathcal{T}(s^*)) = 0$, or equivalently, $s^* = \mathcal{T}(s^*)$. Suppose $\mathcal{T}$ has a second fixed point $\tilde{s}^*$. Then

$$\begin{aligned}
\delta(s^*, \tilde{s}^*) &= \delta(\mathcal{T}(s^*), \mathcal{T}(\tilde{s}^*)) \\
&\leq \alpha\,\delta(s^*, \tilde{s}^*),
\end{aligned}$$

which implies that $\delta(s^*, \tilde{s}^*) = 0$ since $0 < \alpha < 1$. Hence, $s^* = \tilde{s}^*$. ■

Example 2.14 Consider the function

$$f(z) = z + \frac{\pi}{2} - \tan^{-1}(z)$$

on $\mathbb{R}$. The set $\mathbb{R}$ is known to be a complete metric space under the metric $\delta(z, z') = |z - z'|$. Clearly, for every $z \in \mathbb{R}$

$$f'(z) = 1 - \frac{1}{1+z^2} < 1.$$

For any $z, z' \in \mathbb{R}$ with $z < z'$, it follows from the mean-value theorem that there is some $\tilde{z} \in (z, z')$ such that

$$\begin{aligned} |f(z) - f(z')| &= |f'(\tilde{z})|\,|z - z'| \\ &< |z - z'|. \end{aligned} \tag{2.4}$$

This is a slightly *weaker* condition than that required for a contraction. Therefore, Theorem 2.4 does *not* apply. Observe that a fixed point, z^*, would have to satisfy $f(z^*) = z^*$, which in this case is precisely equivalent to requiring that $\tan^{-1}(z^*) = \pi/2$. But no such number z^* exists in $\mathbb{R}$. Hence, (2.4) does not guarantee the existence of a fixed point in general. It does, however, provide for uniqueness when a fixed point is known to exist by other means (see Problem 2.3.6).

2.4 The Shuffle Product

In this section, a new product on $\mathbb{R}^\ell\langle\langle X\rangle\rangle$ is considered, the *shuffle product*. It is probably the most important product after the catenation product for analyzing nonlinear systems in a formal power series setting. The following definition describes the basic idea when only words are involved. But the goal is to eventually define a *shuffle algebra* on $\mathbb{R}^\ell\langle\langle X\rangle\rangle$.

Definition 2.8 *The* ***shuffle*** *of two words $\eta, \xi \in X^*$ is defined to be the language*

$$\begin{aligned} \mathbb{S}_{\eta,\xi} = \{\nu \in X^* : \nu = \eta_1\xi_1\eta_2\xi_2\cdots\eta_n\xi_n,\ \eta_i, \xi_i \in X^*, \\ \eta = \eta_1\eta_2\cdots\eta_n,\ \xi = \xi_1\xi_2\cdots\xi_n,\ n \geq 1\}. \end{aligned}$$

In particular, $\mathbb{S}_{\eta,\emptyset} = \{\eta\}$ *and* $\mathbb{S}_{\emptyset,\xi} = \{\xi\}$.

This shuffle operation most likely derives its name from the manner in which playing cards are mixed, that is, η and ξ are combined so as to preserve the relative ordering of their respective components. For example, the word η_i is to the left of η_{i+1} before and after the shuffle operation is performed. An equivalent definition of the shuffle with a more combinatorial flavor can be given in terms of set bisections. For any integer $n \geq 0$, let $[n] = \{1, 2, \ldots, n\}$ with $[0] = \emptyset$. A pair of subsets (I, J) is a *bisection* of $[n]$ if $I \cup J = [n]$ and $I \cap J = \emptyset$. Given a word $\nu = \nu_1\nu_2\cdots\nu_\ell$ with $\nu_i \in X$ and a subset $I = \{i_1 < i_2 < \cdots < i_k\}$ of $[|\nu|]$, let $\nu_I = \nu_{i_1}\nu_{i_2}\cdots\nu_{i_k}$ denote a *subword* of ν. For example, if

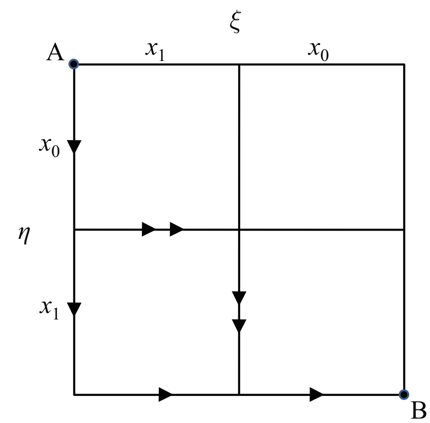


Fig. 2.2. Two paths corresponding to the word $x_0x_1^2x_0$ in the shuffle language in Example 2.15.

$\nu = x_0x_1x_0x_1$ and $I = \{2,4\}$ then $\nu_I = x_1^2$. It is easily verified that $\nu \in \mathbb{S}_{\eta,\xi}$ if and only if there exists a bisection (I,J) of $[|\nu|]$ such that $\nu_I = \eta$ and $\nu_J = \xi$. Therefore, any $\nu \in \mathbb{S}_{\eta,\xi}$ satisfies $|\nu| = |\eta| + |\xi|$, and, in general $\mathbb{S}_{\eta,\xi}$ can contain at most

$$\binom{|\eta| + |\xi|}{|\eta|}$$

distinct words.

Example 2.15 Suppose $X = \{x_0, x_1\}$, $\eta = x_0x_1$ and $\xi = x_1x_0$. A systematic way to construct a word in the shuffle language $\mathbb{S}_{\eta,\xi}$ is to create a table as shown in Figure 2.2, where the rows are labeled from top to bottom with the letters of η, and the columns are labeled from left to right with the letters of ξ. Consider a path connecting point A to point B, where one is only permitted to move down and to the right. With each path one can associate a single word in $\mathbb{S}_{\eta,\xi}$ by keeping track of which rows and columns have been traversed. For example, the two paths corresponding to the word $x_0x_1^2x_0$ are shown in the figure. Proceeding in this way, the complete shuffle language is found to be

$$\mathbb{S}_{\eta,\xi} = \{x_0x_1^2x_0,\ x_0x_1x_0x_1,\ x_1x_0x_1x_0,\ x_1x_0^2x_1\}.$$

□

A third equivalent definition of the shuffle can be given recursively in terms of formal polynomials. This approach has two advantages: it is more computational in nature, and it is easier to determine when words have a multiplicity greater than one, i.e. when they can be generated by more than one bisection or path in the context of the previous definitions.

Definition 2.9 *The **shuffle product** of two words is*

$$(x_i\eta) \sqcup\!\sqcup (x_j\xi) = x_i(\eta \sqcup\!\sqcup (x_j\xi)) + x_j((x_i\eta) \sqcup\!\sqcup \xi), \tag{2.5}$$

where $x_i, x_j \in X$, $\eta, \xi \in X^*$ *and with* $\eta \sqcup\!\sqcup \emptyset = \emptyset \sqcup\!\sqcup \eta = \eta$.[3]

The claim, which is left to the reader to verify, is that

$$\mathbb{S}_{\eta,\xi} = \mathrm{supp}(\eta \sqcup\!\sqcup \xi).$$

Example 2.16 Reconsider the previous example where $\eta = x_0x_1$ and $\xi = x_1x_0$. Then

$$\begin{aligned} x_0x_1 \sqcup\!\sqcup x_1x_0 &= x_0(x_1 \sqcup\!\sqcup x_1x_0) + x_1(x_0x_1 \sqcup\!\sqcup x_0) \\ &= x_0[x_1^2x_0 + x_1(x_1 \sqcup\!\sqcup x_0)] + x_1[x_0(x_1 \sqcup\!\sqcup x_0) + x_0^2x_1] \\ &= 2x_0x_1^2x_0 + x_0x_1x_0x_1 + x_1x_0x_1x_0 + 2x_1x_0^2x_1. \end{aligned}$$

Thus, six words are generated by this product, but two of them have multiplicity 2. Furthermore,

$$\begin{aligned} \mathrm{supp}(x_0x_1 \sqcup\!\sqcup x_1x_0) &= \{x_0x_1^2x_0,\ x_0x_1x_0x_1,\ x_1x_0x_1x_0,\ x_1x_0^2x_1\} \\ &= \mathbb{S}_{\eta,\xi}. \end{aligned}$$

□

Example 2.17 Given a language $L \subseteq X^*$, the *characteristic series* of L is the element in $\mathbb{R}\langle\langle X\rangle\rangle$ defined by $\mathrm{char}(L) = \sum_{\nu \in L} \nu$. Suppose, for example, $X = \{x_0, x_1\}$. Then

$$\begin{aligned} \mathrm{char}(X) &= x_0 + x_1 = x_0 \sqcup\!\sqcup \emptyset + \emptyset \sqcup\!\sqcup x_1 \\ &= \sum_{\substack{r_0, r_1 \geq 0 \\ r_0 + r_1 = 1}} x_0^{r_0} \sqcup\!\sqcup x_1^{r_1}. \end{aligned}$$

[3] The symbol $\sqcup\!\sqcup$ is the letter *Sha* in the Cyrillic alphabet, which is used, for example, in the Bulgarian, Russian and Ukrainian written languages. It is used to represent a sound which is roughly equivalent to the *sh* sound in the English word *shuffle*.

Similarly,

$$\begin{aligned}\mathrm{char}(X^2) &= x_0^2 + x_0x_1 + x_1x_0 + x_1^2 \\ &= x_0^2 \,\sqcup\!\sqcup\, \emptyset + x_0 \,\sqcup\!\sqcup\, x_1 + \emptyset \,\sqcup\!\sqcup\, x_1^2 \\ &= \sum_{\substack{r_0,r_1\geq 0\\ r_0+r_1=2}} x_0^{r_0} \,\sqcup\!\sqcup\, x_1^{r_1}.\end{aligned}$$

An inductive argument for an arbitrary alphabet $X = \{x_0, x_1, \ldots, x_m\}$ produces the useful identity

$$\mathrm{char}(X^k) = \sum_{\substack{r_0,r_1,\ldots,r_m\geq 0\\ r_0+r_1+\cdots+r_m=k}} x_0^{r_0} \,\sqcup\!\sqcup\, x_1^{r_1} \,\sqcup\!\sqcup\, \cdots \,\sqcup\!\sqcup\, x_m^{r_m}, \quad k \geq 0 \tag{2.6}$$

(see Problem 2.4.6). □

The definition of the shuffle product is extended linearly to any two series $c, d \in \mathbb{R}\langle\langle X\rangle\rangle$ by letting

$$c \,\sqcup\!\sqcup\, d = \sum_{\eta,\xi\in X^*} (c,\eta)(d,\xi)\, \eta \,\sqcup\!\sqcup\, \xi. \tag{2.7}$$

For a fixed $\nu \in X^*$, the coefficient

$$(c \,\sqcup\!\sqcup\, d, \nu) = (c,\eta)(d,\xi)(\eta \,\sqcup\!\sqcup\, \xi, \nu) = 0, \quad |\eta| + |\xi| \neq |\nu|.$$

Hence, the infinite sum in (2.7) is well defined since the family of polynomials $\{\eta \,\sqcup\!\sqcup\, \xi\}_{(\eta,\xi)\in X^*\times X^*}$ is locally finite, i.e., $I(\nu) \subseteq \{(\eta,\xi) \in X^* \times X^* : |\eta| + |\xi| = |\nu|\}$ is finite for every $\nu \in X^*$. Given two series $c, d \in \mathbb{R}^\ell\langle\langle X\rangle\rangle$ the shuffle product $c \,\sqcup\!\sqcup\, d$ is defined componentwise, i.e., the i-th component series of $c \,\sqcup\!\sqcup\, d$ is $(c \,\sqcup\!\sqcup\, d)_i = c_i \,\sqcup\!\sqcup\, d_i$, where $1 \leq i \leq \ell$.

Example 2.18 Reconsider Example 2.6 where $X = \{x_0, x_1\}$ and

$$c = 2x_0x_1, \quad d = x_0 + x_1.$$

Observe that

$$\begin{aligned}c \,\sqcup\!\sqcup\, d &= 2[x_0x_1 \,\sqcup\!\sqcup\, x_0] + 2[x_0x_1 \,\sqcup\!\sqcup\, x_1] \\ &= 2\,(x_0[x_1 \,\sqcup\!\sqcup\, x_0] + x_0[x_0x_1 \,\sqcup\!\sqcup\, \emptyset]) + 2\,(x_0[x_1 \,\sqcup\!\sqcup\, x_1] + x_1[x_0x_1 \,\sqcup\!\sqcup\, \emptyset]) \\ &= 2x_0(x_1x_0 + x_0x_1 + x_0x_1) + 2(2x_0x_1^2 + x_1x_0x_1)\end{aligned}$$

$$= 4x_0^2x_1 + 2x_0x_1x_0 + 2x_1x_0x_1 + 4x_0x_1^2$$

and

$$\begin{aligned}
d \sqcup\!\sqcup\, c &= 2[x_0 \sqcup\!\sqcup\, x_0x_1] + 2[x_1 \sqcup\!\sqcup\, x_0x_1] \\
&= 2\left(x_0[\emptyset \sqcup\!\sqcup\, x_0x_1] + x_0[x_0 \sqcup\!\sqcup\, x_1]\right) + 2\left(x_1[\emptyset \sqcup\!\sqcup\, x_0x_1] + x_0[x_1 \sqcup\!\sqcup\, x_1]\right) \\
&= 2(x_0^2x_1 + x_0(x_0x_1 + x_1x_0)) + 2(x_1x_0x_1 + 2x_0x_1^2) \\
&= 4x_0^2x_1 + 2x_0x_1x_0 + 2x_1x_0x_1 + 4x_0x_1^2 \\
&= c \sqcup\!\sqcup\, d.
\end{aligned}$$

□

In general, the shuffle product is commutative. It is also associative and distributes over addition. Thus, $\mathbb{R}\langle\langle X\rangle\rangle$ forms a commutative $\mathbb{R}$-algebra, the *shuffle algebra*, with multiplicative identity element 1. In addition, the shuffle algebra is an *integral domain*, namely, it has the property that $c \sqcup\!\sqcup\, d = 0$ if and only if at least one of its arguments is the zero series (see Problem 2.4.1). This fact yields another basic property of the shuffle product. Using the property that $\mathrm{ord}(c + d) \geq \min(\mathrm{ord}(c), \mathrm{ord}(d))$, it is easy to see in general that $\mathrm{ord}(c \sqcup\!\sqcup\, d) \geq \mathrm{ord}(c) + \mathrm{ord}(d)$. But the following stronger claim holds.

Lemma 2.2 *For any $c, d \in \mathbb{R}\langle\langle X\rangle\rangle$, $\mathrm{ord}(c \sqcup\!\sqcup\, d) = \mathrm{ord}(c) + \mathrm{ord}(d)$.*

Proof: Consider only the nontrivial case where both c and d are not zero. Define the nonzero polynomials $p, q \in \mathbb{R}\langle X\rangle$ by $c = p + c'$ and $d = q + d'$, where p is the homogeneous part of c satisfying $\mathrm{ord}(p) = \mathrm{ord}(c)$ and $\mathrm{ord}(c') = \mathrm{ord}(c - p) > \mathrm{ord}(c)$, and likewise for q. In which case,

$$c \sqcup\!\sqcup\, d = (p + c') \sqcup\!\sqcup\, (q + d') = p \sqcup\!\sqcup\, q + p \sqcup\!\sqcup\, d' + c' \sqcup\!\sqcup\, q + c' \sqcup\!\sqcup\, d'.$$

As the shuffle algebra is an integral domain, it is immediate that $p \sqcup\!\sqcup\, q \neq 0$ and $\mathrm{ord}(p \sqcup\!\sqcup\, q) = \mathrm{ord}(p) + \mathrm{ord}(q) = \mathrm{ord}(c) + \mathrm{ord}(d)$. Furthermore, each of the last three terms has order exceeding $\mathrm{ord}(c) + \mathrm{ord}(d)$, for example, $\mathrm{ord}(p \sqcup\!\sqcup\, d') \geq \mathrm{ord}(p) + \mathrm{ord}(d') > \mathrm{ord}(c) + \mathrm{ord}(d)$. Therefore, the identity in question must hold. ■

As alluded to earlier, the shuffle algebra plays a central role in the analysis of nonlinear systems. To better understand this connection, let $\{u_0, u_1, \ldots, u_m\}$ be a fixed set of piecewise continuous real-valued functions defined on a finite interval $[t_0, t_1]$. For any word $\eta \in X^+$ define recursively the iterated integral

$$E_\eta[u](t,t_0) = E_{x_i\eta'}[u](t,t_0) = \int_{t_0}^{t} u_i(\tau)E_{\eta'}[u](\tau,t_0)\,d\tau$$

with $E_\emptyset[u](t,t_0) = 1$ for all $t \in [t_0, t_1]$. For any polynomial $p \in \mathbb{R}\langle X\rangle$ extend this definition as

$$E_p[u](t,t_0) = \sum_{\eta\in X^*} (p,\eta)\, E_\eta[u](t,t_0).$$

Let $\mathcal{E}(\mathbb{R}\langle X\rangle)$ denote the set of all such finite linear combinations of iterated integrals. Clearly, $\mathcal{E}(\mathbb{R}\langle X\rangle)$ forms an $\mathbb{R}$-vector space. The following lemma provides additional algebraic structure.

Lemma 2.3 *For any* $\eta, \xi \in X^*$

$$E_\eta[u](t,t_0)E_\xi[u](t,t_0) = E_{\eta \,\sqcup\!\sqcup\, \xi}[u](t,t_0).$$

Proof: The claim is trivially true when $|\eta| + |\xi| = 0$ and $|\eta| + |\xi| = 1$. Assume it holds up to the case where $|\eta| + |\xi| = n \geq 1$, and suppose for example that ξ is nonempty. Then via integration by parts formula

$$UV = \int dU\,V + \int U\,dV$$

it follows for any $x_i \in X$ that

$$\begin{aligned} E_{x_i\eta}[u](t,t_0)\; E_\xi[u](t,t_0) &= \int_{t_0}^{t} u_i(\tau)E_\eta[u](\tau,t_0)\; E_\xi[u](\tau,t_0)\,d\tau + \\ &\qquad \int_{t_0}^{t} u_j(\tau)E_{x_i\eta}[u](\tau,t_0)\; E_{\xi'}[u](\tau,t_0)\,d\tau \\ &= E_{x_i(\eta \,\sqcup\!\sqcup\, \xi)+x_j((x_i\eta) \,\sqcup\!\sqcup\, \xi')}[u](t,t_0) \\ &= E_{(x_i\eta) \,\sqcup\!\sqcup\, \xi}[u](t,t_0), \end{aligned}$$

where $\xi = x_j\xi'$. Hence, by induction, the identity holds for all words $\eta, \xi \in X^*$. ■

The vector space $\mathcal{E}(\mathbb{R}\langle X\rangle)$ thus forms an associative and commutative $\mathbb{R}$-algebra with product $E_pE_q = E_{p \,\sqcup\!\sqcup\, q}$ and multiplicative identity element $E_1 = 1$. It will be shown in Chapter 3 (see Theorems 3.5 and 3.38) that the mapping

$$\begin{aligned}\rho : \mathcal{E}(\mathbb{R}\langle X\rangle) &\to \mathbb{R}\langle X\rangle \\ : E_p &\mapsto p\end{aligned}$$

is well defined and bijective. Thus, this $\mathbb{R}$-algebra is isomorphic to the shuffle $\mathbb{R}$-algebra on $\mathbb{R}\langle X\rangle$ since

$$\begin{aligned}\rho(E_p E_q) &= \rho(E_{p \sqcup\!\sqcup q}) = p \sqcup\!\sqcup q \\ &= \rho(E_p) \sqcup\!\sqcup \rho(E_q)\end{aligned}$$

for all $p, q \in \mathbb{R}\langle X\rangle$ and $\rho(E_1) = 1$. This *shuffle isomorphism*, or more precisely its extension in Chapter 3 to $\mathbb{R}^\ell\langle\langle X\rangle\rangle$, means that the shuffle product underlies any calculation that involves the product of iterated integrals. As will be seen shortly, this happens naturally when input-output systems are interconnected.

The section is concluded by describing how the left-shift operator interacts with the shuffle product.

Theorem 2.5 *The left-shift operator acts as a derivation on the shuffle product, i.e., for $c, d \in \mathbb{R}\langle\langle X\rangle\rangle$ and $x_k \in X$*

$$x_k^{-1}(c \sqcup\!\sqcup d) = x_k^{-1}(c) \sqcup\!\sqcup d + c \sqcup\!\sqcup x_k^{-1}(d).$$

Proof: First consider the identity when restricted to words $\eta, \xi \in X^*$. If either word is the empty word then the claim is trivial. If $\eta, \xi \in X^+$ then let $\eta = x_i\eta'$ and $\xi = x_j\xi'$ and observe for any $x_k \in X$ that

$$\begin{aligned}x_k^{-1}(\eta \sqcup\!\sqcup \xi) &= x_k^{-1}(x_i(\eta' \sqcup\!\sqcup \xi) + x_j(\eta \sqcup\!\sqcup \xi')) \\ &= \delta_{ki}(\eta' \sqcup\!\sqcup \xi) + \delta_{kj}(\eta \sqcup\!\sqcup \xi') \\ &= (\delta_{ki}\eta') \sqcup\!\sqcup \xi + \eta \sqcup\!\sqcup (\delta_{kj}\xi') \\ &= x_k^{-1}(\eta) \sqcup\!\sqcup \xi + \eta \sqcup\!\sqcup x_k^{-1}(\xi),\end{aligned}$$

where

$$\delta_{ij} = \begin{cases} 1 : i = j \\ 0 : \text{otherwise.}\end{cases}$$

Now in the general case,

$$x_k^{-1}(c \sqcup\!\sqcup d) = \sum_{\eta,\xi \in X^*} (c, \eta)(d, \xi)\, x_k^{-1}(\eta \sqcup\!\sqcup \xi)$$

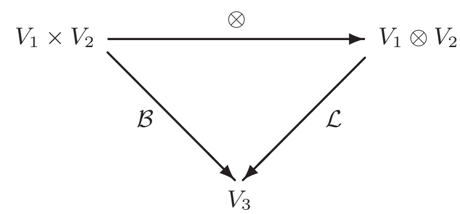


Fig. 2.3. The commutative diagram for the mappings $\mathcal{B}$ and $\mathcal{L}$ in Definition 2.10.

$$= \sum_{\eta,\xi \in X^*} (c,\eta)(d,\xi)\, x_k^{-1}(\eta) \sqcup\!\sqcup\, \xi + \sum_{\eta,\xi \in X^*} (c,\eta)(d,\xi)\, \eta \sqcup\!\sqcup\, x_k^{-1}(\xi)$$
$$= x_k^{-1}(c) \sqcup\!\sqcup\, d + c \sqcup\!\sqcup\, x_k^{-1}(d).$$

■

2.5 Catenation-Shuffle Product Duality

In this section, a duality is presented between the catenation product and the shuffle product when each is viewed as a linear mapping on a tensor product space. A few preliminary concepts need to be established first before the precise sense of this duality can be described.

Let V_1, V_2, and V_3 be three arbitrary vector spaces over $\mathbb{R}$. Consider an $\mathbb{R}$-bilinear map of the form $\mathcal{B} : V_1 \times V_2 \to V_3$, that is, a map where

$$\mathcal{B}(\alpha w + \beta x, y) = \alpha \mathcal{B}(w, y) + \beta \mathcal{B}(x, y)$$
$$\mathcal{B}(x, \alpha y + \beta z) = \alpha \mathcal{B}(x, y) + \beta \mathcal{B}(x, z)$$

for all $\alpha, \beta \in \mathbb{R}$, $w, x \in V_1$ and $y, z \in V_2$. In this context, consider the following definition.

Definition 2.10 *The* ***tensor product space*** $V_1 \otimes V_2$ *is another vector space on which there exists a unique* $\mathbb{R}$*-linear mapping* $\mathcal{L} : V_1 \otimes V_2 \to V_3$ *such that*

$$\mathcal{B}(x, y) = \mathcal{L}(x \otimes y), \quad \forall x \in V_1,\ y \in V_2 \tag{2.8}$$

(see Figure 2.3).

Example 2.19 Let $V_1 = V_2 = \mathbb{R}^2$ and $V_3 = \mathbb{R}$ with their usual $\mathbb{R}$-vector space structures. For any matrix $A \in \mathbb{R}^{2\times 2}$, a corresponding $\mathbb{R}$-bilinear mapping of the form $\mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ is

$$\begin{aligned}
\mathcal{B} : (x,y) \mapsto x^T A^T y &= a_{11}x_1y_1 + a_{21}x_1y_2 + a_{12}x_2y_1 + a_{22}x_2y_2 \\
&= \left[\begin{array}{cccc} a_{11} & a_{21} & a_{12} & a_{22} \end{array}\right] \left[\begin{array}{c} x_1y_1 \\ x_1y_2 \\ x_2y_1 \\ x_2y_2 \end{array}\right] \\
&= \mathrm{vec}(A)^T(x \otimes y),
\end{aligned}$$

where $\mathrm{vec}(\cdot)$ is the matrix column stacking operator, and $\otimes$ is the Kronecker matrix product. Therefore $\mathcal{B}$ can be made to look like a map of the form $\mathbb{R}^4 \to \mathbb{R}$ in which

$$x \otimes y + x' \otimes y = (x + x') \otimes y \tag{2.9}$$

$$x \otimes y + x \otimes y' = x \otimes (y + y') \tag{2.10}$$

$$\alpha(x \otimes y) = (\alpha x) \otimes y = x \otimes (\alpha y) \tag{2.11}$$

for all $x, x', y, y' \in \mathbb{R}^2$ and $\alpha \in \mathbb{R}$. Taking $e_1 = [1\,0]^T$ and $e_2 = [0\,1]^T$ as a basis for $\mathbb{R}^2$, a corresponding basis for the vector space $\mathbb{R}^2 \otimes \mathbb{R}^2$ is $\{e_1 \otimes e_1, e_1 \otimes e_2, e_2 \otimes e_1, e_2 \otimes e_2\}$ so that

$$x \otimes y = \sum_{i,j=1}^{2} x_i y_j (e_i \otimes e_j). \tag{2.12}$$

In this coordinate system, vector addition and scalar multiplication are defined in the usual way to provide an $\mathbb{R}$-vector space structure for $\mathbb{R}^2 \otimes \mathbb{R}^2$. It is then straightforward to verify that

$$\begin{aligned}
\mathcal{L} &: \mathbb{R}^2 \otimes \mathbb{R}^2 \to \mathbb{R} \\
&: x \otimes y \mapsto \mathrm{vec}(A)^T(x \otimes y)
\end{aligned} \tag{2.13}$$

is an $\mathbb{R}$-linear map satisfying the identity (2.8) (see Problem 2.5.1). □

Next consider the following two concepts.

Definition 2.11 *A* ***scalar product*** *on an $\mathbb{R}$-vector space V is an $\mathbb{R}$-bilinear mapping $(\cdot,\cdot)_V : V \times V \to \mathbb{R}$ with the following properties:*

i. $(x,x)_V \geq 0$, $\forall x \in V$

ii. $(x,x)_V = 0$ *if and only if* $x = 0$

iii. $(x,x')_V = (x',x)_V$, $\forall x,x' \in V$.

Example 2.20 Suppose $V_1 = \mathbb{R}^2$ with the scalar product $(x,x')_{V_1} = x^T x'$. Observe from the previous example that (2.12) can be written as

$$x \otimes y = \sum_{i,j=1}^{2} (x \otimes y, e_i \otimes e_j)\, e_i \otimes e_j,$$

where $(x \otimes y, e_i \otimes e_j) := (x^T e_i)(y^T e_j) = x_i y_j$. This in turn induces a scalar product on $V_2 = \mathbb{R}^2 \otimes \mathbb{R}^2$, namely,

$$\begin{aligned}
(x \otimes y, x' \otimes y')_{V_2} &= \sum_{i,j=1}^{2} (x \otimes y, e_i \otimes e_j)(x' \otimes y', e_i \otimes e_j) \\
&= \sum_{i,j=1}^{2} x_i y_j x_i' y_j' \\
&= (x,x')_{V_1}(y,y')_{V_1}.
\end{aligned}$$

□

Definition 2.12 *Given any* $\mathbb{R}$*-linear mapping* $\mathcal{T} : V_1 \to V_2$*, where each vector space has a scalar product, an* ***adjoint*** *of* $\mathcal{T}$ *is any* $\mathbb{R}$*-linear map* $\mathcal{T}^* : V_2 \to V_1$ *which satisfies the identity*

$$(\mathcal{T}(x), y)_{V_2} = (x, \mathcal{T}^*(y))_{V_1}, \quad \forall x \in V_1, y \in V_2.$$

The following example illustrates that for finite dimensional spaces such an adjoint map always exists and corresponds uniquely to the transpose of any matrix representation of $\mathcal{T}$.

Example 2.21 Suppose $V_1 = \mathbb{R}^n$ and $V_2 = \mathbb{R}^m$ with the respective scalar products

$$\begin{aligned}
(x_1,x_2)_{\mathbb{R}^n} &= x_1^T x_2 \\
(y_1,y_2)_{\mathbb{R}^m} &= y_1^T y_2.
\end{aligned}$$

If $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^m : x \mapsto y = Ax$ for some matrix $A \in \mathbb{R}^{m \times n}$ then the adjoint of $\mathcal{A}$ must satisfy

$$(Ax, y)_{\mathbb{R}^m} = x^T A^T y = (x, A^T y)_{\mathbb{R}^n}.$$

Therefore, $\mathcal{A}^* : y \mapsto A^T y$. □

The desired catenation-shuffle duality is now described using scalar products on the infinite dimensional vector spaces $V_1 = \mathbb{R}\langle X \rangle$ and $V_2 = \mathbb{R}\langle X \rangle \otimes \mathbb{R}\langle X \rangle$. Their spanning sets are taken to be X^* and $X^* \otimes X^*$, respectively, so that

$$\begin{aligned}(p, q)_{V_1} &= \sum_{\eta \in X^*} (p, \eta)(q, \eta) \\ (p \otimes q, r \otimes s)_{V_2} &= \sum_{\eta, \eta' \in X^*} (p \otimes q, \eta \otimes \eta')(r \otimes s, \eta \otimes \eta') \\ &= \sum_{\eta, \eta' \in X^*} (p, \eta)(q, \eta')(r, \eta)(s, \eta') \\ &= (p, r)_{V_1}(q, s)_{V_1}.\end{aligned}$$

(Henceforth, the subscripts on these scalar products will be omitted.) Also observe that $X^* \otimes X^*$ forms a monoid under the catenation product $(\eta \otimes \eta')(\xi \otimes \xi') = (\eta\xi \otimes \eta'\xi')$, which can be extended linearly so that $\mathbb{R}\langle X \rangle \otimes \mathbb{R}\langle X \rangle$ is an $\mathbb{R}$-algebra under this catenation product. Likewise, one can define $(\eta \otimes \eta') \sqcup\!\sqcup (\xi \otimes \xi') = (\eta \sqcup\!\sqcup \xi) \otimes (\eta' \sqcup\!\sqcup \xi')$ to yield a shuffle algebra on $\mathbb{R}\langle X \rangle \otimes \mathbb{R}\langle X \rangle$.

The catenation product and the shuffle product can now be identified, respectively, with the $\mathbb{R}$-linear mappings:

$$\begin{aligned}\mathrm{cat} &: \mathbb{R}\langle X \rangle \otimes \mathbb{R}\langle X \rangle \to \mathbb{R}\langle X \rangle \\ &: p \otimes q \mapsto pq\end{aligned}$$

$$\begin{aligned}\mathrm{sh} &: \mathbb{R}\langle X \rangle \otimes \mathbb{R}\langle X \rangle \to \mathbb{R}\langle X \rangle \\ &: p \otimes q \mapsto p \sqcup\!\sqcup q.\end{aligned}$$

The corresponding adjoint mappings cat^* and sh^* are then $\mathbb{R}$-linear mappings of the form $\mathbb{R}\langle X \rangle \to \mathbb{R}\langle X \rangle \otimes \mathbb{R}\langle X \rangle$ which satisfy, respectively, the identities:

$$(\mathrm{cat}(p \otimes q), r) = (p \otimes q, \mathrm{cat}^*(r)) \tag{2.14}$$

$$(\mathrm{sh}(p \otimes q), r) = (p \otimes q, \mathrm{sh}^*(r)) \tag{2.15}$$

for all $p, q, r \in \mathbb{R}\langle X\rangle$. Explicit expressions for sh^* and cat^* can be derived directly from these relations. For example, given any $\xi, \nu \in X^*$ and $r \in \mathbb{R}\langle X\rangle$ it follows that

$$\begin{aligned}(\xi \sqcup\!\sqcup \nu, r) &= (\mathrm{sh}(\xi \otimes \nu), r) = (\xi \otimes \nu, \mathrm{sh}^*(r)) \\ &= (\mathrm{sh}^*(r), \xi \otimes \nu).\end{aligned}$$

Therefore,

$$\begin{aligned}\mathrm{sh}^*(r) &= \sum_{\xi,\nu\in X^*} (\mathrm{sh}^*(r), \xi \otimes \nu)\, \xi \otimes \nu \\ &= \sum_{\xi,\nu\in X^*} (r, \xi \sqcup\!\sqcup \nu)\, \xi \otimes \nu. \end{aligned} \tag{2.16}$$

A similar analysis reveals that

$$\mathrm{cat}^*(r) = \sum_{\xi,\nu\in X^*} (r, \xi\nu)\, \xi \otimes \nu. \tag{2.17}$$

Example 2.22 Using (2.16) and (2.17), respectively, observe that for $x_{i_j} \in X^*$:

$$\begin{aligned}\mathrm{sh}^*(\mathbf{1}) &= \mathbf{1} \otimes \mathbf{1} \\ \mathrm{sh}^*(x_{i_1}) &= x_{i_1} \otimes \mathbf{1} + \mathbf{1} \otimes x_{i_1} \\ \mathrm{sh}^*(x_{i_2}x_{i_1}) &= x_{i_2}x_{i_1} \otimes \mathbf{1} + x_{i_2} \otimes x_{i_1} + x_{i_1} \otimes x_{i_2} + \mathbf{1} \otimes x_{i_2}x_{i_1} \\ &= (x_{i_2} \otimes \mathbf{1} + \mathbf{1} \otimes x_{i_2})(x_{i_1} \otimes \mathbf{1} + \mathbf{1} \otimes x_{i_1}) \\ &= \mathrm{sh}^*(x_{i_2})\mathrm{sh}^*(x_{i_1}) \\ &\ \vdots \\ \mathrm{sh}^*(x_{i_k} \cdots x_{i_1}) &= \mathrm{sh}^*(x_{i_k}) \cdots \mathrm{sh}^*(x_{i_1})\end{aligned}$$

and

$$\begin{aligned}\mathrm{cat}^*(\mathbf{1}) &= \mathbf{1} \otimes \mathbf{1} \\ \mathrm{cat}^*(x_{i_1}) &= x_i \otimes \mathbf{1} + \mathbf{1} \otimes x_{i_1} \\ \mathrm{cat}^*(x_{i_1}x_{i_2}) &= x_{i_1}x_{i_2} \otimes \mathbf{1} + x_{i_1} \otimes x_{i_2} + \mathbf{1} \otimes x_{i_1}x_{i_2} \\ &\ \vdots \\ \mathrm{cat}^*(x_{i_1} \cdots x_{i_k}) &= x_{i_1} \cdots x_{i_k} \otimes \mathbf{1} + \mathbf{1} \otimes x_{i_1} \cdots x_{i_k} +\end{aligned}$$

$$\sum_{j=1}^{k-1} x_{i_1}\cdots x_{i_j}\otimes x_{i_{j+1}}\cdots x_{i_k},$$

where **1** denotes the unit polynomial $1\emptyset$. Note that this last identity above can be written in the inductive form

$$\text{cat}^*(x_i\eta) = (x_i\otimes \mathbf{1})\text{cat}^*(\eta) + \mathbf{1}\otimes x_i\eta$$

(see Problem 2.5.2). □

The next theorem states the desired duality in terms of $\mathbb{R}$-algebra homomorphisms.[4]

Theorem 2.6 *The adjoint map* sh* *is an* $\mathbb{R}$*-algebra homomorphism for the catenation product* cat*, and the adjoint map* cat* *is an* $\mathbb{R}$*-algebra homomorphism for the shuffle product* sh*. Specifically, this means that*

$$\text{sh}^*(pq) = \text{sh}^*(p)\,\text{sh}^*(q) \tag{2.18}$$
$$\text{cat}^*(p ⧢ q) = \text{cat}^*(p) ⧢ \text{cat}^*(q) \tag{2.19}$$

for all $p,q\in\mathbb{R}\langle X\rangle$.

Proof: It is shown that sh* is an $\mathbb{R}$-algebra homomorphism for the catenation product by first showing via induction that for all $k\geq 1$ and $x_{i_j}\in X$

$$\begin{aligned}\text{sh}^*(x_{i_k}\cdots x_{i_1}) &= (x_{i_k}\otimes\mathbf{1}+\mathbf{1}\otimes x_{i_k})\text{sh}^*(x_{i_{k-1}}\cdots x_{i_1})\\ &= \text{sh}^*(x_{i_k})\text{sh}^*(x_{i_{k-1}}\cdots x_{i_1}).\end{aligned} \tag{2.20}$$

In which case, as indicated in the previous example,

$$\text{sh}^*(x_{i_k}\cdots x_{i_1}) = \text{sh}^*(x_{i_k})\cdots\text{sh}^*(x_{i_1}).$$

Therefore, (2.18) holds for all words since the identity clearly holds if one or both words are empty. The $k=1$ case is trivial. If the claim holds for some fixed $k\geq 1$ then it follows that

$$\text{sh}^*(x_{i_{k+1}}x_{i_k}\cdots x_{i_1}) = \sum_{\xi,\nu\in X^*}(x_{i_{k+1}}x_{i_k}\cdots x_{i_1}, \xi ⧢ \nu)\,\xi\otimes\nu$$

[4] This concept will be defined more precisely in Definition 2.13.

$$
\begin{aligned}
&= \sum_{\xi,\nu\in X^*} (x_{i_k}\cdots x_{i_1}, x_{i_{k+1}}^{-1}(\xi \sqcup\!\sqcup\, \nu))\ \xi\otimes\nu \\
&= \sum_{\xi,\nu\in X^*} \Big[(x_{i_k}\cdots x_{i_1}, x_{i_{k+1}}^{-1}(\xi)\sqcup\!\sqcup\,\nu)+ \\
&\qquad (x_{i_k}\cdots x_{i_1}, \xi\sqcup\!\sqcup\, x_{i_{k+1}}^{-1}(\nu))\Big]\ \xi\otimes\nu \\
&= \sum_{\xi,\nu\in X^*} (x_{i_k}\cdots x_{i_1}, \xi\sqcup\!\sqcup\,\nu)\ (x_{i_{k+1}}\xi)\otimes\nu + \\
&\qquad \sum_{\xi,\nu\in X^*} (x_{i_k}\cdots x_{i_1}, \xi\sqcup\!\sqcup\,\nu)\ \xi\otimes(x_{i_{k+1}}\nu) \\
&= (x_{i_{k+1}}\otimes\mathbf{1})\,\mathrm{sh}^*(x_{i_k}\cdots x_{i_1}) + \\
&\qquad (\mathbf{1}\otimes x_{i_{k+1}})\,\mathrm{sh}^*(x_{i_k}\cdots x_{i_1}) \\
&= \mathrm{sh}^*(x_{i_{k+1}})\,\mathrm{sh}^*(x_{i_k}\cdots x_{i_1}).
\end{aligned}
$$

Thus, one must conclude that (2.20) holds for all $k\geq 1$. Finally, by linearity, $\mathrm{sh}^*(pq)=\mathrm{sh}^*(p)\,\mathrm{sh}^*(q)$ for all $p,q\in\mathbb{R}\langle X\rangle$.

Next it is shown that cat^* is a shuffle algebra homomorphism. For any $\eta,\eta'\in X^*$, equation (2.16) and the above result imply that

$$
\mathrm{sh}^*(\eta\eta') = \sum_{\xi,\nu\in X^*} (\eta\eta', \xi\sqcup\!\sqcup\,\nu)\ \xi\otimes\nu \tag{2.21}
$$

$$
\begin{aligned}
&= \left[\sum_{\xi,\nu\in X^*} (\eta, \xi\sqcup\!\sqcup\,\nu)\ \xi\otimes\nu\right]\cdot\left[\sum_{\xi',\nu'\in X^*} (\eta', \xi'\sqcup\!\sqcup\,\nu')\ \xi'\otimes\nu'\right]
\end{aligned}
$$

$$
= \sum_{\xi,\nu,\xi',\nu'\in X^*} (\eta,\xi\sqcup\!\sqcup\,\nu)(\eta',\xi'\sqcup\!\sqcup\,\nu')\ \xi\xi'\otimes\nu\nu'. \tag{2.22}
$$

Taking the scalar product of the right-hand sides of equations (2.21) and (2.22) with $\bar\xi\otimes\bar\nu$ gives

$$
(\eta\eta',\bar\xi\sqcup\!\sqcup\,\bar\nu) = \sum_{\xi,\nu,\xi',\nu'\in X^*} (\eta,\xi\sqcup\!\sqcup\,\nu)(\eta',\xi'\sqcup\!\sqcup\,\nu')(\bar\xi,\xi\xi')(\bar\nu,\nu\nu').
$$

This identity, combined with (2.17), produces the following:

$$
\begin{aligned}
\mathrm{cat}^*(\bar\xi\sqcup\!\sqcup\,\bar\nu) &= \sum_{\eta,\eta'\in X^*} (\bar\xi\sqcup\!\sqcup\,\bar\nu, \eta\eta')\ \eta\otimes\eta' \\
&= \sum_{\xi,\eta,\nu,\xi',\eta',\nu'\in X^*} (\eta,\xi\sqcup\!\sqcup\,\nu)(\eta',\xi'\sqcup\!\sqcup\,\nu')(\bar\xi,\xi\xi')(\bar\nu,\nu\nu')\ \eta\otimes\eta'
\end{aligned}
$$

$$= \sum_{\xi,\nu,\xi',\nu' \in X^*} (\bar{\xi}, \xi\xi')(\bar{\nu}, \nu\nu') \cdot$$

$$\left(\sum_{\eta \in X^*} (\eta, \xi \sqcup\!\sqcup \nu)\, \eta \right) \otimes \left(\sum_{\eta' \in X^*} (\eta', \xi' \sqcup\!\sqcup \nu')\, \eta' \right)$$

$$= \sum_{\xi,\nu,\xi',\nu' \in X^*} (\bar{\xi}, \xi\xi')(\bar{\nu}, \nu\nu')\, (\xi \sqcup\!\sqcup \nu) \otimes (\xi' \sqcup\!\sqcup \nu')$$

$$= \left(\sum_{\xi,\xi' \in X^*} (\bar{\xi}, \xi\xi')\, \xi \otimes \xi' \right) \sqcup\!\sqcup \left(\sum_{\nu,\nu' \in X^*} (\bar{\nu}, \nu\nu')\, \nu \otimes \nu' \right)$$

$$= \mathrm{cat}^*(\bar{\xi}) \sqcup\!\sqcup \mathrm{cat}^*(\bar{\nu}).$$

Again by linearity, this last equality holds for all $p, q \in \mathbb{R}\langle X \rangle$. ■

This basic duality theory can be generalized in several useful ways. First consider the tensor product space

$$\mathbb{R}\langle X \rangle^{\otimes k} := \mathbb{R}\langle X \rangle \otimes \mathbb{R}\langle X \rangle \otimes \cdots \otimes \mathbb{R}\langle X \rangle,$$

where $\mathbb{R}\langle X \rangle$ appears $k \geq 1$ times on the right-hand side. Define the k-shuffle product

$$\begin{aligned} \mathrm{sh}_k &: \mathbb{R}\langle X \rangle^{\otimes k} \to \mathbb{R}\langle X \rangle \\ &: p_1 \otimes p_2 \otimes \cdots \otimes p_k \mapsto p_1 \sqcup\!\sqcup p_2 \sqcup\!\sqcup \cdots \sqcup\!\sqcup p_k. \end{aligned}$$

A straight forward generalization of Theorem 2.6 gives

$$\mathrm{sh}_k^*(q_1 q_2 \cdots q_\ell) = \mathrm{sh}_k^*(q_1)\mathrm{sh}_k^*(q_2) \cdots \mathrm{sh}_k^*(q_\ell),$$

where $\mathrm{sh}_k^*(\cdot)$ satisfies

$$(p_1 \sqcup\!\sqcup p_2 \sqcup\!\sqcup \cdots \sqcup\!\sqcup p_k, q) = (p_1 \otimes p_2 \cdots \otimes p_k, \mathrm{sh}_k^*(q)), \tag{2.23}$$

and, in particular,

$$\mathrm{sh}_k^*(x_i) = x_i \otimes \mathbf{1} \otimes \cdots \otimes \mathbf{1} + \mathbf{1} \otimes x_i \otimes \cdots \otimes \mathbf{1} + \cdots + \mathbf{1} \otimes \mathbf{1} \otimes \cdots \otimes x_i.$$

This latter identity is written more compactly as

$$\mathrm{sh}_k^*(x_i) = \sum_{j=1}^{k} x_i^{\otimes j},$$

where

$$x_i^{\otimes j} := \overbrace{\underbrace{\mathbf{1} \otimes \cdots \otimes x_i}_{j\text{-th position}} \otimes \cdots \otimes \mathbf{1}}^{k \text{ terms}}.$$

One can also generalize the various scalar products discussed above to at least partially admit power series and not just polynomials. For example, given any $c, d \in \mathbb{R}\langle\langle X \rangle\rangle$ and $p, q \in \mathbb{R}\langle X \rangle$, define

$$\begin{aligned}(c,p) &= \sum_{\eta \in X^*} (c,\eta)(p,\eta) \\ (c \otimes d, p \otimes q) &= \sum_{\eta,\eta' \in X^*} (c \otimes d, \eta \otimes \eta')(p \otimes q, \eta \otimes \eta') \\ &= \sum_{\eta,\eta' \in X^*} (c,\eta)(d,\eta')(p,\eta)(q,\eta') \\ &= (c,p)(d,q).\end{aligned}$$

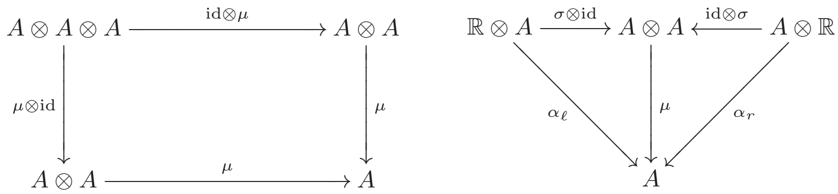
Since each summation above is finite, there are no convergence issues to consider, as would be the case if one tried to define a scalar product on $\mathbb{R}\langle\langle X \rangle\rangle \times \mathbb{R}\langle\langle X \rangle\rangle$. In this context, all of the results presented so far extend in the expected manner and, in fact, can be combined with the shuffle product generalization above. A particularly important example of this, which is used in Chapter 5, is the generalization of identity (2.23):

$$(c_1 ⧢ c_2 ⧢ \cdots ⧢ c_k, q) = (c_1 \otimes c_2 \otimes \cdots \otimes c_k, \mathrm{sh}_k^*(q)), \tag{2.24}$$

where $c_i \in \mathbb{R}\langle\langle X \rangle\rangle$ for $i = 1, 2, \ldots, k$ and $q \in \mathbb{R}\langle X \rangle$.

2.6 Hopf Algebras

The catenation-shuffle product duality described in the previous section is just part of a larger algebraic picture, one involving *Hopf algebras*. So this perspective is presented in this section. The starting point is the standard definition of a Hopf algebra, which at first glance seems like a rather complex mathematical object. It will then be shown that the catenation-shuffle product duality provides two Hopf algebras, which not unexpectedly are duals of each other in a certain sense. Next, a canonical construction of a Hopf algebra from a group is described. The Faà di Bruno Hopf algebra is presented as a specific example of



(a) Associative property (b) Unitary property

Fig. 2.4. Defining properties of an $\mathbb{R}$-algebra (A, μ, σ).

this construction. This particular algebra is important as it strongly motivates the synthesis of a second Hopf algebra which plays a central role in the analysis of feedback systems. The latter will be presented in Chapter 3.

First consider what it means for a set A to be a unital associative $\mathbb{R}$-algebra. Let A be an $\mathbb{R}$-vector space with two $\mathbb{R}$-linear maps

$$\mu : A \otimes A \to A,$$

and

$$\sigma : \mathbb{R} \to A,$$

which satisfy the associative and unitary properties, respectively, as described by the commutative diagrams in Figure 2.4.[5] Here id is the identity map on A, $\alpha_\ell : k \otimes a \mapsto ka$, and $\alpha_r : a \otimes k \mapsto ak$, where $k \in \mathbb{R}$ and $a \in A$. Therefore, $\mathbb{R} \otimes A$ and $A \otimes \mathbb{R}$ are each canonically identified with A. These diagrams are equivalent to, respectively, the identities

$$(ab)c = a(bc), \quad a, b, c \in A$$
$$\mathbf{1}a = a = a\mathbf{1}, \quad a \in A,$$

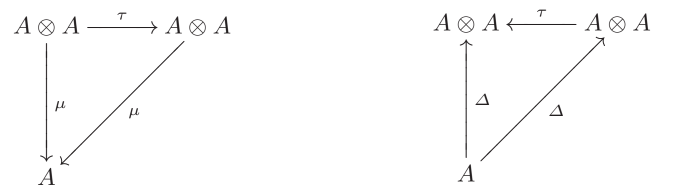
where $ab := \mu(a \otimes b)$ and $\mathbf{1} := \sigma(1)$ is the unit of A. Traditionally, μ is called the *multiplication map*, and σ is called the *unit map*. The triple (A, μ, σ) is a unital associative algebra. The algebra A is said to be *commutative* when $ab = ba$. The corresponding commutative diagram is shown in Figure 2.5(a), where $\tau : a \otimes b \mapsto b \otimes a$ for any $a, b \in A$.

Next suppose there exist two $\mathbb{R}$-linear maps

$$\Delta : A \to A \otimes A,$$

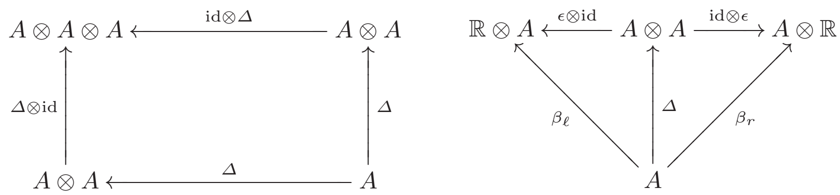
and

[5] It is more traditional to use m and u for the multiplication and unit maps, respectively. But these symbols clash with their use in system theory.



(a) Commutative property (b) Noncommutative property

Fig. 2.5. Defining the commutative and cocommutative properties.



(a) Coassociative property (b) Counitary property

Fig. 2.6. Defining properties of an $\mathbb{R}$-coalgebra (A, Δ, ϵ).

$$\epsilon : A \to \mathbb{R}$$

which satisfy the coassociative and counitary properties, respectively, as illustrated in Figure 2.6. These commutative diagrams are the same as the ones depicted in Figure 2.4 except that the directions of the arrows have been reversed. In this case, Δ is called the *comultiplication map*, and ϵ is the *counit map*. Here $\beta_\ell : a \to 1 \otimes a$ and $\beta_r : a \to a \otimes 1$ for $a \in A$. These diagrams are equivalent, respectively, to the identities

$$(\mathrm{id} \otimes \Delta) \circ \Delta = (\Delta \otimes \mathrm{id}) \circ \Delta$$
$$(\epsilon \otimes \mathrm{id}) \circ \Delta \sim (\mathrm{id} \otimes \epsilon) \circ \Delta,$$

where $\sim$ denotes the canonical equivalence between $\mathbb{R} \otimes A$ and $A \otimes \mathbb{R}$. The triple (A, Δ, ϵ) is called a counital coassociative *coalgebra*. A common notation known as Sweedler's notation is useful for representing coproducts in a calculation. It has several variations, for example,

$$\Delta(a) = \sum_{(a)} a_{(1)} \otimes a_{(2)} = \sum a_{(1)} \otimes a_{(2)} = a_{(1)} \otimes a_{(2)},$$

depending on the level of brevity desired. They all represent the sum of all possible *pieces* of $a \in A$ generated by applying the coproduct Δ. A coalgebra is said to be *cocommutative* when $\tau \circ \Delta = \Delta$ as shown in Figure 2.5(b).

Consider now the following definition.

Definition 2.13 *A* ***homomorphism between*** $\mathbb{R}$***-algebras*** (A_1, μ_1, σ_1) *and* (A_2, μ_2, σ_2) *is any* $\mathbb{R}$*-linear map* $\psi : A_1 \to A_2$ *such that*

$$\psi \circ \mu_1 = \mu_2 \circ (\psi \otimes \psi)$$
$$\psi \circ \sigma_1 = \sigma_2.$$

An analogous definition can be given for a homomorphism between two $\mathbb{R}$-coalgebras. Using either concept, one can produce the notion of a bialgebra as described next.

Definition 2.14 *The five-tuple* $(A, \mu, \sigma, \Delta, \epsilon)$ *is called an* $\mathbb{R}$***-bialgebra*** *when* Δ *and* ϵ *are both* $\mathbb{R}$*-algebra homomorphisms.*

Specifically this means that the mapping $\Delta : A \to A \otimes A$ must be an $\mathbb{R}$-algebra homomorphism between the $\mathbb{R}$-algebras (A, μ, σ) and $(A \otimes A, \mu_{A\otimes A}, \sigma_{A\otimes A})$, where [6]

$$\begin{aligned}
\mu_{A\otimes A} &: (A \otimes A) \otimes (A \otimes A) \to A \otimes A \\
&: (a_1 \otimes a_2) \otimes (a_3 \otimes a_4) \mapsto \mu(a_1 \otimes a_3) \otimes \mu(a_2 \otimes a_4) \\
\sigma_{A\otimes A} &: \mathbb{R} \otimes \mathbb{R} \to A \otimes A \\
&: k_1 \otimes k_2 \mapsto \sigma(k_1) \otimes \sigma(k_2).
\end{aligned}$$

In which case, it follows directly that

1. $\Delta \circ \mu = \mu_{A\otimes A} \circ (\Delta \otimes \Delta) = (\mu \otimes \mu) \circ (\mathrm{id} \otimes \tau \otimes \mathrm{id}) \circ (\Delta \otimes \Delta)$
2. $\Delta \circ \sigma = \sigma_{A\otimes A} = \sigma \otimes \sigma$

(see Problem 2.6.1). Similarly, $\epsilon : A \to \mathbb{R}$ must be an $\mathbb{R}$-algebra homomorphism between the $\mathbb{R}$-algebras (A, μ, σ) and $(\mathbb{R}, \mu_{\mathbb{R}}, \sigma_{\mathbb{R}})$. Therefore,

3. $\epsilon \circ \mu = \mu_{\mathbb{R}} \circ (\epsilon \otimes \epsilon) = \epsilon \cdot \epsilon$
4. $\epsilon \circ \sigma = \sigma_{\mathbb{R}} = 1.$

Note that properties 1 and 2 can be expressed in terms of the commutative diagrams shown in Figure 2.7, and, likewise, properties 3 and 4 are shown in Figure 2.8. They are often written in a more abbreviated notation as:

1. $\Delta(aa') = \Delta(a)\Delta(a')$, $a, a' \in A$
2. $\Delta(\mathbf{1}) = \mathbf{1} \otimes \mathbf{1}$
3. $\epsilon(aa') = \epsilon(a)\epsilon(a')$, $a, a' \in A$
4. $\epsilon \circ \sigma(k) = k$, $k \in \mathbb{R}$.

[6] In the definition of $\sigma_{A\otimes A}$, $\mathbb{R} \otimes \mathbb{R}$ is being identified with $\mathbb{R}$ via the mapping β.

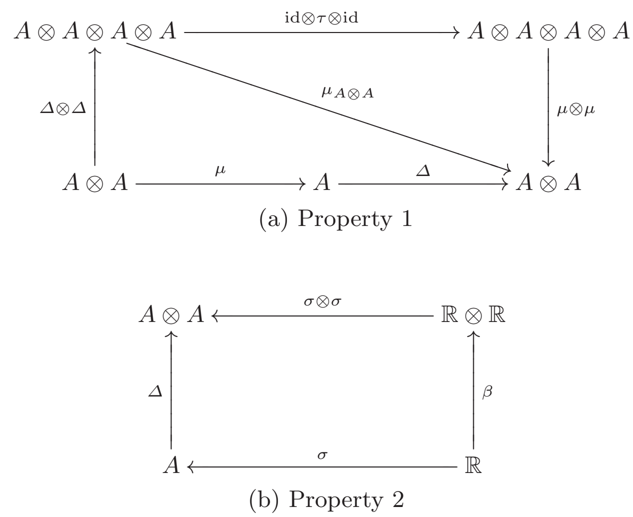


Fig. 2.7. Commutative diagrams describing Δ as an $\mathbb{R}$-algebra homomorphism.

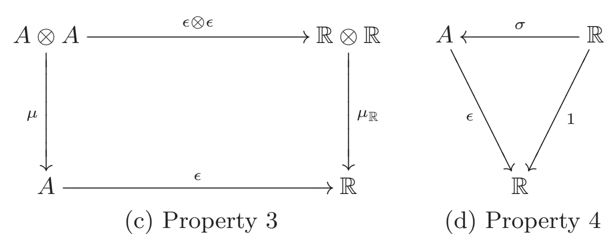


Fig. 2.8. Commutative diagrams describing ϵ as an $\mathbb{R}$-algebra homomorphism.

If instead one introduces the notion of an $\mathbb{R}$-coalgebra homomorphism as suggested above, then an equivalent characterization of a bialgebra is one where μ and σ are both $\mathbb{R}$-coalgebra homomorphisms, yielding properties 1 and 3, and properties 2 and 4, respectively. That exercise is left to the reader.

To complete the development of the Hopf algebra definition, consider the set of all $\mathbb{R}$-linear maps taking vector space A back to itself, denoted by $\mathrm{End}(A)$.[7] Given two arbitrary $f, g \in \mathrm{End}(A)$, the *Hopf convolution product*,

$$f \star g := \mu \circ (f \otimes g) \circ \Delta,$$

defines another element of $\mathrm{End}(A)$. The following theorem is central to the theory

[7] Such maps are called *endomorphisms* on vector space A.

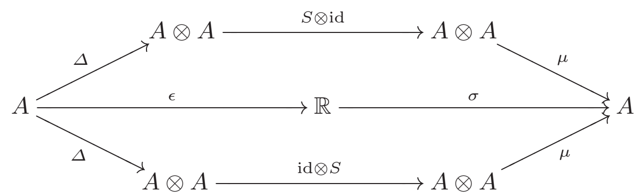


Fig. 2.9. Commutative diagram describing the antipode, S.

Theorem 2.7 *The triple* $(\mathrm{End}(A), \star, \vartheta)$ *forms an associative* $\mathbb{R}$*-algebra with unit* $\vartheta = \sigma \circ \epsilon$.

Proof: The associativity of the convolution product follows directly from the associativity of μ and the coassociativity of Δ:

$$\begin{aligned} f \star (g \star h) &= \mu \circ (f \otimes (g \star h)) \circ \Delta \\ &= \mu \circ (f \otimes (\mu \circ (g \otimes h) \circ \Delta)) \circ \Delta \\ &= \mu \circ ((\mathrm{id} \otimes \mu)(f \otimes (g \otimes h))(\mathrm{id} \otimes \Delta)) \circ \Delta \\ &= \mu \circ ((\mu \otimes \mathrm{id})((f \otimes g) \otimes h))(\Delta \otimes \mathrm{id})) \circ \Delta \\ &= \mu \circ ((\mu \circ (f \otimes g) \circ \Delta) \otimes h) \circ \Delta \\ &= \mu \circ ((f \star g) \otimes h) \circ \Delta \\ &= (f \star g) \star h. \end{aligned}$$

To show that ϑ is the convolution unit, it is necessary to use the counit identity $(\mathrm{id} \otimes \epsilon) \circ \Delta = \mathrm{id} \otimes 1$ (see Figure 2.6(b)). Observe that

$$\begin{aligned} f \star \vartheta &= \mu \circ (f \otimes (\sigma \circ \epsilon)) \circ \Delta \\ &= \mu \circ ((\mathrm{id} \otimes \sigma)(f \otimes 1)(\mathrm{id} \otimes \epsilon)) \circ \Delta \\ &= \mu \circ ((\mathrm{id} \otimes \sigma)(f \otimes 1)(\mathrm{id} \otimes 1)). \end{aligned}$$

Thus, for any $a \in A$

$$\begin{aligned} (f \star \vartheta)(a) &= \mu((\mathrm{id} \otimes \sigma)(f \otimes 1)(a \otimes 1)) \\ &= \mu(f(a) \otimes \mathbf{1}) \\ &= f(a)\mathbf{1} \\ &= f(a). \end{aligned}$$

Likewise, $\vartheta \star f = f$. ■

Finally, an element $S \in \mathrm{End}(A)$ satisfying

$$S \star \mathrm{id} = \mathrm{id} \star S = \vartheta. \tag{2.25}$$

is called an *antipode* of the bialgebra. The corresponding commutative diagram is shown in Figure 2.9. Equation (2.25) implies that S is the convolution inverse of id, so formally

$$S = \mathrm{id}^{\star -1} = (\vartheta - (\vartheta - \mathrm{id}))^{\star -1} = \vartheta + \sum_{k=1}^{\infty} (\vartheta - \mathrm{id})^{\star k}. \tag{2.26}$$

It can be shown that when an antipode exists, it must be unique. It also follows that $S(\mathbf{1}) = \mathbf{1}$ and $S(aa') = S(a')S(a)$ for any $a, a' \in A$ (see Problem 2.6.3). This final bit of structure culminates in the definition below.

Definition 2.15 *The six-tuple* $(A, \mu, \sigma, \Delta, \epsilon, S)$ *is an* ***ℝ-Hopf algebra*** *if it is an* $\mathbb{R}$*-bialgebra with an antipode.*

Two $\mathbb{R}$-Hopf algebras can be introduced on $\mathbb{R}\langle X\rangle$, one associated with the catenation product and the other with the shuffle product. They are duals of each other in the sense described by Theorem 2.6. First, the relevant $\mathbb{R}$-bialgebras are described.

Theorem 2.8 $(\mathbb{R}\langle X\rangle, \mathrm{cat}, \sigma, \mathrm{sh}^*, \epsilon)$ *is a noncommutative cocommutative* $\mathbb{R}$*-bialgebra, where*

$$\begin{aligned} \sigma &: \mathbb{R} \to \mathbb{R}\langle X\rangle : k \mapsto k\mathbf{1} \\ \epsilon &: \mathbb{R}\langle X\rangle \to \mathbb{R} : p \mapsto (p, \emptyset). \end{aligned}$$

Likewise, $(\mathbb{R}\langle X\rangle, \mathrm{sh}, \sigma, \mathrm{cat}^*, \epsilon)$ *is a commutative noncocommutative* $\mathbb{R}$*-bialgebra.*

Proof: The defining properties are easy to check. As an example, consider the bialgebra $(\mathbb{R}\langle X\rangle, \mathrm{cat}, \sigma, \mathrm{sh}^*, \epsilon)$. Here it is necessary to verify that sh^* and ϵ are $\mathbb{R}$-algebra homomorphisms. Specifically, this means that the following identities must be satisfied:

1. $\mathrm{sh}^* \circ \mathrm{cat} = (\mathrm{cat} \otimes \mathrm{cat}) \circ (\mathrm{id} \otimes \tau \otimes \mathrm{id}) \circ (\mathrm{sh}^* \otimes \mathrm{sh}^*)$
2. $\mathrm{sh}^* \circ \sigma = \sigma \otimes \sigma$
3. $\epsilon \circ \mathrm{cat} = \epsilon \cdot \epsilon$
4. $\epsilon \circ \sigma = 1,$

where

$$\begin{aligned}
\text{cat} &: p \otimes q \to pq \\
\text{sh}^* &: r \to \sum_{\xi,\nu \in X^*} (r, \xi \sqcup\!\sqcup \nu)\, \xi \otimes \nu.
\end{aligned}$$

The details of each calculation are presented below.

1. On the left-hand side, using Theorem 2.6, observe that

$$\begin{aligned}
\text{sh}^*(\text{cat}(p \otimes q)) &= \text{sh}^*(pq) = sh^*(p)\, \text{sh}^*(q) \\
&= \sum_{\xi,\nu \in X^*} (p, \xi \sqcup\!\sqcup \nu)\, \xi \otimes \nu \sum_{\xi',\nu' \in X^*} (q, \xi' \sqcup\!\sqcup \nu')\, \xi' \otimes \nu' \\
&= \sum_{\xi,\nu,\xi',\nu' \in X^*} (p, \xi \sqcup\!\sqcup \nu)(q, \xi' \sqcup\!\sqcup \nu')\, \xi\xi' \otimes \nu\nu'.
\end{aligned}$$

   While on the right-side, one has

$$\begin{aligned}
&(\text{cat} \otimes \text{cat}) \circ (\text{id} \otimes \tau \otimes \text{id}) \circ (\text{sh}^* \otimes \text{sh}^*)(p \otimes q) \\
&= (\text{cat} \otimes \text{cat}) \circ (\text{id} \otimes \tau \otimes \text{id})(\text{sh}^*(p) \otimes \text{sh}^*(q)) \\
&= (\text{cat} \otimes \text{cat}) \circ (\text{id} \otimes \tau \otimes \text{id}) \left(\sum_{\xi,\nu,\xi',\nu' \in X^*} (p, \xi \sqcup\!\sqcup \nu)(q, \xi' \sqcup\!\sqcup \nu') \right. \\
&\qquad \left. \xi \otimes \nu \otimes \xi' \otimes \nu' \right) \\
&= (\text{cat} \otimes \text{cat}) \left(\sum_{\xi,\nu,\xi',\nu' \in X^*} (p, \xi \sqcup\!\sqcup \nu)(q, \xi' \sqcup\!\sqcup \nu') \cdot \right. \\
&\qquad \left. \xi \otimes \xi' \otimes \nu \otimes \nu' \right) \\
&= \sum_{\xi,\nu,\xi',\nu' \in X^*} (p, \xi \sqcup\!\sqcup \nu)(q, \xi' \sqcup\!\sqcup \nu')\, \xi\xi' \otimes \nu\nu'.
\end{aligned}$$

   Hence, the first identity is satisfied.

2. Here the left-hand side evaluates to

$$\begin{aligned}(\mathrm{sh}^* \circ \sigma)(k) &= \mathrm{sh}^*(k\mathbf{1}) \\ &= \sum_{\xi,\nu \in X^*} (k\mathbf{1}, \xi \sqcup\!\sqcup \nu)\, \xi \otimes \nu = k(\mathbf{1} \otimes \mathbf{1}) = k\mathbf{1} \otimes \mathbf{1}.\end{aligned}$$

As expected, the right-hand side gives

$$(\sigma \otimes \sigma)(k \otimes \mathbf{1}) = \sigma(k) \otimes \sigma(1) = k\mathbf{1} \otimes \mathbf{1}.$$

3. In this case, the left-hand side is

$$(\epsilon \circ \mathrm{cat})(p \otimes q) = \epsilon(pq) = (pq, \emptyset) = (p, \emptyset)(q, \emptyset).$$

While the right-hand side is

$$(\epsilon \cdot \epsilon)(p, q) = \epsilon(p)\epsilon(q) = (p, \emptyset)(q, \emptyset).$$

4. This identity is especially simple. Observe

$$(\epsilon \circ \sigma)(k) = \epsilon(k\mathbf{1}) = k.$$

It is obvious that the algebra $(\mathbb{R}\langle X\rangle, \mathrm{cat}, \sigma)$ is not commutative. To see that $(\mathbb{R}\langle X\rangle, \mathrm{sh}^*, \epsilon)$ is cocommutative observe from (2.16) that for any $r \in \mathbb{R}\langle X\rangle$ that

$$\begin{aligned}\tau \circ \mathrm{sh}^*(r) &= \tau\left(\sum_{\xi,\nu \in X^*} (r, \xi \sqcup\!\sqcup \nu)\, \xi \otimes \nu\right) \\ &= \sum_{\xi,\nu \in X^*} (r, \xi \sqcup\!\sqcup \nu)\, \nu \otimes \xi \\ &= \sum_{\xi,\nu \in X^*} (r, \nu \sqcup\!\sqcup \xi)\, \nu \otimes \xi \\ &= \mathrm{sh}^*(r).\end{aligned}$$

The analogous arguments regarding the bialgebra $(\mathbb{R}\langle X\rangle, \mathrm{sh}, \epsilon, \mathrm{cat}^*, \sigma)$ are left to the reader to verify. ■

Next, the corresponding Hopf algebras are described.

Theorem 2.9 $(\mathbb{R}\langle X\rangle, \mathrm{cat}, \sigma, \mathrm{sh}^*, \epsilon, S)$ *is a noncommutative cocommutative $\mathbb{R}$-Hopf algebra, where the convolution product is*

$$\begin{aligned} f \star g &: \mathbb{R}\langle X\rangle \to \mathbb{R}\langle X\rangle \\ &: p \mapsto \sum_{\eta,\xi\in X^*} (p, \eta \sqcup\!\sqcup \xi)\, f(\eta)g(\xi), \end{aligned}$$

and S is the unique $\mathbb{R}$-linear map satisfying

$$S(x_{i_1}x_{i_2}\cdots x_{i_k}) = (-1)^k x_{i_k}x_{i_{k-1}}\cdots x_{i_1}$$

for every $x_{i_1}x_{i_2}\cdots x_{i_k} \in X^*$. Likewise, $(\mathbb{R}\langle X\rangle, \mathrm{sh}, \sigma, \mathrm{cat}^*, \epsilon, S)$ is a commutative noncocommutative $\mathbb{R}$-Hopf algebra, where the convolution product is

$$\begin{aligned} f \star' g &: \mathbb{R}\langle X\rangle \to \mathbb{R}\langle X\rangle \\ &: p \mapsto \sum_{\eta,\xi\in X^*} (p, \eta\xi)\, f(\eta) \sqcup\!\sqcup g(\xi), \end{aligned}$$

and S is as above.

Proof: Regarding the first claim, it is first necessary to verify that

$$f \star g := (\mathrm{cat} \circ (f \otimes g) \circ \mathrm{sh}^*)(p) = \sum_{\eta,\xi\in X^*} (p, \eta \sqcup\!\sqcup \xi)\, f(\eta)g(\xi).$$

Recalling that

$$sh^*(p) = \sum_{\eta,\xi\in X^*} (p, \eta \sqcup\!\sqcup \xi)\, \eta \otimes \xi,$$

it follows directly that

$$((f \otimes g) \circ \mathrm{sh}^*)(p) = \sum_{\eta,\xi\in X^*} (p, \eta \sqcup\!\sqcup \xi)\, f(\eta) \otimes g(\xi).$$

Whereupon the desired equality follows immediately. The only remaining task is to verify that $I \star S = S \star I = \sigma \circ \epsilon$. Since in this case, $(\sigma \circ \epsilon)(p) = (p, \emptyset)1$, $\forall p \in \mathbb{R}\langle X\rangle$, and, for example, $I \star S$ is $\mathbb{R}$-linear, it is sufficient to show that

$$(I \star S)(\mathbf{1}) = 1, \;\; (I \star S)(\nu) = 0, \;\; \forall \nu \in X^+.$$

The first identity is trivial. The second identity follows from induction. Observe that any $x_i \in X$

$$
\begin{aligned}
(I \star S)(x_i) &= \sum_{\eta,\xi \in X^*} (x_i, \eta \sqcup\!\sqcup \xi)\, \eta(-1)^{|\xi|}\tilde{\xi} \\
&= x_i(-1)^0 1 + 1(-1)^1 x_i \\
&= 0,
\end{aligned}
$$

where $\tilde{\xi}$ denotes ξ with the letters written in reverse order. Now assume the identity holds for all words up to some fixed length $k \geq 0$. Then for any $x_i \in X$ and $\nu \in X^k$

$$
\begin{aligned}
(I \star S)(x_i\nu) &= \sum_{\eta,\xi \in X^*} (x_i\nu, \eta \sqcup\!\sqcup \xi)\, \eta(-1)^{|\xi|}\tilde{\xi} \\
&= \sum_{\eta,\xi \in X^*} (\nu, x_i^{-1}(\eta \sqcup\!\sqcup \xi))\, \eta(-1)^{|\xi|}\tilde{\xi} \\
&= \sum_{\eta,\xi \in X^*} (\nu, x_i^{-1}(\eta) \sqcup\!\sqcup \xi)\, \eta(-1)^{|\xi|}\tilde{\xi} + \\
&\qquad \sum_{\eta,\xi \in X^*} (\nu, \eta \sqcup\!\sqcup x_i^{-1}(\xi))\, \eta(-1)^{|\xi|}\tilde{\xi} \\
&= x_i \left(\sum_{\eta,\xi \in X^*} (\nu, \eta \sqcup\!\sqcup \xi)\, \eta(-1)^{|\xi|}\tilde{\xi} \right) - \\
&\qquad \left(\sum_{\eta,\xi \in X^*} (\nu, \eta \sqcup\!\sqcup \xi)\, \eta(-1)^{|\xi|}\tilde{\xi} \right) x_i \\
&= x_i\, (I \star S)(\nu) - (I \star S)(\nu)\, x_i \\
&= 0.
\end{aligned}
$$

The claim regarding the second Hopf algebra is left to the reader. The identity from Problem 2.4.3(d) is useful in this case. ■

The following lemma provides an interesting interpretation of the antipode in the context of iterated integrals defined for m functions which are absolutely integrable over $[0,T]$, denoted here by $L_1^m[0,T]$ This identity will reappear later in Section 5.4 when Chen series are introduced.

Lemma 2.4 *Let $X = \{x_1, x_2, \ldots, x_m\}$. For any given $u \in L_1^m[0,T]$ and fixed $t \in [0,T]$ define the input function $u_{S,i}(\tau) = -u_i(t-\tau)$ on $[0,t]$ for $i = 1, 2, \ldots, m$. Then for any $\eta \in X^*$ it follows that*

$$E_{S(\eta)}[u](t,0) = E_\eta[u_S](t,0).$$

Proof: The claim is trivial when $\eta = \emptyset$. In the case where $\eta = x_{i_k}x_{i_{k-1}}\cdots x_{i_1}$, observe:

$$\begin{aligned}
E_{S(x_{i_k}x_{i_{k-1}}\cdots x_{i_1})}[u](t,0) &= (-1)^k E_{x_{i_1}x_{i_2}\cdots x_{i_k}}[u](t,0) \\
&= (-1)^k \int_0^t \int_0^{\tau_1} \cdots \int_0^{\tau_{k-1}} u_{i_1}(\tau_1)u_{i_2}(\tau_2)\cdots \\
&\qquad u_{i_k}(\tau_k)\, d\tau_k d\tau_{k-1}\cdots d\tau_1 \\
&= (-1)^k \int_0^t \cdots \int_0^t u_{i_1}(\tau_1)\cdots u_{i_k}(\tau_k) \\
&\qquad \mathbb{U}(\tau_1-\tau_2)\cdots\mathbb{U}(\tau_{k-1}-\tau_k)\, d\tau_k\cdots d\tau_1,
\end{aligned}$$

where $\mathbb{U}$ denotes the unit step function

$$\mathbb{U}(t) = \begin{cases} 1 : t \geq 0 \\ 0 : t < 0. \end{cases}$$

Interchanging the order of integration gives

$$\begin{aligned}
E_{S(\eta)}[u](t,0) = (-1)^k \int_0^t \int_{\tau_k}^t \cdots \int_{\tau_2}^t u_{i_1}(\tau_1)u_{i_2}(\tau_2)\cdots \\
u_{i_k}(\tau_k)\, d\tau_1 d\tau_2\cdots d\tau_k.
\end{aligned}$$

Finally, substituting $t-\tau_1$ for τ_1 followed by $t-\tau_2$ for τ_2, etc., yields the desired result, namely,

$$\begin{aligned}
E_{S(\eta)}[u](t,0) &= \int_0^t \int_0^{\tau_k} \cdots \int_0^{\tau_2} (-u_{i_1}(t-\tau_1))(-u_{i_2}(t-\tau_2))\cdots \\
&\qquad (-u_{i_k}(t-\tau_k))\, d\tau_1 d\tau_2\cdots d\tau_k \\
&= E_\eta[u_S](t,0).
\end{aligned}$$

■

In some cases it is possible to introduce additional structure on a bialgebra A to guarantee that it has a well-defined antipode, and thus is a Hopf algebra. The following definitions are essential in this regard.

Definition 2.16 *An $\mathbb{R}$-bialgebra $(A,\mu,\sigma,\Delta,\epsilon)$ is **filtered** if there exists a nested sequence of $\mathbb{R}$-vector subspaces of A, say $A_{(0)} \subsetneq A_{(1)} \subsetneq \cdots$, such that $A = \cup_{n\geq 0}A_{(n)}$ and*

$$\Delta A_{(n)} \subseteq \sum_{i=0}^{n} A_{(i)} \otimes A_{(n-i)}.$$

The collection $\{A_{(n)}\}_{n\geq 0}$ *is called a* ***filtration*** *of* A.

Definition 2.17 *An* $\mathbb{R}$*-bialgebra that is filtered such that* $A_{(0)} = \sigma(\mathbb{R})$ *is said to be* ***connected****.*

Definition 2.18 *An* $\mathbb{R}$*-bialgebra is* ***graded*** *if there exists a set of* $\mathbb{R}$*-vector subspaces of* A*, say* $\{A_n\}_{n\geq 0}$*, such that* $A = \oplus_{n\geq 0} A_n$ *with*

$$A_i A_j \subseteq A_{i+j}, \quad \Delta A_n \subseteq \bigoplus_{i=0}^{n} A_i \otimes A_{n-i},$$

and $\epsilon(A_n) = 0$*,* $n > 0$*.*

Definition 2.19 *Let* A *be an* $\mathbb{R}$*-bialgebra. An element* $g \in A$ *is* ***group-like*** *if* $\epsilon(g) = 1$ *and* $\Delta g = g \otimes g$*. If* A *has only one group-like element, then any other element* $a \in A$ *is* ***primitive*** *if* $\Delta a = a \otimes g + g \otimes a$*.*

A number of useful results follow from these definitions. For example, if A has a grading $\{A_n\}_{n\geq 0}$, then a natural filtration of A is $\{A_{(n)}\}_{n\geq 0}$, where

$$A_{(n)} = \bigoplus_{i=0}^{n} A_i.$$

Furthermore, if $A_0 = \sigma(\mathbb{R})$ then A has only one group-like element. Perhaps the most important aspect concerning a connected bialgebra is a key property of its coalgebra. If

$$A^+ := \ker \epsilon, \;\; A^+_{(n)} := A^+ \cap A_{(n)} \tag{2.27}$$

then for any $a \in A^+_{(n)}$ it follows that

$$\Delta a = a \otimes \mathbf{1} + \mathbf{1} \otimes a + \Delta' a, \tag{2.28}$$

where the *reduced* coproduct $\Delta' a \in A^+_{(n-1)} \otimes A^+_{(n-1)}$ (see Problem 2.6.4). This leads to the following central result.

Theorem 2.10 *Let* $(A, \mu, \sigma, \Delta, \epsilon)$ *be a connected* $\mathbb{R}$*-bialgebra. Then* $(A, \mu, \sigma, \Delta, \epsilon, S)$ *is an* $\mathbb{R}$*-Hopf algebra, where the antipode is given on* A^+ *by*

$$Sa = -a - \sum (Sa'_{(1)})a'_{(2)}, \tag{2.29}$$

or alternatively,

$$Sa = -a - \sum a'_{(1)} Sa'_{(2)} \tag{2.30}$$

with $\Delta' a = \sum a'_{(1)} \otimes a'_{(2)}$ *being the reduced coproduct in Sweedler's notation.*

Proof: It is first shown by induction that if $a \in A^+_{(n)}$ then the series representation (2.25) for Sa is finite, specifically,

$$Sa = \left[\vartheta + \sum_{k=1}^{n} (\vartheta - \mathrm{id})^{\star k} \right] (a).$$

Therefore, since antipodes are unique, this must be the antipode for all of A. The claim clearly holds when $a \in A_{(0)}$ since $(\vartheta - \mathrm{id})\mathbf{1} = 0$. If $a \in A^+_1$ observe that for any $k \geq 2$

$$\begin{aligned}
(\vartheta - \mathrm{id})^{\star k}(a) &= [(\vartheta - \mathrm{id}) \star (\vartheta - \mathrm{id})^{\star k-1}](a) \\
&= \mu[(\vartheta - \mathrm{id}) \otimes (\vartheta - \mathrm{id})^{\star k-1}]\Delta a \\
&= \mu[(\vartheta - \mathrm{id}) \otimes (\vartheta - \mathrm{id})^{\star k-1}](a \otimes \mathbf{1} + \mathbf{1} \otimes a) \\
&= 0
\end{aligned}$$

since again $(\vartheta - \mathrm{id})\mathbf{1} = 0$. Now assume the claim holds up to some fixed $n \geq 1$. If $a \in A^+_{(n+1)}$ then

$$\begin{aligned}
(\vartheta - \mathrm{id})^{\star n+1}(a) &= \mu[(\vartheta - \mathrm{id}) \otimes (\vartheta - \mathrm{id})^{\star n}](a \otimes \mathbf{1} + \mathbf{1} \otimes a + \Delta' a) \\
&= \mu[(\vartheta - \mathrm{id}) \otimes (\vartheta - \mathrm{id})^{\star n}]\Delta' a \\
&= 0,
\end{aligned}$$

where the induction hypothesis was used to arrive at the final equality. Therefore, the result holds for all $n \geq 0$.

Having established that the bialgebra is a Hopf algebra, it is shown next that S has the recursive forms given in (2.29) and (2.30). The focus will be on the first formula, the other one follows similarly. The case where $a \in A^+_{(1)}$ is trivial. Assume the identity in question hold up to some fixed $n \geq 1$. Then for $a \in A^+_{n+1}$ observe

$$Sa = (\vartheta - \mathrm{id})a + \sum_{k=1}^{n} (\vartheta - \mathrm{id})^{\star k} \star (\vartheta - \mathrm{id})a$$

$$= -a + \mu \left(\sum_{k=1}^{n} (\vartheta - \mathrm{id})^{\star k} \otimes (\vartheta - \mathrm{id}) \right) \Delta a$$
$$= -a + \mu \left(\sum_{k=1}^{n} (\vartheta - \mathrm{id})^{\star k} \otimes (\vartheta - \mathrm{id}) \right) \sum a'_{(1)} \otimes a'_{(2)}$$
$$= -a + \sum \left[\sum_{k=1}^{n} (\vartheta - \mathrm{id})^{\star k} a'_{(1)} \right] (\vartheta - \mathrm{id}) a'_{(2)}$$
$$= -a - \sum \left[\sum_{k=1}^{n} (\vartheta - \mathrm{id})^{\star k} a'_{(1)} \right] a'_{(2)}$$
$$= -a - \sum S(a'_{(1)}) a'_{(2)},$$

where the induction hypothesis is employed to get the last equality. This proves the final part of the theorem. ■

One way that Hopf algebras naturally arise is in the context of groups. Suppose V is a finite dimensional vector space over $\mathbb{R}$. Let $GL(V)$ denote the *general linear group* on V, namely, the group of all invertible $\mathbb{R}$-linear maps taking V back to itself. A group G with unit 1_G is said to have a *representation* if there exists a group homomorphism $\pi : G \to GL(V)$, that is,

$$\pi(gg') = \pi(g_1)\pi(g_2), \quad \forall g_1, g_2 \in G. \tag{2.31}$$

The representation is *faithful* if π is injective. Given some fixed basis for V, $A = \pi(g)$ is an invertible matrix with real coefficients. In which case, there exists a set of *coordinate functions* of the form $a_{ij} : G \to \mathbb{R}$. This collection of functions, $R(G)$, forms a commutative algebra under pointwise defined operations, for example,

$$(a_{ij}a_{kl})(g) := a_{ij}(g)a_{kl}(g), \quad \forall g \in G.$$

From (2.31) and the definition of matrix multiplication there is a well defined coproduct

$$\Delta : R(G) \to R(G) \otimes R(G)$$
$$: a_{ij} \mapsto \sum_k a_{ik} \otimes a_{kj}.$$

It can be shown that $R(G)$ constitutes a commutative $\mathbb{R}$-Hopf algebra, H, where the unit, counit, and antipode maps are given, respectively, by $\sigma(1) = \mathbf{1}$ with $\mathbf{1}(g) = 1$, $\forall g \in G$, $\epsilon(a_{ij}) = a_{ij}(1_G)$, and

$$Sa_{ij}(g) = a_{ij}(g^{-1}), \quad \forall g \in G. \tag{2.32}$$

For any $g \in G$ one can define an $\mathbb{R}$-linear map $\Phi_g : H \to \mathbb{R}$ via

$$\Phi_g : a_{ij} \mapsto a_{ij}(g),$$

and $\Phi_g(\mathbf{1}) = 1$ so that

$$\Phi_g(a_{ij}a_{kl}) = a_{ij}(g)a_{kl}(g) = \Phi_g(a_{ij})\Phi_g(a_{kl}).$$

These are usually called the *characters* of the Hopf algebra, and they form a group under the Hopf convolution product. Specifically, the group product satisfies

$$\begin{aligned}(\Phi_{g_1} \star \Phi_{g_2})(a_{ij}) &= \mu \circ (\Phi_{g_1} \otimes \Phi_{g_2}) \circ \Delta a_{ij} \\ &= \sum_k \Phi_{g_1}(a_{ik})\, \Phi_{g_2}(a_{kj}) \\ &= \sum_k a_{ik}(g_1)a_{kj}(g_2) \\ &= a_{ij}(g_1 g_2) \end{aligned} \tag{2.33}$$

(apply (2.31) in the final step), and the identity element of the group is Φ_{1_G}. In addition, from (2.32)

$$\begin{aligned}\Phi_{g^{-1}}(a_{ij}) &= a_{ij}(g^{-1}) = (Sa_{ij})(g) \\ &= (\Phi_g \circ S)(a_{ij}) =: \Phi_g^{\star -1}(a_{ij}). \end{aligned} \tag{2.34}$$

In which case, the bijective map $\psi : g \mapsto \Phi_g$ is a group isomorphism between G and the character group.

A specific example of this construction is the *Faà di Bruno Hopf algebra*. The name refers to the well known Faà di Bruno formula from calculus which describes the composition of two functions in terms of their Taylor series. Let f_c and f_d be two functions with convergent Taylor series about the point $z = 0$ having the property that $f_c(0) = 0$ and $f_d(0) = 0$. In which case,

$$f_c(z) = \sum_{n=1}^{n} c(n)\frac{z^n}{n!}, \quad f_d(z) = \sum_{n=1}^{n} d(n)\frac{z^n}{n!}. \tag{2.35}$$

It is further assumed that these functions are normalized in the sense that $c(1) = d(1) = 1$. It is easy to show that the composition $f_{c\circ d} :=$

$f_c \circ f_d$ has the same nature as f_c and f_d, namely, it has a convergent Taylor series representation

$$f_{c\circ d}(z) = \sum_{n=1}^{\infty} (c \circ d)(n) \frac{z^n}{n!}$$

for some set of coefficients $(c \circ d)(n) \in \mathbb{R}$, $n \geq 1$ with $(c \circ d)(1) = 1$, as does the composition inverse of any such function, for example, f_c^{-1}, where $f_c^{-1} \circ f_c = f_c \circ f_c^{-1} = I$ with $I(z) := z$. In which case, this class of functions forms a group G_{FdB} under composition. The well known Faà di Bruno formula provides the Taylor series coefficients of $f_{c\circ d}$, specifically,

$$(c \circ d)(n) = \sum_{k=1}^{n} \frac{c(k)}{k!} \sum_{j} \frac{n!k!}{j_1!j_2!\cdots j_n!} \frac{d(1)^{j_1} d(2)^{j_2} \cdots d(n)^{j_n}}{(1!)^{j_1}(2!)^{j_2}\cdots(n!)^{j_n}}, \quad (2.36)$$

where the second sum is over all $j_1, j_2, \ldots, j_n \geq 0$ such that $j_1 + j_2 + \cdots + j_n = k$ and $j_1 + 2j_2 + \cdots + nj_n = n$.

To construct the underlying Hopf algebra, first let $\mathbb{R}_p\,[[X]]$ be the set of all proper series over the alphabet $X = \{x\}$. Now identify a given Taylor series $f_c(z) = \sum_{n\geq 1} c(n) z^n/n!$ with its corresponding formal power series $c \in \mathbb{R}_p\,[[X]]$. Let $c \circ d$ denote the formal power series corresponding to the function $f_c \circ f_d$, where f_c and f_d are given in (2.35). For any $n \geq 1$ define the coordinate function

$$\begin{aligned} a_n &: \mathbb{R}_p\,[[X]] \to \mathbb{R} \\ &: c \mapsto (c, x^n) = c(n). \end{aligned}$$

As described above, the set of these real-valued mappings defines an $\mathbb{R}$-vector space, H, and a commutative algebra where the product is given by

$$\mu : a_n \otimes a_m \mapsto a_n a_m$$

with unit $a_1 \sim \mathbf{1}$.[8] Given that the underlying group representation of G_{FdB} associated with this Hopf algebra is not finite dimensional (see Problem 2.6.5), Theorem 2.10 will be utilized to ensure that the construction is successful. The *degree* of a_n is defined to be $\deg(a_n) = n - 1$, $n \geq 1$, and $\deg(a_n a_m) = \deg(a_n) + \deg(a_m)$. Therefore, $H =$

[8] Here $\mathbf{1}$ is the unit of a new algebra and should not be confused with the monomial $1\emptyset$ by the same name, which was the unit for the catenation and shuffle algebras.

$\oplus_{n\geq 0} H_n$, where H_n denotes all elements of degree n, constitutes a grading of H. Since $H_0 = \mathbb{R}\mathbf{1}$, this grading is connected. The key idea is that (2.36) can be used to define a coproduct on H:

$$\begin{aligned}\Delta a_n(c,d) &= a_n(c\circ d) = (c\circ d, x^n)\\ &= \sum_{k=1}^{n} \frac{a_k(c)}{k!} \sum_j \frac{n!k!}{j_1!j_2!\cdots j_n!} \frac{a_1^{j_1}(d)a_2^{j_2}(d)\cdots a_n^{j_n}(d)}{(1!)^{j_1}(2!)^{j_2}\cdots(n!)^{j_n}}\\ &= \sum_{k=1}^{n} \frac{1}{k!} \sum_j \frac{n!k!}{j_1!j_2!\cdots j_n!} \frac{a_k\otimes a_1^{j_1}a_2^{j_2}\cdots a_n^{j_n}}{(1!)^{j_1}(2!)^{j_2}\cdots(n!)^{j_n}}(c,d).\end{aligned}$$

For example, the first few coproducts ordered by degree are:

$$\begin{aligned}\Delta a_1 &= a_1\otimes a_1\\ \Delta a_2 &= a_1\otimes a_2 + a_2\otimes a_1^2\\ \Delta a_3 &= a_1\otimes a_3 + a_2\otimes 3a_1a_2 + a_3\otimes a_1^3\\ \Delta a_4 &= a_1\otimes a_4 + a_2\otimes(4a_1a_3+3a_2^2) + a_3\otimes 6a_1^2a_2 + a_4\otimes a_1^4\\ &\;\;\vdots\end{aligned}$$

Thus, $(H,\mu,\sigma,\Delta,\epsilon)$ forms a connected graded commutative noncocommutiative bialgebra with $\sigma(1)=\mathbf{1}$, $\epsilon(a_n)=0$ for $n>1$, and $\epsilon(\mathbf{1})=1$. From Theorem 2.10 it follows that this bialgebra is a Hopf algebra. In this case, the reduced coproduct is

$$\begin{aligned}\Delta' a_2 &= 0\otimes 0\\ \Delta' a_3 &= a_2\otimes 3a_2\\ \Delta' a_4 &= a_2\otimes(4a_3+3a_2^2) + a_3\otimes 6a_2\\ &\;\;\vdots\end{aligned}$$

and the antipode computed using (2.29) is:

$$Sa_1 = a_1 \tag{2.37a}$$

$$Sa_2 = -a_2 \tag{2.37b}$$

$$Sa_3 = -a_3 - S(a_2)3a_2 = -a_3 + 3a_2^2 \tag{2.37c}$$

$$\begin{aligned}Sa_4 &= -a_4 - S(a_2)(4a_3+3a_2^2) - S(a_3)6a_2\\ &= -a_4 - (-a_2)(4a_3+3a_2^2) - (-a_3+3a_2^2)6a_2\\ &= -a_4 + 4a_2a_3 + 3a_2^3 + 6a_2a_3 - 18a_2^3\end{aligned}$$

$$= -a_4 + 10a_2a_3 - 15a_2^3 \tag{2.37d}$$
$$\vdots$$

Note, in particular, that in the Sa_4 calculation above there is some inter-term cancellation involving the monomial a_2^3. If this calculation is repeated using instead (2.30), one will observe that there is no such cancellations *ever* (see Problem 2.6.6). In fact, it is known that the number of inter-term cancellations increases dramatically with degree, which from a computational point of view is wasteful. The cancellation free *right* antipode formula (2.30) is known in the literature as a *Zimmermann formula*. In addition, observe from (2.32) that

$$(Sa_n)(c) = a_n(c^{-1}), \;\; n \geq 1,$$

where $c^{-1} \in \mathbb{R}[[X]]$ denotes the generating series for f_c^{-1}, that is, $f_{c^{-1}} = f_c^{-1}$. In which case, the antipode of this Hopf algebra in affect yields a recursive form of the Lagrange series inversion formula.

Example 2.23 Consider the function $f_c(z) = \log(1+z)$, where $f_c(0) = 0$ and $f_c'(0) = 1$. Then $f_c^{-1}(z) = \mathrm{e}^z - 1$ so that

$$f_c(z) = z - \frac{z^2}{2!} + 2\frac{z^3}{3!} - 6\frac{z^4}{4!} + 24\frac{z^5}{5!} + \cdots$$

and

$$f_{c^{-1}}(z) = z + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{z^4}{4!} + \frac{z^5}{5!} + \cdots$$

Observe that

$$\begin{aligned}
a_1(c^{-1}) &= a_1(c) = 1 \\
a_2(c^{-1}) &= -a_2(c) = -(-1) = 1 \\
a_3(c^{-1}) &= -a_3(c) + 3a_2^2(c) = -(2) + 3(-1)^2 = 1 \\
a_4(c^{-1}) &= -a_4(c) - 15a_2^3(c) + 10a_2(c)a_3(c) \\
&= -(-6) + 10(-1)(2) - 15(-1)^3 = 1 \\
&\;\;\vdots
\end{aligned}$$

as expected. □

Finally, recall that the set of characters $\{\Phi_c : H \to \mathbb{R} : c \in \mathbb{R}_p\,[[X]], (c,x) = 1\}$ forms a group under convolution. From (2.33)-(2.34) it follows for $n \geq 1$ that

$$\begin{aligned}(\Phi_c \star \Phi_d)(a_n) &= a_n(c \circ d) = (c \circ d, x^n)\\ \Phi_c^{\star -1}(a_n) &= (Sa_n)(c) = (c^{-1}, x^n).\end{aligned}$$

Therefore, G_{FdB} is isomorphic to the character group associated with the Hopf algebra H.

2.7 Composition Products

Motivated by the final example in the previous section, a *class* of formal power series products known as composition products is considered in this section. They all come from the same basic construction process. involving two alphabet $X = \{x_0, x_1, \ldots, x_m\}$ and $\tilde{X} = \{\tilde{x}_0, \tilde{x}_1, \ldots, \tilde{x}_{\tilde{m}}\}$ and two formal power series $c \in \mathbb{R}^{\tilde{\ell}}\langle\langle \tilde{X}\rangle\rangle$ and $d \in \mathbb{R}^{\ell}\langle\langle X\rangle\rangle$. Mathematically there is no need for any type of *compatibility* between the parameters m, ℓ, $\tilde{m}$ and $\tilde{\ell}$. But in many applications there generally is some kind of natural matching requirement such as $\ell = \tilde{m}$.

Definition 2.20 *Fix two alphabets, $X = \{x_0, x_1, \ldots, x_m\}$ and $\tilde{X} = \{\tilde{x}_0, \tilde{x}_1, \ldots, \tilde{x}_{\tilde{m}}\}$, and assume that $\mathbb{R}\langle\langle X\rangle\rangle$ is an associative $\mathbb{R}$-algebra with product $\square$ and multiplicative identity element $\mathbf{1}$. The associated **composition product** is the binary operation*

$$\begin{aligned}\mathbb{R}^{\tilde{\ell}}\langle\langle \tilde{X}\rangle\rangle \times \mathbb{R}^{\ell}\langle\langle X\rangle\rangle &\to \mathbb{R}^{\tilde{\ell}}\langle\langle X\rangle\rangle\\ (c,d) &\mapsto c \circ d = \sum_{\tilde{\eta} \in \tilde{X}^*} (c, \tilde{\eta})\, \tilde{\eta} \circ d,\end{aligned}$$

where $\tilde{\eta} \circ d$ is the unique extension of

$$\tilde{x}_i \circ d = \rho_i(d), \;\; i = 0, 1, \ldots, \tilde{m}$$

to $\tilde{X}^*$ given by

$$(\tilde{x}_{i_k} \tilde{x}_{i_{k-1}} \cdots \tilde{x}_{i_1}) \circ d = \rho_{i_k}(d) \square \rho_{i_{k-1}}(d) \square \cdots \square \rho_{i_1}(d)$$

with $\rho_i : \mathbb{R}^{\ell}\langle\langle X\rangle\rangle \to \mathbb{R}\langle\langle X\rangle\rangle$ such that $\rho_i(\emptyset) = \mathbf{1}$, $i = 0, 1, \ldots, \tilde{m}$ and $\emptyset \circ d = \mathbf{1}$.

Composition products arise in many forms when systems are interconnected to produce new systems. Their particular form depends on the nature of the systems involved. Whenever possible, the same notation will be used, and the specific definition will be evident from the context. It is easily verified that any composition product is $\mathbb{R}$-linear in its left argument, that is, for any real numbers α and β

$$(\alpha c + \beta d) \circ e = \alpha(c \circ e) + \beta(d \circ e),$$

but in general $c \circ (\alpha d + \beta e) \neq \alpha(c \circ d) + \beta(c \circ e)$ (see Problem 2.7.1). Before tackling the more technical issues, such as the conditions under which a composition product is well defined, some important examples are introduced. These examples will appear frequently in later chapters as they are all inspired by system interconnections.

Example 2.24 Suppose $X = \{x_1, x_2, \ldots, x_m\}$, $\tilde{X} = \{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_{\tilde{m}}\}$, $c \in \mathbb{R}^{\tilde{\ell}}\langle\langle \tilde{X} \rangle\rangle$ and $d \in \mathbb{R}^{\tilde{m}}\langle\langle X \rangle\rangle$. Note that number of components series in d, namely $\tilde{m}$, is equal to the number of letters in $\tilde{X}$. Let d_i denote the i-th component series of d, that is, $(d_i, \xi) = (d, \xi)_i$ for every $\xi \in X^*$. Consider the composition product defined by letting $\rho_i(d) = d_i$, $i = 1, \ldots, \tilde{m}$, and

$$(\underbrace{\tilde{x}_{i_1}\tilde{x}_{i_2}\cdots\tilde{x}_{i_k}}_{\tilde{\eta}}) \circ d = d_{i_1} d_{i_2} \cdots d_{i_k} =: d^{\tilde{\eta}}.$$

In which case,

$$c \circ d = \sum_{\tilde{\eta} \in \tilde{X}^*} (c, \tilde{\eta})\, d^{\tilde{\eta}}.$$

If the letters in $\tilde{X}$ commute, then the composition product is written in the exponential form

$$c \circ d = \sum_{\tilde{\eta} \in \tilde{X}^*} (c, \tilde{\eta})\, \frac{d^{\tilde{\eta}}}{\tilde{\eta}!}$$

(review Example 2.7). If, in addition, the letters in the alphabet X also commute, then the power $d^{\tilde{\eta}}$ is taken to be the multinomial catenation power having coefficients

$$(d^{\tilde{\eta}}, \eta) = \sum_{\eta_1 \cdots \eta_k = \eta} (d_{i_1}, \eta_1) \cdots (d_{i_k}, \eta_k)\, \frac{\eta!}{\eta_1! \cdots \eta_k!}, \quad \eta \in X^*.$$

This type of composition product describes function composition in the sense described in the previous section, but in the multivariable setting. That is, suppose U and V are two neighborhoods of the origin. Let $f_c : U \subset \mathbb{R}^{\tilde{m}} \to \mathbb{R}^{\tilde{\ell}}$ and $f_d : V \subset \mathbb{R}^m \to \mathbb{R}^{\tilde{m}}$ be real analytic functions with $f_d(V) \subset U$ and having Taylor series about $\tilde{z} = 0$ and $z = 0$,

$$
\begin{aligned}
f_c(\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_{\tilde{m}}) &= \sum_{\tilde{\eta} \in \tilde{X}^*} (c, \tilde{\eta}) \frac{\tilde{z}^{\tilde{\eta}}}{\tilde{\eta}!} \\
f_d(z_1, z_2, \ldots, z_m) &= \sum_{\eta \in X^*} (d, \eta) \frac{z^{\eta}}{\eta!},
\end{aligned}
$$

respectively. Here $c \in \mathbb{R}^{\tilde{\ell}}\,[[\tilde{X}]]$, and $d \in \mathbb{R}^{\tilde{m}}\,[[X]]$ is assumed to be proper, i.e., $f_d(0) = 0$.[9] The composite function $f_c \circ f_d$ corresponds to setting $\tilde{z}_i = f_{d,i}(z)$, $i = 1, 2, \ldots, \tilde{m}$. By direct substitution observe

$$
\begin{aligned}
& f_c \circ f_d\,(z) \\
&= \sum_{\tilde{\eta} \in \tilde{X}^*} (c, \tilde{\eta}) \left. \frac{\tilde{z}^{\tilde{\eta}}}{\tilde{\eta}!} \right|_{\tilde{z} = f_d(z)} \\
&= \sum_{\tilde{\eta} \in \tilde{X}^*} (c, \tilde{\eta}) \frac{f_{d,i_1}(z) \cdots f_{d,i_k}(z)}{\tilde{\eta}!} \\
&= \sum_{\tilde{\eta} \in \tilde{X}^*} (c, \tilde{\eta}) \frac{1}{\tilde{\eta}!} \left[\sum_{\eta_1 \in X^*} (d_{i_1}, \eta_1) \frac{z^{\eta_1}}{\eta_1!} \right] \cdots \left[\sum_{\eta_k \in X^*} (d_{i_k}, \eta_k) \frac{z^{\eta_k}}{\eta_k!} \right] \\
&= \sum_{\tilde{\eta} \in \tilde{X}^*} (c, \tilde{\eta}) \frac{1}{\tilde{\eta}!} \left[\sum_{\eta_1, \ldots, \eta_k \in X^*} (d_{i_1}, \eta_1) \cdots (d_{i_k}, \eta_k) \frac{z^{\eta_1} \ldots z^{\eta_k}}{\eta_1! \cdots \eta_k!} \right] \\
&= \sum_{\tilde{\eta} \in \tilde{X}^*} (c, \tilde{\eta}) \frac{1}{\tilde{\eta}!} \left[\sum_{\eta \in X^*} \left[\sum_{\eta_1 \cdots \eta_k = \eta} (d_{i_1}, \eta_1) \cdots (d_{i_k}, \eta_k) \frac{\eta!}{\eta_1! \cdots \eta_k!} \right] \frac{z^{\eta}}{\eta!} \right] \\
&= \sum_{\tilde{\eta} \in \tilde{X}^*} (c, \tilde{\eta}) \frac{1}{\tilde{\eta}!} \left[\sum_{\eta \in X^*} (d^{\tilde{\eta}}, \eta) \frac{z^{\eta}}{\eta!} \right]
\end{aligned}
$$

[9] Otherwise, it is more natural to write f_c as a Taylor series about $\tilde{z}_0 = (d, \emptyset)$ so that $f_c(\tilde{z}_0) = (c, \emptyset)$, i.e., this value is not determined by an infinite sum.

$$= \sum_{\eta \in X^*} \left[\sum_{\tilde{\eta} \in \tilde{X}^*} (c, \tilde{\eta})\, (d^{\tilde{\eta}}, \eta) \frac{1}{\tilde{\eta}!} \right] \frac{z^\eta}{\eta!}$$
$$= \sum_{\eta \in X^*} (c \circ d, \eta) \frac{z^\eta}{\eta!}$$
$$= f_{c \circ d}(z).$$

In which case, $f_c \circ f_d = f_{c \circ d}$. So the underlying composition product for the commutative alphabets X and $\tilde{X}$ is that induced by function composition.

□

Example 2.25 In this example, the situation is mixed. One alphabet is commutative, while the other is not. Let $X = \{x_0, x_1, \ldots, x_m\}$, $\tilde{X} = \{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_{\tilde{m}}\}$, $c \in \mathbb{R}^\ell\,[[\tilde{X}]]$ and $d \in \mathbb{R}^{\tilde{m}}\langle\langle X \rangle\rangle$. If $\rho_i(d) = d_i$, $i = 1, \ldots, \tilde{m}$, and

$$(\underbrace{\tilde{x}_{i_1} \tilde{x}_{i_2} \cdots \tilde{x}_{i_k}}_{\tilde{\eta}}) \circ d = d_{i_1} \sqcup\!\sqcup\, d_{i_2} \sqcup\!\sqcup \cdots \sqcup\!\sqcup\, d_{i_k} = d^{\sqcup\!\sqcup\, \tilde{\eta}}$$

then

$$c \circ d = \sum_{\tilde{\eta} \in \tilde{X}^*} (c, \tilde{\eta})\, \frac{d^{\sqcup\!\sqcup\, \tilde{\eta}}}{\tilde{\eta}!}.$$

This type of composition product describes the interconnection of an integral operator $E_p \in \mathcal{E}(\mathbb{R}\langle X \rangle)$, as described in Section 2.4, followed by a function f_c, specifically, $f_c \circ E_p = E_{c \circ p}$. The composite system will be called a *Wiener-Fliess system* in Chapter 3, where in general the polynomial p can be replaced with a formal power series d to produce a well defined map F_d. A special case of such compositions is the class of state space systems considered in Chapter 6. In this case, F_d will represent the solution to the state equation, and f_c will be the output function. So the mapping from input to output is given by the composition $f_c \circ F_d = F_{c \circ d}$.

□

Example 2.26 This example describes a type of composition product involving only a single noncommutative alphabet, $X = \{x_0, x_1, \ldots, x_m\}$. Suppose $c \in \mathbb{R}^\ell\langle\langle X \rangle\rangle$ and $d \in \mathbb{R}^m\langle\langle X \rangle\rangle$. Define the family of linear operators

$$\begin{aligned}\psi_d(x_i) &: \mathbb{R}\langle\langle X\rangle\rangle \to \mathbb{R}\langle\langle X\rangle\rangle \\ &: e \mapsto x_0(d_i \sqcup\!\sqcup\, e),\end{aligned} \tag{2.38}$$

$i = 0, 1, \ldots, m$ with $d_0 := \mathbf{1}$. Extend the definition inductively to words by letting

$$\psi_d(x_i x_j) = \psi_d(x_i) \bullet \psi_d(x_j),$$

where '$\bullet$' above denotes operator composition, and $\psi_d(\emptyset) :=$ id. Since this product is associative and $\mathbb{R}$-bilinear, in effect ψ_d is a continuous (in the ultrametric sense) algebra homomorphism mapping the associative algebra $(\mathbb{R}\langle\langle X\rangle\rangle, \mathrm{cat}, \mathbf{1})$ to the associative operator algebra $(\mathrm{End}(\mathbb{R}\langle\langle X\rangle\rangle), \bullet, \mathrm{id})$. In this setting, define a formal power series composition product by setting $\rho_i(d) = \psi_d(x_i)(\mathbf{1})$, $i = 0, 1, \ldots, m$, and

$$\begin{aligned}(x_{i_k} x_{i_{k-1}} \cdots x_{i_1}) \circ d &= \psi_d(x_{i_k}) \bullet \psi_d(x_{i_{k-1}}) \bullet \cdots \bullet \psi_d(x_{i_1})(\mathbf{1}) \\ &= \psi_d(x_{i_k} x_{i_{k-1}} \cdots x_{i_1})(\mathbf{1}).\end{aligned}$$

Therefore,

$$\begin{aligned}c \circ d &= \sum_{\eta \in X^*} (c, \eta)\, \eta \circ d \\ &= \sum_{\eta \in X^*} (c, \eta)\, \psi_d(\eta)(\mathbf{1}).\end{aligned}$$

This type of composition product describes the composition of two integral operators in the class $\mathcal{E}(\mathbb{R}\langle X\rangle)$. Namely, if $E_p, E_q \in \mathcal{E}(\mathbb{R}\langle X\rangle)$ then $E_p \circ E_q = E_{p \circ q}$. For example,

$$(E_{x_1} \circ E_{x_1})[u] = E_{x_1}[E_{x_1}[u]] = E_{x_0 x_1}[u],$$

and in fact

$$x_1 \circ x_1 = \psi_{x_1}(x_1)(\mathbf{1}) = x_0(x_1 \sqcup\!\sqcup\, \mathbf{1}) = x_0 x_1.$$

A variation of this product, say $c \,\tilde{\circ}\, d$, uses the following linear operators

$$\begin{aligned}\phi_d(x_i) &: \mathbb{R}\langle\langle X\rangle\rangle \to \mathbb{R}\langle\langle X\rangle\rangle \\ &: e \mapsto x_i e + x_0(d_i \sqcup\!\sqcup\, e)\end{aligned} \tag{2.39}$$

for $i = 0, 1, 2, \ldots, m$ with $d_0 := 0$. For example, $x_1 \,\tilde{\circ}\, x_1 = x_1 + x_0 x_1$. This product can be used to describe the feedback connection of two integral operators. Both of these composition products will be further developed and applied in Chapter 3. □

The well definedness of a composition product is considered next. The following two theorems will cover all the examples described above. The only difference in their assumptions is the inequalities in their first item. If the inequality is the strict sense, then the composition product is always well defined. If not, then an additional condition involving properness is needed.

Theorem 2.11 *Consider a composition product defined by* $(\rho, \square, \mathbf{1})$, *where*

1. $\mathrm{ord}(\rho_i(d)) > \mathrm{ord}(d)$, $i = 0, 1, \ldots, \tilde{m}$, $d \in \mathbb{R}^{\ell}\langle\langle X\rangle\rangle$
2. $\mathrm{ord}(d \square d') = \mathrm{ord}(d) + \mathrm{ord}(d')$, $d, d' \in \mathbb{R}\langle\langle X\rangle\rangle$.

For any $c \in \mathbb{R}^{\tilde{\ell}}\langle\langle \tilde{X}\rangle\rangle$ *and* $d \in \mathbb{R}^{\ell}\langle\langle X\rangle\rangle$ *the composition* $c \circ d$ *is a well defined series in* $\mathbb{R}^{\tilde{\ell}}\langle\langle X\rangle\rangle$.

Proof: It suffices to show that the family of formal power series $\{\tilde{\eta} \circ d\}_{\tilde{\eta} \in \tilde{X}^*}$ is locally finite, and hence, summable. For a fixed $d \in \mathbb{R}^{\ell}\langle\langle X\rangle\rangle$ defined the integers $r_i = \mathrm{ord}(\rho_i(d)) - \mathrm{ord}(d) > 0$, $i = 0, 1, \ldots, \tilde{m}$, and $r = \min_i r_i > 0$. Then given any word $\tilde{\eta} \in \tilde{X}^+$:

$$
\begin{aligned}
\mathrm{ord}(\tilde{\eta} \circ d) &= \mathrm{ord}(\underbrace{\tilde{x}_{i_k} \tilde{x}_{i_{k-1}} \cdots \tilde{x}_{i_1}}_{\tilde{\eta}} \circ d) \\
&= \mathrm{ord}(\rho_{i_k}(d) \square \rho_{i_{k-1}}(d) \square \cdots \square \rho_{i_1}(d)) \\
&= \sum_{j=1}^{k} \mathrm{ord}(\rho_{i_j}(d)) \\
&= \sum_{j=1}^{k} \mathrm{ord}(d) + r_{i_j} \\
&\geq |\tilde{\eta}| \, (\mathrm{ord}(d) + r).
\end{aligned}
$$

Since $\mathrm{ord}(d) + r \geq 1$, $\mathrm{ord}(\tilde{\eta} \circ d)$ increases at least proportionally as the length of $\tilde{\eta}$ is increased. So for a fixed $\xi \in X^*$ the set

$$I_d(\xi) := \{\tilde{\eta} \in \tilde{X}^* : (\tilde{\eta} \circ d, \xi) \neq 0\}$$

must be finite since $(\tilde{\eta} \circ d, \xi) = 0$ when the length of $\tilde{\eta}$ is such that

$$|\tilde{\eta}| \, (\mathrm{ord}(d) + r) > |\xi| \, .$$

In which case, the family in question is locally finite. ■

Theorem 2.12 *Consider a composition product defined by* $(\rho, \square, \mathbf{1})$, *where*

1. $\mathrm{ord}(\rho_i(d)) \geq \mathrm{ord}(d)$, $i = 0, 1, \ldots, \tilde{m}$, $d \in \mathbb{R}^{\ell}\langle\langle X \rangle\rangle$
2. $\mathrm{ord}(d \square d') = \mathrm{ord}(d) + \mathrm{ord}(d')$, $d, d' \in \mathbb{R}\langle\langle X \rangle\rangle$.

For any $c \in \mathbb{R}^{\tilde{\ell}}\langle\langle \tilde{X} \rangle\rangle$ *and a proper* $d \in \mathbb{R}^{\ell}\langle\langle X \rangle\rangle$ *the composition* $c \circ d$ *is a well defined series in* $\mathbb{R}^{\tilde{\ell}}\langle\langle X \rangle\rangle$.

Proof: Following the same logic as in the proof of the previous theorem, one can conclude here that

$$\mathrm{ord}(\tilde{\eta} \circ d) \geq |\tilde{\eta}| \; \mathrm{ord}(d).$$

Since d is proper, $\mathrm{ord}(d) \geq 1$. In which case,

$$I_d(\xi) = \{\tilde{\eta} \in \tilde{X}^* : (\tilde{\eta} \circ d, \xi) \neq 0\}$$

is finite since $(\tilde{\eta} \circ d, \xi) = 0$ when the length of $\tilde{\eta}$ is such that

$$|\tilde{\eta}| \; \mathrm{ord}(d) > |\xi| \, .$$

This proves the theorem. ■

Example 2.27 Suppose $\rho_i(d) = d_i$, $i = 1, 2, \ldots, \tilde{m}$ as in Examples 2.24 and 2.25. Clearly, $\mathrm{ord}(\rho_i(d)) = \mathrm{ord}(d_i) \geq \mathrm{ord}(d)$ as required by Theorem 2.12. Furthermore, the catenation product and shuffle product both satisfy item *2* in Theorem 2.12 (see Lemma 2.2 in the latter case). Therefore, the corresponding composition products are well defined if d is proper. □

Example 2.28 Consider the first composition product defined in Example 2.26. From Lemma 2.2 it follows that for any $e \in \mathbb{R}\langle\langle X \rangle\rangle$

$$\mathrm{ord}(\psi_d(x_i)(e)) = \mathrm{ord}(d_i) + \mathrm{ord}(e) + 1.$$

Therefore,

$$\mathrm{ord}(\rho_i(d)) = \mathrm{ord}(\psi_d(x_i)(\mathbf{1})) = \mathrm{ord}(d_i) + 1 > \mathrm{ord}(d), \;\; i = 0, 1, \ldots, m.$$

Furthermore,

$$\begin{aligned}\operatorname{ord}(\psi_d(x_i)\bullet\psi_d(x_j)(\mathbf{1})) &= \operatorname{ord}(x_0(d_i \sqcup\!\sqcup (x_0(d_j \sqcup\!\sqcup \mathbf{1})))) \\ &= \operatorname{ord}(d_i)+1+\operatorname{ord}(d_j)+1 \\ &= \operatorname{ord}(\psi_d(x_i)(\mathbf{1}))+\operatorname{ord}(\psi_d(x_j)(\mathbf{1})).\end{aligned}$$

In which case, Theorem 2.11 applies, and the composition product $c\circ d$ is well defined everywhere on $\mathbb{R}^\ell\langle\langle X\rangle\rangle\times\mathbb{R}^m\langle\langle X\rangle\rangle$. $\square$

The next theorem gives conditions under which a composition product is associative. Of course, this only makes sense if the composition is defined for two series coming from the same underlying set. The theorem is stated here under the simplifying assumption that all the series are coming from $\mathbb{R}\langle\langle X\rangle\rangle$ with $X=\{x_0,x_1\}$. It can be stated and proved in a much more general setting, like that of Definition 2.20, but this only complicates the notation while obscuring the fundamental idea.

Theorem 2.13 *Consider a composition product defined by* $(\rho,\square,\mathbf{1})$ *on* $\mathbb{R}\langle\langle X\rangle\rangle\times\mathbb{R}\langle\langle X\rangle\rangle$ *with* $X=\{x_0,x_1\}$. *If for every* $c,d,e\in\mathbb{R}\langle\langle X\rangle\rangle$

$$(c\square d)\circ e=(c\circ e)\square(d\circ e), \tag{2.40}$$

then the composition product is associative.

Proof: It is first shown by induction that for any word $\eta\in X^*$ and series $d,e\in\mathbb{R}\langle\langle X\rangle\rangle$

$$(\eta\circ d)\circ e=\eta\circ(d\circ e).$$

If $\eta=\emptyset$ then directly

$$(\emptyset\circ d)\circ e=\mathbf{1}\circ e=\mathbf{1}=\emptyset\circ(d\circ e).$$

Now suppose the claim holds for words up to some fixed length $k\geq 0$. Select any $x_i\in X$, $\eta\in X^k$ and observe from Definition 2.20 and (2.40) that

$$\begin{aligned}((x_i\eta)\circ d)\circ e &= ((x_i\circ d)\square(\eta\circ d))\circ e \\ &= ((x_i\circ d)\circ e)\square((\eta\circ d)\circ e).\end{aligned}$$

Applying the induction hypothesis and Definition 2.20 one more time gives

$$\begin{aligned}((x_i\eta)\circ d)\circ e &= (x_i\circ(d\circ e))\Box(\eta\circ(d\circ e))\\ &= (x_i\eta)\circ(d\circ e).\end{aligned}$$

Hence, the proposed identity holds for all $\eta\in X^*$. Finally, for any $c\in\mathbb{R}\langle\langle X\rangle\rangle$ it follows that

$$\begin{aligned}(c\circ d)\circ e &= \left(\sum_{\eta\in X^*}(c,\eta)\,\eta\circ d\right)\circ e = \sum_{\eta\in X^*}(c,\eta)(\eta\circ d)\circ e\\ &= \sum_{\eta\in X^*}(c,\eta)\,\eta\circ(d\circ e) = c\circ(d\circ e).\end{aligned}$$

Therefore, the composition product on $\mathbb{R}\langle\langle X\rangle\rangle$ is associative. ■

Example 2.29 Consider the composition product defined in Example 2.24. In this case, the product $\Box$ corresponds to the catenation product, which is bilinear, and $\rho_i(d)=d_i$. Therefore, it is sufficient to check a reduced version (2.40), namely,

$$(\eta\Box\xi)\circ e = (\eta\circ e)\Box(\xi\circ e),\quad \eta,\xi\in X^*. \tag{2.41}$$

Observe that if $\eta=x_{i_k}\cdots x_{i_1}$ and $\xi=x_{j_\ell}\cdots x_{j_1}$, then

$$\begin{aligned}(\eta\xi)\circ e &= ((x_{i_k}\cdots x_{i_1})(x_{j_\ell}\cdots x_{j_1}))\circ e\\ &= \rho_{i_k}(e)\cdots\rho_{i_1}(e)\rho_{j_\ell}(e)\cdots\rho_{j_1}(e)\\ &= e_{i_k}\cdots e_{i_1}e_{j_\ell}\cdots e_{j_1}\\ &= (\rho_{i_k}(e)\cdots\rho_{i_1}(e))(\rho_{j_\ell}(e)\cdots\rho_{j_1}(e))\\ &= (\eta\circ e)(\xi\circ e).\end{aligned}$$

Hence, this composition product is associative, which comes as no surprise since it is induced by function composition, which is well known to be associative. □

Example 2.30 Consider the first composition product defined in Example 2.26. Here $\Box$ is effectively operator composition on $\mathrm{End}(\mathbb{R}\langle\langle X\rangle\rangle)$, which is $\mathbb{R}$-bilinear, and $\rho_i(d) = \psi_d(x_i)(\mathbf{1})$, $i = 0, 1$. So again it is sufficient to check (2.41), which in this case is

$$(\psi_d(\eta) \bullet \psi_d(\xi)(\mathbf{1})) \circ e = \psi_{d\circ e}(\eta) \bullet \psi_{d\circ e}(\xi)(\mathbf{1}). \tag{2.42}$$

Suppose, for example, that $\eta = \xi = x_1$. Then, using the identity in Problem 2.7.7(d), it follows that

$$\begin{aligned}(\psi_d(x_1) \bullet \psi_d(x_1)(\mathbf{1})) \circ e &= (x_0(d_1 \sqcup\!\sqcup (x_0(d_1 \sqcup\!\sqcup \mathbf{1})) \circ e \\ &= x_0((d_1 \circ e) \sqcup\!\sqcup (x_0(d_1 \circ e)) \\ &= \psi_{d\circ e}(x_1) \bullet \psi_{d\circ e}(x_1)(\mathbf{1}),\end{aligned}$$

which is in agreement with (2.42). The general case can be proved by induction, and thus, this composition product is associative (see Problem 2.7.4).

In contrast, the second composition product in Example 2.26 does not satisfy the condition (2.41) and is in fact *not* associative. Continuing the earlier example, it is evident that

$$(x_1 \,\tilde{\circ}\, x_1) \,\tilde{\circ}\, x_1 = x_1 + 2x_0x_1 + x_0^2x_1,$$

while

$$x_1 \,\tilde{\circ}\, (x_1 \,\tilde{\circ}\, x_1) = x_1 + x_0x_1 + x_0^2x_1.$$

□

The section is concluded by developing a bit further the first composition product described in Example 2.26. As indicated earlier, it will appear shortly in the context of system interconnections. In the analysis that follows, it is useful to write an arbitrary $\eta \in X^*$ in the form

$$\eta = x_0^{n_k} x_{i_k} x_0^{n_{k-1}} x_{i_{k-1}} \cdots x_0^{n_1} x_{i_1} x_0^{n_0}, \tag{2.43}$$

where $i_j \neq 0$ for $j = 1, \dots, k$. In which case,

$$\begin{aligned}\eta \circ d &= \psi_d(x_0^{n_k}) \bullet \psi_d(x_{i_k}) \bullet \psi_d(x_0^{n_{k-1}}) \bullet \psi_d(x_{i_{k-1}}) \bullet \cdots \\ &\qquad \bullet \psi_d(x_0^{n_1}) \bullet \psi_d(x_{i_1}) \bullet \psi_d(x_0^{n_0}) \\ &= x_0^{n_k+1}[d_{i_k} \sqcup\!\sqcup x_0^{n_{k-1}+1}[d_{i_{k-1}} \sqcup\!\sqcup \cdots x_0^{n_1+1}[d_{i_1} \sqcup\!\sqcup x_0^{n_0}] \cdots]].\end{aligned}$$

It is easily verified that

$$\begin{aligned}
\operatorname{ord}(\eta\circ d) &= n_0+k+\sum_{j=1}^{k} n_j+\operatorname{ord}(d_{i_j})\\
&= |\eta| + \sum_{j=1}^{|\eta|-|\eta|_{x_0}} \operatorname{ord}(d_{i_j}) \qquad (2.44)\\
&\geq |\eta| + (|\eta|-|\eta|_{x_0})\operatorname{ord}(d). \qquad (2.45)
\end{aligned}$$

Alternatively, for any $\eta \in X^*$, one can uniquely associate a set of right factors $\{\eta_0, \eta_1, \ldots, \eta_k\}$ by the iteration

$$\eta_{j+1} = x_0^{n_{j+1}} x_{i_{j+1}} \eta_j, \quad \eta_0 = x_0^{n_0}, \quad i_{j+1} \neq 0, \qquad (2.46)$$

so that $\eta = \eta_k$ with $k = |\eta| - |\eta|_{x_0}$. In which case, $\eta \circ d = \eta_k \circ d$, where

$$\eta_{j+1} \circ d = x_0^{n_{j+1}+1}[d_{i_{j+1}} \,⧢\, (\eta_j \circ d)]$$

and $\eta_0 \circ d = x_0^{n_0}$. Then for any $c \in \mathbb{R}^\ell\langle\langle X\rangle\rangle$ and $d \in \mathbb{R}^m\langle\langle X\rangle\rangle$, the composition product can be written using the set of all right factors as described by (2.46). For each word $\eta \in X^i$, the j-th right factor, η_j, has exactly j letters not equal to x_0. Therefore, given any $\nu \in X^*$:

$$(c \circ d, \nu) = \sum_{i=0}^{|\nu|} \sum_{j=0}^{i} \sum_{\eta_j \in X^i} (c, \eta_j)(\eta_j \circ d, \nu). \qquad (2.47)$$

The third summation is understood to be the sum over the set of all possible j-th right factors of words of length i. This set has a familiar combinatoric interpretation. A *composition* of a positive integer N is an ordered set of positive integers $\{a_1, a_2, \ldots, a_K\}$ such that $N = a_1 + a_2 + \cdots + a_K$. (For example, the integer 3 has the compositions $1+1+1$, $1+2$, $2+1$ and 3). For a given N and K, it is well known that there are $\mathcal{C}_K(N) = \binom{N-1}{K-1}$ possible compositions. Now each factor $\eta_j \in X^i$, when written in the form

$$\eta_j = x_0^{n_j} x_{i_j} x_0^{n_{j-1}} x_{i_{j-1}} \cdots x_0^{n_1} x_{i_1} x_0^{n_0},$$

maps to a unique composition of $i+1$ with $j+1$ elements:

$$i+1 = (n_0+1) + (n_1+1) + \cdots + (n_j+1).$$

Thus, there are exactly $\mathcal{C}_{j+1}(i+1)m^j = \binom{i}{j} m^j$ possible factors η_j in X^i, and the total number of terms in the summations appearing in (2.47) is

$$\frac{((m+1)^{|\nu|+1}-1)}{m} \leq \frac{m+1}{m}(m+1)^{|\nu|}.$$

This provides in general a very conservative estimate on the growth rate of the coefficients of $c \circ d$.

Example 2.31 A series $c \in \mathbb{R}^{\ell}\langle\langle X\rangle\rangle$ is called *linear* if

$$\mathrm{supp}(c) \subseteq \{\eta \in X^* : \eta = x_0^{n_1} x_i x_0^{n_0},\ i \in \{1,2,\ldots,m\},\ n_1, n_0 \geq 0\}. \tag{2.48}$$

Since the shuffle product is $\mathbb{R}$-bilinear on the vector space $\mathbb{R}^{\ell}\langle\langle X\rangle\rangle$, it follows for any $\eta = x_0^{n_1} x_i x_0^{n_0}$ that

$$\begin{aligned}
\eta \circ (\alpha d + \beta e) &= x_0^{n_1+1}[(\alpha d + \beta e)_i \sqcup\!\sqcup\, x_0^{n_0}] \\
&= \alpha x_0^{n_1+1}(d_i \sqcup\!\sqcup\, x_0^{n_0}) + \beta x_0^{n_1+1}(e_i \sqcup\!\sqcup\, x_0^{n_0}) \\
&= \alpha(\eta \circ d) + \beta(\eta \circ e).
\end{aligned}$$

Therefore, if c is a linear series then

$$\begin{aligned}
c \circ (\alpha d + \beta e) &= \sum_{\eta \in X^*} (c,\eta)\ \eta \circ (\alpha d + \beta e) \\
&= \sum_{\eta \in X^*} \alpha(c,\eta)\ \eta \circ d + \beta(c,\eta)\ \eta \circ e \\
&= \alpha(c \circ d) + \beta(c \circ e).
\end{aligned}$$

In other words, the composition product (2.47) in this special situation is linear in its right argument as well as its left. □

Additional observations regarding the composition product (2.47) include the fact it is neither commutative nor has an identity element. Therefore, $(\mathbb{R}^{\ell}\langle\langle X\rangle\rangle, \circ)$ and $(\mathbb{R}^{\ell}\langle X\rangle, \circ)$ form only semigroups. A summary of other useful elementary properties is given below.

Lemma 2.5 *The following identities hold for the composition product defined in (2.47):*

1. $0 \circ d = 0$, $\forall d \in \mathbb{R}^m\langle\langle X\rangle\rangle$.
2. $c \circ 0 = c_0 := \sum_{n \geq 0}(c, x_0^n)\ x_0^n$. *(Therefore,* $c \circ 0 = 0$ *if and only if* $c_0 = 0$.*)*
3. $c_0 \circ d = c_0$, $\forall d \in \mathbb{R}^m\langle\langle X\rangle\rangle$. *(In particular,* $\mathbf{1} \circ d = \mathbf{1}$.*)*
4. $c \circ \mathbb{1} = c_{\mathbb{1}} := \sum_{\eta \in X^*}(c,\eta)\ x_0^{|\eta|}$. *(Therefore,* $c \circ \mathbb{1} = c$ *if and only if* $c = c_0$.*)*

Here $\mathbb{1}$ *denotes a column vector where all* m *component series are* $\mathbf{1}$.

Example 2.32 Suppose $c = \frac{1}{2}x_0x_1x_0 - \frac{1}{3}x_1x_0^2$ and $d = x_0$. Observe $\psi_{x_0}(x_0)(e) = x_0(\mathbf{1} \sqcup\!\sqcup e) = x_0e$ and $\psi_{x_0}(x_1)(e) = x_0(x_0 \sqcup\!\sqcup e)$ so that

$$\begin{aligned}\frac{1}{2}x_0x_1x_0 \circ x_0 &= \frac{1}{2}\psi_{x_0}(x_0) \bullet \psi_{x_0}(x_1) \bullet \psi_{x_0}(x_0)(\mathbf{1}) \\ &= \frac{1}{2}x_0^2(x_0 \sqcup\!\sqcup x_0) \\ &= x_0^4\end{aligned}$$

and

$$\begin{aligned}\frac{1}{3}x_1x_0^2 \circ x_0 &= \frac{1}{3}\psi_{x_0}(x_1) \bullet \psi_{x_0}(x_0) \bullet \psi_{x_0}(x_0)(\mathbf{1}) \\ &= \frac{1}{3}x_0(x_0 \sqcup\!\sqcup x_0^2) \\ &= \frac{1}{3}x_0(3x_0^3) \\ &= x_0^4.\end{aligned}$$

Therefore, $c \circ d = 0$. That is, it is possible to have $c \circ d = 0$ when both c and d are *not* zero. □

Example 2.33 Let $X = \{x_0, x_1\}$ and consider the two linear series c and d with $(c, x_0^{n_1}x_1x_0^{n_0}) = (d, x_0^{n_1}x_1x_0^{n_0}) = 0$ for all $n_0 > 0$. Then

$$\begin{aligned}c \circ d &= \sum_{\eta \in X^*} (c, \eta)\ \eta \circ d \\ &= \sum_{i=0}^{\infty} (c, x_0^i x_1)\ x_0^{i+1} d \\ &= \sum_{i,j=0}^{\infty} (c, x_0^i x_1)(d, x_0^j x_1)\ x_0^{i+j+1} x_1.\end{aligned}$$

For any $k \geq 1$ observe

$$\begin{aligned}(c \circ d, x_0^k x_1) &= \sum_{i,j=0}^{\infty} (c, x_0^i x_1)(d, x_0^j x_1)\ (x_0^{i+j+1}x_1, x_0^k x_1) \\ &= \sum_{j=0}^{k-1} (c, x_0^{k-j-1}x_1)(d, x_0^j x_1).\end{aligned}$$

This last expression is the familiar convolution sum (see (1.29)) and is similar to what appears for single letter alphabets under the Cauchy product in Example 2.5. □

Finally, some more advanced properties of the composition product (2.47) are considered. The first theorem states that this composition product on $\mathbb{R}^m\langle\langle X\rangle\rangle \times \mathbb{R}^m\langle\langle X\rangle\rangle$ is continuous in its left argument. (Right argument continuity will be addressed shortly.)

Theorem 2.14 *Let $\{c_i\}_{i\geq 1}$ be a sequence in $\mathbb{R}^m\langle\langle X\rangle\rangle$ with $\lim_{i\to\infty} c_i = c$ in the ultrametric sense. Then $\lim_{i\to\infty}(c_i \circ d) = c \circ d$ for any $d \in \mathbb{R}^m\langle\langle X\rangle\rangle$.*

Proof: Define the sequence of non-negative integers $k_i = \mathrm{ord}(c_i - c)$ for $i \geq 1$. Since c is the limit of the sequence $\{c_i\}_{i\geq 1}$, the sequence $\{k_i\}_{i\geq 1}$ can have no upper bound. Observe that

$$\mathrm{dist}(c_i \circ d, c \circ d) = \sigma^{\mathrm{ord}((c_i - c)\, \circ d)}$$

and, in light of (2.45),

$$\begin{aligned}
\mathrm{ord}((c_i - c)\circ d) &= \mathrm{ord}\left(\sum_{\eta\in \mathrm{supp}(c_i-c)} (c_i - c, \eta)\, \eta \circ d\right) \\
&\geq \min_{\eta\in \mathrm{supp}(c_i-c)} \mathrm{ord}(\eta\circ d) \\
&\geq \min_{\eta\in \mathrm{supp}(c_i-c)} |\eta| + (|\eta| - |\eta|_{x_0})\,\mathrm{ord}(d) \\
&\geq k_i.
\end{aligned}$$

Thus, $\mathrm{dist}(c_i \circ d, c\circ d) \leq \sigma^{k_i}$ for all $i \geq 1$, and $\lim_{i\to\infty}(c_i \circ d) = c\circ d$. ■

The next theorem describes an ultrametric contraction induced on $\mathbb{R}\langle\langle X\rangle\rangle$ by this composition product.

Theorem 2.15 *For any $c \in \mathbb{R}^m\langle\langle X\rangle\rangle$, the mapping $d \mapsto c\circ d$ is an ultrametric contraction on $\mathbb{R}^m\langle\langle X\rangle\rangle$. Specifically,*

$$\mathrm{dist}(c\circ d, c\circ e) \leq \sigma\, \mathrm{dist}(d, e), \quad \forall d, e \in \mathbb{R}^m\langle\langle X\rangle\rangle.$$

Proof: First observe that the claim is exactly equivalent to the inequality

$$\operatorname{ord}(c \circ d - c \circ e) \geq 1 + \operatorname{ord}(d - e), \tag{2.49}$$

which is trivially true when $c = 0$ since $c \circ d - c \circ e = 0$. So assume $c \neq 0$ such that $\operatorname{supp}(c)$ is nonempty. Now if all the words in $\operatorname{supp}(c)$ have the form $\eta = x_0^k$, $k \geq 0$, then the claim is also trivially true since $\eta \circ d - \eta \circ e = 0$ for all $\eta \in \operatorname{supp}(c)$ giving again that $c \circ d - c \circ e = 0$. Thus, further assume that $\operatorname{supp}(c)$ contains at least one word η utilizing one or more letters from the subalphabet $\{x_1, x_2, \ldots, x_m\}$. In which case,

$$\begin{aligned} \operatorname{ord}(c \circ d - c \circ e) &= \operatorname{ord}\left(\sum_{\eta \in X^*} (c, \eta)(\eta \circ d - \eta \circ e) \right) \\ &\geq \min_{\eta \in \operatorname{supp}(c)} \operatorname{ord}(\eta \circ d - \eta \circ e). \end{aligned}$$

From the definition of the composition product, it is clear that the shorted possible word generated by a series of the form $\eta \circ d - \eta \circ e$ has length $\operatorname{ord}(x_i \circ d - x_i \circ e) = \operatorname{ord}(x_0(d_i - e_i)) = 1 + \operatorname{ord}(d_i - e_i)$ for some $i \neq 0$. This would directly establish the equality in (2.49) if this $x_i \in \operatorname{supp}(c)$. But if $x_i \notin \operatorname{supp}(c)$, then this simply means that

$$\min_{\eta \in \operatorname{supp}(c)} \operatorname{ord}(\eta \circ d - \eta \circ e) > 1 + \operatorname{ord}(d - e).$$

Thus, either way, the theorem is proved. ■

An immediate result of this theorem is the *right* argument continuity property alluded to earlier.

Theorem 2.16 *Let $\{d_i\}_{i \geq 1}$ be a sequence in $\mathbb{R}^m\langle\langle X \rangle\rangle$ with $\lim_{i \to \infty} d_i = d$ in the ultrametric sense. Then $\lim_{i \to \infty}(c \circ d_i) = c \circ d$ for all $c \in \mathbb{R}^m\langle\langle X \rangle\rangle$.*

Proof: Observe

$$\lim_{i \to \infty} \operatorname{dist}(c \circ d_i, c \circ d) \leq \sigma \lim_{i \to \infty} \operatorname{dist}(d_i, d) = 0.$$

■

Problems

Section 2.1

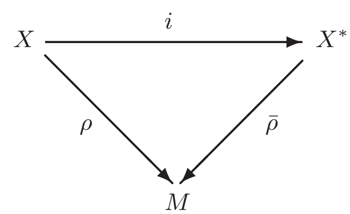


Fig. 2.10. The mapping $\rho : X \to M$ and its associated monoid homomorphism $\bar{\rho} : X^* \to M$.

Problem 2.1.1 Suppose M is the set of real-valued functions which have a well defined one-sided Laplace transformation $H(s) = \mathscr{L}\{h(t)\}$. Let $M' = \mathscr{L}(M)$.

(a) Describe specifically how M and M' can be given the structure of a monoid.
(b) Is $\mathscr{L} : M \to M'$ a monoid homomorphism, a coding, an isomorphism? Explain.

Problem 2.1.2 In the definition for a monoid homomorphism $\rho : M \to M'$, the second of two requirements is that the units e and e' of monoids M and M', respectively, must be related by $\rho(e) = e'$. Can this identity be deduced from the first requirement in the definition, i.e., is the definition redundant? If true, prove it. If false, provide a specific counterexample.

Problem 2.1.3 Given an alphabet $X = \{x_0, x_1, \ldots, x_m\}$ and a monoid M, show that any mapping $\rho : X \to M$ can be uniquely extended to produce a monoid homomorphism $\bar{\rho} : X^* \to M$. That is, if $i : X \to X^*$ denotes the natural injection of X into X^*, then there exists a unique monoid homomorphism $\bar{\rho}$ such that $\rho = \bar{\rho} \circ i$ (see Figure 2.10).

Section 2.2

Problem 2.2.1 Verify that $\mathbb{R}\langle\langle X \rangle\rangle$ and $\mathbb{R}\langle X \rangle$ with the usual notions of addition, scalar multiplication, and the catenation product each constitute:

(a) a ring,

(b) a module over $\mathbb{R}\langle X\rangle$,
(c) an associative $\mathbb{R}$-algebra.

Problem 2.2.2 Let $X = \{x\}$ and assume f_c and f_d are real analytic functions with generating series $c, d \in \mathbb{R}[[X]]$, respectively.

(a) Verify that for the pointwise product $f_c f_d = f_{cd}$, where cd is the binomial convolution product defined in (2.2).
(b) Show that $x^{-1}(cd) = x^{-1}(c)d + cx^{-1}(d)$.
(c) Is this identity above consistent with Lemma 2.1? Explain.

Problem 2.2.3 Prove the following propositions for an arbitrary non-commutative alphabet $X = \{x_0, x_1, \ldots, x_m\}$:

(a) The left-shift operator $\xi^{-1}(\cdot)$ is a linear operator on the $\mathbb{R}$-vector space $\mathbb{R}^\ell\langle\langle X\rangle\rangle$ for any $\xi \in X^*$.
(b) If $p \in \mathbb{R}\langle X\rangle$ then $x_k^{-1}(p) = 0$ for all $x_k \in X$ if and only if $p = (p, \emptyset)\emptyset$.

Section 2.3

Problem 2.3.1 An ultrametric space (S, δ) is *bounded* if there exists a real number $B \geq 0$ such that $\delta(s, s') \leq B$ for all $s, s' \in S$. Show that $(\mathbb{R}^\ell\langle\langle X\rangle\rangle, \mathrm{dist})$ is a bounded ultrametric space.

Problem 2.3.2 Let $\{s_1, s_2, \ldots\}$ be a convergent sequence in a metric space (S, δ).

(a) Show that its limit point s is unique.
(b) Prove that it is a Cauchy sequence.

Problem 2.3.3 Let $\{s_1, s_2, \ldots\}$ be a sequence in an ultrametric space (S, δ). Show that the sequence is a Cauchy sequence if and only if for every $\epsilon > 0$ there exists a natural number N_ϵ such that $\delta(s_i, s_{i+1}) < \epsilon$ when $i \geq N_\epsilon$.

Problem 2.3.4 Let $c, d, e, f \in \mathbb{R}\langle\langle X\rangle\rangle$ and consider the ultrametric dist defined on $\mathbb{R}\langle\langle X\rangle\rangle$. Show that in general:

(a) $\mathrm{dist}(c + e, c + f) = \mathrm{dist}(e, f)$
(b) $\mathrm{dist}(c + e, d + f) \leq \max\{\mathrm{dist}(c, d), \mathrm{dist}(e, f)\}$
(c) dist is continuous in both arguments, for example, $\lim_{i\to\infty} \mathrm{dist}(c_i, d) = \mathrm{dist}(c, d)$, where $c = \lim_{i\to\infty} c_i$.

Problem 2.3.5 Solutions of an equation $g(z) = 0$, where $z \in \mathbb{R}$, can often be computed numerically by converting the equation into the form $z = f(z)$, selecting an initial condition z_0, and then iterating as follows:

$$z_{i+1} = f(z_i), \quad i = 0, 1, \ldots.$$

(a) Assume $g(z) = 0$ has the unique solution z^*. Use Theorem 2.4 to show that a sufficient condition for convergence of this iteration to z^* is having f differentiable and $|f'(z)| \leq \alpha < 1$ everywhere on $\mathbb{R}$.
(b) A simple geometric argument can be used to show that the transcendental equation $4z + 2\sin(z) + 1 = 0$ has a unique solution. Devise an iterative method to compute it and determine the first five iterates after $z_0 = 0$.

Remark: In the event that $g(x) = 0$ has more than one solution, a *local* version of Theorem 2.4 exists. It provides for convergence of the iterates $\{z_0, z_1, \ldots\}$ to a solution when z_0 is selected in an interval containing a solution and on which f is (locally) a contraction. See, for example, [133].

Problem 2.3.6 Let (S, δ) be a complete nonempty metric space and $\mathcal{T} : S \to S$ be a mapping such that $\delta(\mathcal{T}(z), \mathcal{T}(z')) < \delta(z, z')$ for all distinct $z, z' \in S$. Show that if $\mathcal{T}$ has a fixed point, it must be unique.

Problem 2.3.7 Reconsider Problem 2.3.6, where δ is now an ultrametric. Is the given condition enough to guarantee that $\mathcal{T}$ always has a fixed point? Explain.

Problem 2.3.8 A normed $\mathbb{R}$-vector space V is one where there is a real-valued function $\|\cdot\| : V \to \mathbb{R}$ satisfying the properties:

i. $\|x\| \geq 0$
ii. $\|x\| = 0$ if and only if $x = 0$
iii. $\|\alpha x\| = |\alpha| \, \|x\|$
iv. $\|x + y\| \leq \|x\| + \|y\|$

for any $x, y \in V$ and $\alpha \in \mathbb{R}$. For a fixed $R > 0$ define the following subsets of $\mathbb{R}\langle\langle X \rangle\rangle$:

$$S_p(R) = \{c \in \mathbb{R}\langle\langle X \rangle\rangle : \|c\|_p < \infty\},$$

where

$$\|c\|_1 = \sum_{\eta \in X^*} |(c,\eta)| \frac{|\eta|!}{R^{|\eta|}}$$

$$\|c\|_\infty = \sup_{\eta \in X^*} \left\{ |(c,\eta)| \frac{R^{|\eta|}}{|\eta|!} \right\}.$$

(a) Verify that each space $S_p(R)$ is a normed $\mathbb{R}$-vector space.
(b) Give some explicit examples of series that are in and not in $S_1(R)$ and $S_\infty(R)$.
(c) Explain in what sense these spaces are *duals* of each other.
(d) A sequence $\{c_1, c_2, \ldots\}$ in $S_p(R)$ is said to converge to $c \in S_p(R)$ if $\|c_i - c\|_p \to 0$ as $i \to \infty$. Explain any differences between convergence in the ultrametric sense and convergence as defined here.

Section 2.4

Problem 2.4.1 Consider an arbitrary alphabet X. Verify the following propositions:

(a) $\mathbb{R}\langle X\rangle$ with the shuffle product forms an associative $\mathbb{R}$-algebra.
(b) The shuffle product is commutative on $\mathbb{R}\langle X\rangle$.
(c) The shuffle algebra on $\mathbb{R}\langle X\rangle$ is an *integral domain*, that is, $p \sqcup\!\sqcup\, q = 0$ if and only if at least one polynomial is zero.
(d) The shuffle product on $\mathbb{R}\langle\langle X\rangle\rangle$ is (ultrametric) continuous in both arguments, for example, $\lim_{i\to\infty}(c_i \sqcup\!\sqcup\, d) = (\lim_{i\to\infty} c_i) \sqcup\!\sqcup\, d$.

Problem 2.4.2 The shuffle product defined inductively by (2.5) could also be called the *left* shuffle product because at each iteration a left-most letter is extracted from a word and moved to the far left position. Analogously, one could define a *right* shuffle product via

$$(\eta x_i) \sqcap\!\sqcap\, (\xi x_j) = (\eta \sqcap\!\sqcap\, (\xi x_j))x_i + ((\eta x_i) \sqcap\!\sqcap\, \xi)x_j,$$

where $x_i, x_j \in X$, $\eta, \xi \in X^*$ and with $\eta \sqcap\!\sqcap\, \emptyset = \emptyset \sqcap\!\sqcap\, \eta = \eta$. Is it true that $\eta \sqcup\!\sqcup\, \xi = \eta \sqcap\!\sqcap\, \xi$? If so, prove this conjecture. If not, provide a simple counterexample.

Problem 2.4.3 Verify the following relations for arbitrary $c, d \in \mathbb{R}\langle\langle X\rangle\rangle$ and $\nu \in X^*$:

(a) $(c \sqcup\!\sqcup d, \nu) = \sum_{i=0}^{|\nu|} \sum_{\substack{\eta \in X^i \\ \xi \in X^{|\nu|-i}}} (c,\eta)(d,\xi)(\eta \sqcup\!\sqcup \xi, \nu)$

(b) $\sum_{\substack{\eta \in X^i \\ \xi \in X^{|\nu|-i}}} (\eta \sqcup\!\sqcup \xi, \nu) = \binom{|\nu|}{i}, \quad 0 \leq i \leq |\nu|$

(c) $\sum_{\eta,\xi \in X^*} (-1)^{|\xi|}\, (\eta \sqcup\!\sqcup \xi, \nu)\, \eta\tilde{\xi} = \begin{cases} 1 & : \quad |\nu| = 0 \\ 0 & : \quad |\nu| > 0 \end{cases}$

(d) $\sum_{\eta,\xi \in X^*} (-1)^{|\xi|}\, (\eta\xi, \nu)\, \eta \sqcup\!\sqcup \tilde{\xi} = \begin{cases} 1 & : \quad |\nu| = 0 \\ 0 & : \quad |\nu| > 0. \end{cases}$

Remark: Part (b) can be proved using the left-shift operator and the derivation property described in Theorem 2.5. The word $\tilde{\xi}$ denotes ξ with the letters written in the reverse order.

Problem 2.4.4 Suppose the binomial coefficient for two words $\nu, \eta \in X^*$ is defined as

$$\binom{\nu}{\eta} = \sum_{\xi \in X^{|\nu|-|\eta|}} (\eta \sqcup\!\sqcup \xi, \nu)$$

when $|\eta| \leq |\nu|$ and zero otherwise.

(a) Show that

$$\sum_{\eta \in X^i} \binom{\nu}{\eta} = \binom{|\nu|}{i}, \quad i \geq 0.$$

(b) Let A be an arbitrary language in X^* and $a = \mathrm{char}(A)$. Show for any word $\eta \in A$ that

$$(\eta \sqcup\!\sqcup a, \nu) = \binom{\nu}{\eta}.$$

Remark: The integer $\binom{\nu}{\eta}$ is equivalent to the number of times the word η appears as a subword of ν. For example, if $\nu = x_0x_1x_0x_1$ and $\eta = x_0x_1$ then $\binom{\nu}{\eta} = 3$.

Problem 2.4.5 Let X be an arbitrary alphabet. Define the shuffle power of a series $c \in \mathbb{R}\langle\langle X \rangle\rangle$ to be

$$c^{\sqcup\!\sqcup k} = \underbrace{c \sqcup\!\sqcup c \sqcup\!\sqcup \cdots \sqcup\!\sqcup c}_{c \text{ appears } k \text{ times}}, \quad k > 0$$

and $c^{\sqcup\!\sqcup\, 0} = 1$. Verify the following identities for an arbitrary letter $x \in X$ and $c, d \in \mathbb{R}\langle\langle X\rangle\rangle$:

(a) $x^{\sqcup\!\sqcup\, k} = k!\, x^k, \quad k \geq 0$

(b) $x^k \sqcup\!\sqcup\, x^{n-k} = \binom{n}{k} x^n, \quad 0 \leq k \leq n$

(c) $x^{\sqcup\!\sqcup\, k} x^{\sqcup\!\sqcup\, (n-k)} = \binom{n}{k}^{-1} x^{\sqcup\!\sqcup\, n}, \quad 0 \leq k \leq n$

(d) $(x^j)^{\sqcup\!\sqcup\, k} = \dfrac{(jk)!}{(j!)^k} x^{jk}, \quad j, k \geq 0$

(e) $(x^{\sqcup\!\sqcup\, j})^k = \dfrac{(j!)^k}{(jk)!} x^{\sqcup\!\sqcup\, jk}, \quad j, k \geq 0$

(f) $\left(\displaystyle\sum_{j=0}^{\infty} x^j\right)^{\sqcup\!\sqcup\, k} = \displaystyle\sum_{j=0}^{\infty} (kx)^j$

(g) $\left(\displaystyle\sum_{j=0}^{\infty} x^{\sqcup\!\sqcup\, j}\right)^{\sqcup\!\sqcup\, k} = \displaystyle\sum_{j=0}^{\infty} \binom{k+j-1}{j} x^{\sqcup\!\sqcup\, j}$

(h) $x^{-1}(c^{\sqcup\!\sqcup\, k}) = kc^{\sqcup\!\sqcup\, (k-1)} \sqcup\!\sqcup\, x^{-1}(c), \quad k \geq 1$

(i) $x^{-n}(c \sqcup\!\sqcup\, d) = \displaystyle\sum_{k=0}^{n} \binom{n}{k} x^{-k}(c) \sqcup\!\sqcup\, x^{-(n-k)}(d), \quad n \geq 0$

(j) $x^{-1}(\mathrm{e}^{\sqcup\!\sqcup\, d}) = x^{-1}(d) \sqcup\!\sqcup\, \mathrm{e}^{\sqcup\!\sqcup\, d}$, where $\mathrm{e}^{\sqcup\!\sqcup\, d} := \displaystyle\sum_{n=0}^{\infty} \frac{d^{\sqcup\!\sqcup\, k}}{k!}$.

Problem 2.4.6 Let X be an arbitrary alphabet.

(a) Verify the identity (2.6).
(b) Show that $\mathrm{char}(X^k) = (\mathrm{char}(X))^k = (\mathrm{char}(X))^{\sqcup\!\sqcup\, k}/k!$.

Problem 2.4.7 Let $X = \{x_0, x_1\}$. Verify the following identities:

(a) $(\alpha x_0 + \beta x_1)^{\sqcup\!\sqcup\, k} = k!\,(\alpha x_0 + \beta x_1)^k, \quad \alpha, \beta \in \mathbb{R}, \quad k \geq 0$

(b) $(c+d)^{\sqcup\!\sqcup\, n} = \displaystyle\sum_{k=0}^{n} \binom{n}{k} c^{\sqcup\!\sqcup\, k} \sqcup\!\sqcup\, d^{\sqcup\!\sqcup\, (n-k)}, \quad n \geq 0, \quad c, d \in \mathbb{R}\langle\langle X\rangle\rangle$

(c) $(x_1 x_0^i \sqcup\!\sqcup\, x_0^j, x_0^k x_1 x_0^{i+j-k}) = \binom{i+j-k}{i}, \quad i, j \geq 0, \quad 0 \leq k \leq j.$

Problem 2.4.8 Suppose K and M are fixed real numbers. Consider a series $d \in \mathbb{R}\langle\langle X\rangle\rangle$ with coefficients $(d,\nu) = KM^{|\nu|}|\nu|!$ for all $\nu \in X^*$. Show that for any $n \geq 1$

$$(d^{\sqcup\!\sqcup\, n},\nu) = K^n M^{|\nu|}\binom{(n-1)+|\nu|}{n-1}|\nu|!, \quad \forall \nu \in X^*.$$

Problem 2.4.9 Let x be an arbitrary letter in an alphabet X and $k_1, k_2, \ldots, k_\ell$ a finite sequence of nonnegative integers. Verify the following shuffle product identity:

$$\coprod_{i=1}^{\ell} x^{k_i} = x^{S_\ell} \prod_{i=1}^{\ell}\binom{S_i}{k_i},$$

where $S_i := \sum_{j=1}^{i} k_j$. Note here that $\coprod$ and $\prod$ denote the shuffle product on X and scalar product on $\mathbb{R}$, respectively.

Problem 2.4.10 Let X be an arbitrary alphabet. Suppose $c \in \mathbb{R}\langle\langle X\rangle\rangle$ is proper.

(a) Verify that $\lim_{i\to\infty} c^{\sqcup\!\sqcup\, i} = 0$.
(b) Show that the summation $\sum_{i=0}^{\infty} \alpha_i c^{\sqcup\!\sqcup\, i}$ is well defined for any sequence of scalars $\{\alpha_i\}_{i\in\mathbb{N}_0}$.

Problem 2.4.11 Show that the set of non proper series in $\mathbb{R}\langle\langle X\rangle\rangle$ is a group under the shuffle product, where the shuffle inverse of any such series c is

$$c^{\sqcup\!\sqcup\, -1} = ((c,\emptyset)(1-c'))^{\sqcup\!\sqcup\, -1} = (c,\emptyset)^{-1}(c')^{\sqcup\!\sqcup\, *}$$

with $c' = 1 - c/(c,\emptyset)$ proper, and $(c')^{\sqcup\!\sqcup\, *} := \sum_{k\geq 0}(c')^{\sqcup\!\sqcup\, k}$.

Problem 2.4.12 Let $X = \{x_0, x_1, \ldots, x_m\}$. A mapping $\mathcal{T} : \mathbb{R}^\ell\langle\langle X\rangle\rangle \to \mathbb{R}^\ell\langle\langle X\rangle\rangle$ is said to have an *eigen-series* $c_\lambda \in \mathbb{R}\langle\langle X\rangle\rangle$ if there exists a nonzero series $c_p \in \mathbb{R}^\ell\langle\langle X\rangle\rangle$ such that $\mathcal{T}(c_p) = c_\lambda c_p$. For a fixed $n \geq 0$ consider the mapping

$$\mathcal{T}_n : \mathbb{R}^\ell\langle\langle X\rangle\rangle \to \mathbb{R}^\ell\langle\langle X\rangle\rangle : c \mapsto x_0^n \sqcup\!\sqcup\, c,$$

where the shuffle product is defined componentwise. For simplicity assume $\ell = 1$ and $m = 0$. Determine whether the following series have a corresponding eigen-series for any $n \geq 0$:

(a) $c_p = (1 - x_0)^{-1}$
(b) $c_p = \exp(x_0)$.

Section 2.5

Problem 2.5.1 Consider the vector space $V = \mathbb{R}^2 \otimes \mathbb{R}^2$ in Example 2.19.

(a) Explicitly define vector addition and scalar multiplication on V.
(b) Verify that $\mathcal{L} : V \to \mathbb{R}$ as defined in (2.13) is an $\mathbb{R}$-linear map.

Problem 2.5.2 For any finite alphabet X and $x_i \in X$ verify the following identities:

(a) $\mathrm{cat}^*(x_i\eta) = (x_i \otimes \mathbf{1})\mathrm{cat}^*(\eta) + 1 \otimes x_i\eta$

(b) $\mathrm{sh}^*(\mathrm{char}(X^{k+1})) = \mathrm{sh}^*(\mathrm{char}(X))\mathrm{sh}^*(\mathrm{char}(X^k))$

(c) $(k+1)\,\mathrm{cat}^*(\mathrm{char}(X^{k+1})) = \mathrm{cat}^*(\mathrm{char}(X)) \sqcup\!\sqcup\, \mathrm{cat}^*(\mathrm{char}(X^k))$.

Section 2.6

Problem 2.6.1 In the context of Definition 2.14, verify that

$$\mu_{A\otimes A} \circ (\Delta \otimes \Delta) = (\mu \otimes \mu) \circ (\mathrm{id} \otimes \tau \otimes \mathrm{id}) \circ (\Delta \otimes \Delta)$$
$$\sigma_{A\otimes A} = \sigma \otimes \sigma$$

defines an $\mathbb{R}$-algebra homomorphism.

Problem 2.6.2 Let $(A, \mu, \sigma, \Delta, \epsilon)$ denote an arbitrary $\mathbb{R}$-bialgebra with unit $\mathbf{1}$.

(a) Show that $\epsilon(\mathbf{1}) = 1$.
(b) Prove that $\Delta(\mathbf{1}) = \mathbf{1} \otimes \mathbf{1}$.
(c) Verify the identities in parts (a)-(b) in the context of Theorem 2.8.

Problem 2.6.3 Let $(A, \mu, \sigma, \Delta, \epsilon, S)$ denote an arbitrary $\mathbb{R}$-Hopf algebra with unit $\mathbf{1}$.

(a) Verify that $S(\mathbf{1}) = \mathbf{1}$.
(b) Show that if $a \in A$ has the coproduct $\Delta(a) = \mathbf{1} \otimes a + a \otimes \mathbf{1}$ then $S(a) = -a$.
(c) Prove that $S(aa') = S(a')S(a)$ for all $a, a' \in A$.

(d) Verify the identities in parts (a)-(c) in the context of Theorem 2.9.

Problem 2.6.4 Let $(A, \mu, \sigma, \Delta, \epsilon)$ be a connected bialgebra with unit $\mathbf{1}$. The proposition is that every element $a \in A^+_{(n)}$ must have a coproduct of the form

$$\Delta a = a \otimes \mathbf{1} + \mathbf{1} \otimes a + \Delta' a$$

with $a \otimes \mathbf{1} + \mathbf{1} \otimes a$ being the primitive part and $\Delta' a \in A^+_{(n-1)} \otimes A^+_{(n-1)}$ the reduced coproduct.

(a) Show that $(\mathbb{R}\langle X\rangle, \mathrm{cat}, \sigma, \mathrm{sh}^*, \epsilon)$ and $(\mathbb{R}\langle X\rangle, \mathrm{sh}, \sigma, \mathrm{cat}^*, \epsilon)$ are both connected bialgebras using word length to define degree and thus a filtration.
(b) Verify the proposition above for the coproducts sh^* and cat^* of words up to degree (length) two.
(c) Prove the proposition holds in general.

Remark: The counit property is very useful in part (c).

Problem 2.6.5 Let $V = \mathbb{R}^\infty$ be the vector space of infinite sequences of real numbers. Show that G_{FdB} has a faithful representation $\pi : G_{FdB} \to GL(\mathbb{R}^\infty)$ given by

$$\pi(f_c) := \left[\frac{k!}{j!} B_{j,k}(c_1, 2!c_2, \ldots, (j-k+1)!c_{j-k+1})\right]$$
$$= \begin{bmatrix} c_1 & c_2 & c_3 & c_4 & c_5 & \cdots \\ 0 & c_1^2 & 2c_1c_2 & c_2^2 + 2c_1c_3 & 2c_2c_3 + 2c_1c_4 & \cdots \\ 0 & 0 & c_1^3 & 3c_1^2c_2 & 3c_1c_2^2 + 3c_1^2c_3 & \cdots \\ 0 & 0 & 0 & c_1^4 & 4c_1^3c_2 & \cdots \\ 0 & 0 & 0 & 0 & c_1^5 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix},$$

where $f_c = \sum_{n\geq 0} c_n t^n/n!$, $c_1 = 1$, and

$$B_{j,k}(t_1, \ldots, t_l) := \sum_{\substack{k_1+k_2+\cdots+k_l=k \\ k_1+2k_2+\cdots+lk_l=j}} \frac{j!}{k_1! \cdots k_l!} \left(\frac{t_1}{1!}\right)^{k_1} \cdots \left(\frac{t_l}{l!}\right)^{k_l}$$

are the (partial exponential) Bell polynomials with $l = j - k + 1$.

Remark: By definition, the product of two such matrices produces another matrix whose coefficients are that of $f_c \circ f_d$ in the first row. Likewise, the coefficients of f_c^{-1} must appear in the first row of $(\pi(f_c))^{-1}$.

With the structure of the representation matrix $A = \pi(f_c)$ fixed, it is fully specified by $c_n = a_{1n}(c)$, $n \geq 1$. Therefore, one can drop the first subscript and write A in terms of a_n, where $c_n = a_n(c)$.

Problem 2.6.6
In the context of the Faà di Bruno Hopf algebra:

(a) Compute the antipodes Sa_i, $a = 2, 3, 4$ using (2.30) and determine whether the calculation is consistent with (2.37) and is cancellation free as claimed in the text.
(b) Compute the coproduct Δa_5 and reduced coproduct $\Delta' a_5$.
(c) Compute the antipode Sa_5 by any method.
(d) Compute the first five terms of the Taylor series expansion of $\tan^{-1}(z)$ about $z = 0$ using the antipode.
(e) Compare the result in part (d) against a direct calculation of the Taylor series of $\tan^{-1}(z)$.

Section 2.7

Problem 2.7.1 Consider an arbitrary composition product defined by $(\rho, \square, \mathbf{1})$. Show that the composition product is $\mathbb{R}$-linear in its left argument.

Problem 2.7.2 Consider the series

$$c = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$
$$d = x + x^2 + x^3 + x^4 + \cdots$$

as elements of $\mathbb{R}\,[[X]]$, where $X = \{x\}$. Compute the first few terms of the compositions:

(a) $c \circ d$
(b) $d \circ d$
(c) $d \circ c$.

Problem 2.7.3 Consider the noncommutative polynomials $p = 1 + x_1^2$ and $q = x_1 x_0$ over the alphabet $X = \{x_0, x_1\}$. Compute the compositions:

(a) $p \circ q$
(b) $q \circ p$

(c) $p \circ p$.

Remark: This and subsequent problems are referring to the first composition product described in Example 2.26.

Problem 2.7.4 Prove the identity (2.42).

Problem 2.7.5 Verify the elementary properties of the composition product given in Lemma 2.5.

Problem 2.7.6 Let $c, d, e \in \mathbb{R}^m\langle\langle X\rangle\rangle$, where $X = \{x_0, x_1, \ldots, x_m\}$. Either verify that each of the propositions below is true or provide a counterexample.

(a) If c is proper then $c \circ d$ is proper.
(b) For fixed d, the mapping $c \mapsto c \circ d$ is an ultrametric contraction on $\mathbb{R}^m\langle\langle X\rangle\rangle$.

Problem 2.7.7 Let $c, d, e \in \mathbb{R}^m\langle\langle X\rangle\rangle$, where $X = \{x_0, x_1, \ldots, x_m\}$. Verify the following identities:

(a) $x_i^j \circ d = \dfrac{1}{j!}(x_0 d_i)^{\sqcup\!\sqcup\, j}$, $j \geq 0$ assuming $d_0 := \mathbf{1}$
(b) $p \circ d = \displaystyle\sum_{j=0}^{n} (p, x_i^j)\frac{1}{j!}(x_0 d_i)^{\sqcup\!\sqcup\, j}$, where $p = \displaystyle\sum_{j=0}^{n}(p, x_i^j)x_i^j$
(c) $(x_0^j c) \circ d = x_0^j (c \circ d)$, $j \geq 0$
(d) $(c \sqcup\!\sqcup\, d) \circ e = (c \circ e) \sqcup\!\sqcup\, (d \circ e)$
(e) $c^{\sqcup\!\sqcup\, j} \circ d = (c \circ d)^{\sqcup\!\sqcup\, j}$, $j \geq 0$
(f) $x_0^{-1}(c \circ d) = x_0^{-1}(c) \circ d + \displaystyle\sum_{i=1}^{m} d_i \sqcup\!\sqcup\, (x_i^{-1}(c) \circ d)$

$x_i^{-1}(c \circ d) = 0, \quad i = 1, 2, \ldots, m.$

Problem 2.7.8 Let $X = \{x_0, x_1\}$ and $c = x_0 x_1$. Determine the fixed point of the mapping $d \mapsto c \circ d$.

Problem 2.7.9 A series $c \in \mathbb{R}^\ell\langle\langle X\rangle\rangle$ is said to be *exchangeable* if

$$|\eta|_{x_i} = |\xi|_{x_i}, \; i = 0, 1, \ldots, m \;\; \Rightarrow \;\; (c, \eta) = (c, \xi).$$

Show that if c is an exchangeable series then the composition product can be written in the form

$$c \circ d = \sum_{k=0}^{\infty} \sum_{\substack{r_0, \ldots, r_m \geq 0 \\ r_0 + \cdots + r_m = k}} (c, x_0^{r_0} \cdots x_m^{r_m})\, \psi_d(x_0^{r_0})(\mathbf{1}) \sqcup\!\sqcup \cdots \sqcup\!\sqcup\, \psi_d(x_m^{r_m})(\mathbf{1}).$$

Bibliographic Notes

Section 2.1 The following books provide a comprehensive introduction to the theory of formal languages and related topics: Berstel [7], Gross and Lentin [99], Harrison [106], Kuich and Salomaa [134], Rèvèsz [159], Rozenberg and Salomaa [163], and Salomaa [168]. It also worth mentioning the series of books by Lothaire addressing combinatorics on words [142, 143, 144].

Section 2.2 In addition to the books mentioned above, most of which treat formal power series to varying degrees, the books by Conway [46], Berstel and Reutenauer [8, 9], and Salomaa and Soittola [169], in particular, provide excellent introductions to this subject with much the same flavor as this section.

Section 2.3 The first part of this section addressing the ultrametric space $\mathbb{R}^\ell\langle\langle X\rangle\rangle$ is based on the treatment of the subject by Berstel and Reutenauer [8, 9]. The material concerning contractive mappings on metric spaces appears in most texts on functional analysis and linear operators, e.g., [133]. Example 2.14, a well known example introducing a weaker type of contraction, appears in many places, e.g., [196]. More specialized treatments of contractions on ultrametric spaces can be found in the work by Heckmanns [193], Priess et al. [155], and Schorner [173]. These did not influence the presentation in the section, but some of the problems at the end of the chapter were motivated by this material, e.g., Problem 2.3.7.

Section 2.4 The shuffle product as defined in this section first appeared in a paper by Ree [156]. A proof of the integral domain property of the shuffle algebra, for example, appears in this paper. His motivation was clearly the seminal work of K.-T. Chen on iterated integrals of paths, in particular [33]. Some other forms of the definition appeared earlier, as, for example, in the work of Hurwitz, who was effectively considering power series in a single letter [111]. Hence, the shuffle product is sometimes referred to as the *Hurwitz product*. (See [62] for additional details on this point.) It should also be noted that around the same time that Ree's paper appeared, Chen et al. utilized the shuffle product in the context of free differential calculus, see [44], largely inspired by Lyndon's use of the concept in [145]. The now standard treatment of shuffles and the shuffle product appears in the book by Lothaire [142]. Useful results related to these concepts can be found in [52, 69, 158, 160]. The important identity (2.6) can be found in the Ph.D. dissertation of Duffaut Espinosa [52, Lemma IV.2.2]. This result is directly related to Proposition 2.2.8 of Fliess in [61].

Section 2.5 The catenation-shuffle product duality is best understood in the context of combinatorial bialgebras and Hopf algebras. This was observed by K.-T. Chen in [38, Theorem 1.8]. Thus, the full treatment of this topic is deferred to the next section after the notion of a Hopf algebra is developed. The duality theory as described in Theorem 2.6 follows the treatment by Reutenauer in [158].

Section 2.6 Standard references on Hopf algebras include the books by Abe [1], Dăscălescu *et al.* [51], and Sweedler [188]. The classical paper on structural theorems for Hopf algebras by Milnor and Moore provides a full and rigorous view of the subject [151]. The papers by Cartier [27], Figueroa and Gracia-Bondía [60], Grinberg and Reiner [97], and Manchon [150] give very readable comprehensive introductions to the subject. The treatment of the topic in this section was heavily influenced by all of these works. The remaining material on the catenation-shuffle product duality in Theorems 2.8 and 2.9 follows Reutenauer in [158]. Lemma 2.4 is based on the work of Chen in [39] and Sussmann in [186]. The Faà di Bruno Hopf algebra was first introduced by Joni and Rota in [123, 124]. The treatment in this chapter follows from the presentation in [60]. Finally, Hopf algebras have appeared in system theory prior to their use in this book. See, for example, [80, 100, 101, 102, 161].

Section 2.7 The composition product induced by real analytic function composition is a classic topic in analysis. See, for example, the book by Knopp [132] concerning the single variable case and the book by Rudin [165] for the multivariable case. The composition of a Fliess operator followed by a memoryless function, i.e., a Wiener-Fliess system, was first described in a state space setting via the *fundamental formula* of Fliess in [69, 74]. See also [113]. A more general treatment is presented by Gray and Thitsa in [94], and additional results regarding convergence and applications are given by Venkatesh in [195]. The composition product induced by Fliess operator composition is due to Ferfera [58, 59]. The idea was further developed by Gray and Li in [91] and [141], which is the source for most of the advanced material in this section. However, Definition 2.20 and the generic treatment of composition products first appeared in [94]. In addition, the proof of Theorem 2.15 is significantly simpler than the approach taken in [91]. Finally, many any other types of noncommutative compositions appear in the literature, for example, see the work by Brouder et al. [22] and Foissy [75]. These concepts are all distinct from the notions utilized through this book as they are induced by compositions of other types of mathematical objects.

3. Fliess Operators

In this chapter, a general class of nonlinear input-output operators is considered with the key property that each member is uniquely specified in terms of a formal power series in $\mathbb{R}^{\ell}\langle\langle X\rangle\rangle$. Such operators are called *Fliess operators*. They can be viewed as a kind of noncommutative Taylor series. The first section introduces the basic definitions and some terminology. The next two sections furnish input sets on which Fliess operators are well defined and describe various properties like continuity and differentiability of the output function and operator continuity. Fliess operators are then compared against the more classical Volterra operator in the subsequent section. In particular, it is shown that a Volterra operator has a Fliess operator representation when each of its kernel functions is real analytic. Hence, all the theory developed for Fliess operators applies directly to this class of Volterra operators. In applications, it is common to construct models of complex systems by interconnecting simpler subsystems. So the next three sections are devoted to the interconnection of Fliess operators, specifically, the parallel, cascade, and feedback connections. The final section describes the notion of a *formal* Fliess operator. In this case, no convergence properties are assumed a priori, and thus, inputs, outputs, and systems are treated as purely algebraic objects.

3.1 Fliess Operators on $L_{\mathfrak{p}}$ Spaces

The goal of this section is to describe a general class of causal input-output operators having m inputs and ℓ outputs. Consider an alphabet $X = \{x_0, x_1, \ldots, x_m\}$. In this chapter, X is viewed as a set of noncommutative indeterminates which is always in one-to-one correspondence with a set of integrable real-valued functions $\{u_0, u_1, \ldots, u_m\}$ defined over an interval $[t_0, t_1]$. The parameter t_1 may or may not be finite. For any word $\eta \in X^+$, one can associate an iterated integral by the iterative calculation

$$E_\eta[u](t,t_0) = E_{x_i\eta'}[u](t,t_0) = \int_{t_0}^{t} u_i(\tau) E_{\eta'}[u](\tau,t_0)\,d\tau,$$

where $E_\emptyset[u](t,t_0) := 1$ for all $t \in [t_0,t_1]$. It will be assumed throughout that $u_0(t) = 1$ on this interval. This fictitious input u_0 is useful, for example, in representing systems that have some kind of stored energy and thus generate a nonzero output even when the input $u := [u_1 \;\cdots\; u_m]^T$ is exactly zero on $[t_0,t_1]$. Given any formal power series over X,

$$c = \sum_{\eta\in X^*} (c,\eta)\,\eta,$$

where each $(c,\eta) \in \mathbb{R}^\ell$, one can uniquely specify an input-output operator as described below.

Definition 3.1 *The* ***Chen-Fliess series*** *associated with any* $c \in \mathbb{R}^\ell\langle\langle X\rangle\rangle$ *is*

$$y(t) = F_c[u](t) = \sum_{\eta\in X^*} (c,\eta)\, E_\eta[u](t,t_0). \tag{3.1}$$

In the event the series converges on some set of inputs $\mathcal{U}$*, the mapping* $F_c : \mathcal{U} \to \mathcal{Y}$ *is called a* ***Fliess operator****.*

Series c is usually referred to as the *generating series* for F_c. A specific input $u \in \mathcal{U}$ is called an *admissible input*. In many applications, a natural class of admissible inputs is the set of Lebesgue measurable functions $L_{\mathfrak{p}}^m[t_0,t_1]$.[1]

Definition 3.2 *For a fixed* $\mathfrak{p} \in [1,\infty]$*, a measurable function* $u : [t_0,t_1] \to \mathbb{R}^m$ *is in the* ***Lebesgue space*** $L_{\mathfrak{p}}^m[t_0,t_1]$ *if its norm*

$$\|u\|_{\mathfrak{p}} = \max_{1\le i\le m} \|u_i\|_{\mathfrak{p}}$$

is finite, where

$$\|u_i\|_{\mathfrak{p}} = \left(\int_{t_0}^{t_1} |u_i(t)|^{\mathfrak{p}}\,dt\right)^{\frac{1}{\mathfrak{p}}}, \;\; \mathfrak{p} \in [1,\infty)$$

$$\|u_i\|_\infty = \operatorname*{ess\,sup}_{t\in[t_0,t_1]} |u_i(t)|.$$

[1] The superscript m will be suppressed when $m = 1$.

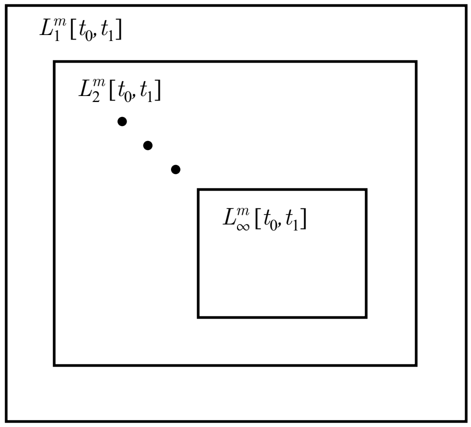


Fig. 3.1. For a finite interval $[t_0, t_1]$, the spaces $L_\mathfrak{p}^m[t_0, t_1]$ for integers $\mathfrak{p} \in [1, \infty)$ and $\mathfrak{p} = \infty$ are nested.

In general, $u \in L_\mathfrak{p}^m[t_0, t_1]$ if and only if $u_i \in L_\mathfrak{p}[t_0, t_1]$ for all $i = 1, \ldots, m$, and clearly $\|u\|_\mathfrak{p} \geq \|u_i\|_\mathfrak{p}$. The set $L_2^m[t_0, t_1]$, for example, is the class of inputs with finite *energy* over $[t_0, t_1]$, while $L_\infty^m[t_0, t_1]$ is the class of inputs that are bounded in magnitude almost everywhere (a.e.) on $[t_0, t_1]$. A *closed ball* in $L_\mathfrak{p}^m[t_0, t_1]$ of radius $R > 0$ and centered at the origin is defined as

$$B_\mathfrak{p}^m(R)[t_0, t_1] = \{u \in L_\mathfrak{p}^m[t_0, t_1] \; : \; \|u\|_\mathfrak{p} \leq R\}.$$

A particularly useful fact illustrated in Figure 3.1 is that this collection of spaces is nested when the interval $[t_0, t_1]$ is finite, i.e., $L_\infty^m[t_0, t_1] \subset L_{\mathfrak{p}+1}^m[t_0, t_1] \subset L_\mathfrak{p}^m[t_0, t_1]$ for all integers $\mathfrak{p} \in [1, \infty)$ (see Problem 3.1.1). As demonstrated in the next example, however, this property does not hold for infinite intervals.

Example 3.1 Consider the function $u(t) = 1/(1 + t)$, which is well defined over the interval $[0, \infty)$. Let $u_{[0,t_1]}$ denote its restriction to the interval $[0, t_1]$. Observe that $\left\|u_{[0,1]}\right\|_1 = \log(2)$ and $\left\|u_{[0,1]}\right\|_2 = 1/\sqrt{2}$. Thus, $u_{[0,1]}$ belongs to both $L_1[0, 1]$ and $L_2[0, 1]$. On the other hand, one can check that $\|u\|_2 = 1$, but u does not have a well defined L_1 norm. Therefore, $L_2[0, \infty)$ is not a subset of $L_1[0, \infty)$. □

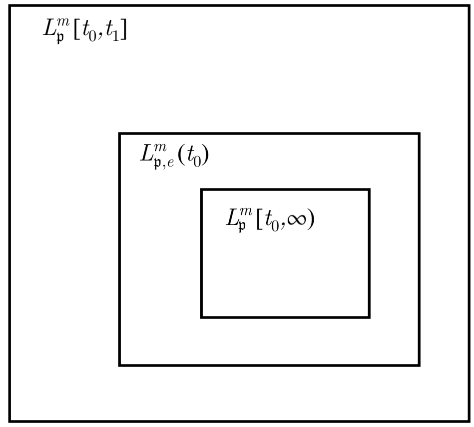


Fig. 3.2. Extended space $L_{\mathfrak{p},e}^m(t_0)$ lies in between $L_{\mathfrak{p}}^m[t_0,t_1]$ with t_1 finite and $L_{\mathfrak{p}}^m[t_0,\infty)$.

Also useful in this chapter is the notion of an *extended* $L_{\mathfrak{p}}$ space. In a certain sense, this function space lies in between $L_{\mathfrak{p}}^m[t_0,t_1]$ with t_1 finite and $L_{\mathfrak{p}}^m[t_0,\infty)$.

Definition 3.3 *For any fixed* $t_0 \in \mathbb{R}$ *and any* $\mathfrak{p} \in [1,\infty]$ *define the* ***extended Lebesgue space*** *as*

$$L_{\mathfrak{p},e}^m(t_0) = \{u : [t_0,\infty) \to \mathbb{R}^m : \; u_{[t_0,t_1]} \in L_{\mathfrak{p}}^m[t_0,t_1] \; \forall t_1 \in (t_0,\infty)\}.$$

Clearly, $L_{\mathfrak{p},e}^m(t_0)$ is a proper subset of $L_{\mathfrak{p}}^m[t_0,t_1]$ for any specific t_1 when its elements are restricted to $[t_0,t_1]$. In addition, the extended spaces are also nested with respect to $\mathfrak{p}$, that is, $L_{\infty,e}^m(t_0) \subset L_{\mathfrak{p}+1,e}^m(t_0) \subset L_{\mathfrak{p},e}^m(t_0)$ for all integers $\mathfrak{p} \in [1,\infty)$. Less obvious is the fact that $L_{\mathfrak{p}}^m[t_0,\infty)$ is a subset of $L_{\mathfrak{p},e}^m(t_0)$ (see Problem 3.1.2).

Example 3.2 Again consider the function $u(t) = 1/(1+t)$ over $[0,\infty)$. For every finite $t_1 > 0$, $u_{[0,t_1]} \in L_1[0,t_1]$ since $\left\|u_{[0,t_1]}\right\|_1 = \log(1+t_1)$. Thus, $u \in L_{1,e}(0)$. But as noted in the previous example, $u \notin L_1[0,\infty)$. So if $L_1[0,\infty)$ is a subset of $L_{1,e}(0)$, it must be a *proper* subset. The exact situation is shown in Figure 3.2. □

3.2 Local Input-Output Properties

The first goal of this section is to describe a sufficient condition for a generating series c which ensures that functions within some closed ball $B_{\mathfrak{p}}^{m}(R)[t_0,t_1]$ are admissible inputs for the operator F_c. Specifically, it will be shown that F_c defines a mapping from $B_{\mathfrak{p}}^{m}(R)[t_0,t_0+T]$ into $B_{\mathfrak{q}}^{\ell}(S)[t_0,t_0+T]$ provided that $R,T>0$ are sufficiently small, and $\mathfrak{p},\mathfrak{q}$ are *conjugate exponents*. That is, $\mathfrak{p},\mathfrak{q}\in[1,\infty]$ such that $1/\mathfrak{p}+1/\mathfrak{q}=1$ with 1 and ∞ being conjugate exponents by convention.[2] In this case the operator F_c will be called *locally convergent* since it is locally well defined in both a temporal sense (finite T) and a spatial sense (finite R). A class of generating series ensuring this property will be given the same name. In the subsequent section, a more restrictive condition is described for c under which F_c maps all of $L_{\mathfrak{p},e}^{m}(t_0)$ into $C[t_0,t_0+T]$, where $C[t_0,t_0+T]$ denotes the set of functions that are continuous on $[t_0,t_0+T]$, and in this case $T>0$ is arbitrary. The operator F_c will be called *global convergent* as will any generating series c which yields this property.

The first theorem states that if a formal power series c has coefficients that satisfy a Cauchy type growth condition then the corresponding operator F_c will converge on $B_1^m(R)[t_0,t_0+T]$ provided that $R,T>0$ are small enough.

Theorem 3.1 *Suppose $c\in\mathbb{R}^{\ell}\langle\langle X\rangle\rangle$ is a series with coefficients that satisfy*

$$|(c,\eta)|\leq KM^{|\eta|}|\eta|!,\quad \forall\eta\in X^{*} \tag{3.2}$$

for some real numbers $K,M>0$. (Here $|z|:=\max_{1\leq i\leq\ell}|z_i|$ when $z\in\mathbb{R}^{\ell}$.) Then there exists real numbers $R,T>0$ such that for each $u\in B_1^m(R)[t_0,t_0+T]$, the series

$$y(t)=F_c[u](t)=\sum_{\eta\in X^{*}}(c,\eta)\,E_{\eta}[u](t,t_0) \tag{3.3}$$

converges absolutely and uniformly on $[t_0,t_0+T]$.

The set of all $c\in\mathbb{R}^{\ell}\langle\langle X\rangle\rangle$ which satisfy a local growth condition of the form (3.2) will be denoted by $\mathbb{R}_{LC}^{\ell}\langle\langle X\rangle\rangle$. For any $\mathfrak{p}\in(1,\infty]$ and on any finite interval $[t_0,t_0+T]$ it can be shown that

[2] The spaces $L_{\mathfrak{p}}^{m}[t_0,t_1]$ and $L_{\mathfrak{q}}^{m}[t_0,t_1]$ are said to be *duals* of each other.

$$\|u\|_1 \leq \|u\|_{\mathfrak{p}}\, T^{1/\mathfrak{q}},$$

when $u \in L_{\mathfrak{p}}^m[t_0, t_0+T]$, and $\mathfrak{p}$ and $\mathfrak{q}$ are conjugate exponents (see Problem 3.2.1). In which case, the following corollary of Theorem 3.1 is immediate and describes other input spaces on which Fliess operators are locally convergent. However, because these spaces are nested for finite T, it is usually most natural to work on the largest space, $L_1^m[t_0, t_0+T]$.

Corollary 3.1 *Suppose $c \in \mathbb{R}_{LC}^{\ell}\langle\langle X\rangle\rangle$ and $\mathfrak{p} \in [1,\infty]$. Then there exists real numbers $R, T > 0$ such that for each $u \in B_{\mathfrak{p}}^m(R)[t_0, t_0+T]$, the series*

$$y(t) = F_c[u](t) = \sum_{\eta \in X^*} (c,\eta)\, E_\eta[u](t,t_0)$$

converges absolutely and uniformly on $[t_0, t_0+T]$.

To prove Theorem 3.1, two upper bounds are needed for iterated integrals over X^*. Both are described in the following lemma. Without loss of generality, it is assumed that $t_0 = 0$, and $E_\eta[u](t,0)$ is abbreviated in this case as $E_\eta[u](t)$.

Lemma 3.1 *For any $u \in L_1^m[0,T]$ and $\eta \in X^*$,*

$$|E_\eta[u](t)| \leq E_\eta[\bar{u}](t), \quad 0 \leq t \leq T,$$

where $\bar{u} \in L_1^m[0,T]$ has components $\bar{u}_j := |u_j|$, $j = 1, 2, \ldots, m$. Furthermore, for any integers $r_j \geq 0$ it follows that

$$\left| E_{x_0^{r_0} \sqcup\!\sqcup\, x_1^{r_1} \sqcup\!\sqcup \cdots \sqcup\!\sqcup\, x_m^{r_m}}[u](t) \right| \leq \prod_{j=0}^{m} \frac{U_j^{r_j}(t)}{r_j!}, \quad 0 \leq t \leq T,$$

where $U_j(t) := \int_0^t |u_j(\tau)|\, d\tau$.[3] *In particular, if on $[0,T]$ it is assumed that* $\max\{\|u\|_1, T\} \leq R$ *then*

$$\left| E_{x_0^{r_0} \sqcup\!\sqcup\, x_1^{r_1} \sqcup\!\sqcup \cdots \sqcup\!\sqcup\, x_m^{r_m}}[u](t) \right| \leq \frac{R^k}{r_0!\, r_1! \cdots r_m!}, \quad 0 \leq t \leq T,$$

where $k = \sum_j r_j$.

[3] For notational convenience, occasionally F_p will be denoted as in the previous chapter by E_p when $p \in \mathbb{R}\langle X\rangle$.

Proof: The first inequality is trivial in the case of the empty word. Suppose it holds for all words up to length $k \geq 0$. Then for any $x_j \in X$ and $\eta \in X^k$ observe that

$$\begin{aligned}\left|E_{x_j\eta}[u](t)\right| &\leq \int_0^t |u_j(\tau)|\,|E_\eta[u](\tau)|\;d\tau \\ &\leq \int_0^t \bar{u}_j(\tau)E_\eta[\bar{u}](\tau)\,d\tau \\ &= E_{x_j\eta}[\bar{u}](t).\end{aligned}$$

Hence, the claim holds for all $\eta \in X^*$.

Concerning the second inequality, note that

$$\left|E_{x_0^{r_0} \sqcup\!\sqcup\, x_1^{r_1} \sqcup\!\sqcup \cdots \sqcup\!\sqcup\, x_m^{r_m}}[u](t)\right| = \prod_{j=0}^{m}\left|E_{x_j^{r_j}}[u](t)\right|$$

(see Lemma 2.3). Thus, it is sufficient to show that

$$\left|E_{x_j^{r_j}}[u](t)\right| \leq \frac{U_j^{r_j}(t)}{r_j!}. \tag{3.4}$$

This claim is clearly true when $r_j = 0$. If it holds up to some fixed integer $r_j \geq 0$ then

$$\begin{aligned}\left|E_{x_j^{r_j+1}}[u](t)\right| &\leq \int_0^t |u_j(\tau)|\left|E_{x_j^{r_j}}[u](\tau)\right|\,d\tau \\ &\leq \int_0^t |u_j(\tau)|\,\frac{U_j^{r_j}(\tau)}{r_j!}\,d\tau \\ &= \frac{U_j^{r_j+1}(t)}{(r_j+1)!}.\end{aligned}$$

Thus, the inequality (3.4) holds for all $r_j \geq 0$, and the lemma is proved. ■

Proof of Theorem 3.1: Suppose the coefficients of c satisfy the local growth condition (3.2). Fix some $T > 0$. Pick any $u \in L_1^m[0,T]$ and let $R = \max\{\|u\|_1, \|u_0\|_1\} = \max\{\|u\|_1, T\}$. Observe that with the help of identity (2.6)

$$\sum_{\eta\in X^*} |(c,\eta)E_\eta[u](t)|$$

$$
\begin{aligned}
&\leq \sum_{k=0}^{\infty} \sum_{\eta \in X^k} |(c,\eta)|\, E_\eta[\bar{u}](t) \\
&\leq \sum_{k=0}^{\infty} KM^k k! \sum_{\substack{r_0,r_1,\ldots,r_m \geq 0 \\ r_0+r_1+\cdots+r_m=k}} E_{x_0^{r_0} \,\sqcup\!\sqcup\, x_1^{r_1} \,\sqcup\!\sqcup\, \cdots \,\sqcup\!\sqcup\, x_m^{r_m}}[\bar{u}](t) \\
&\leq \sum_{k=0}^{\infty} KM^k k! \sum_{\substack{r_0,r_1,\ldots,r_m \geq 0 \\ r_0+r_1+\cdots+r_m=k}} \frac{R^k}{r_0! r_1! \cdots r_m!} \\
&= \sum_{k=0}^{\infty} K(MR)^k \sum_{\substack{r_0,r_1,\ldots,r_m \geq 0 \\ r_0+r_1+\cdots+r_m=k}} \frac{k!}{r_0! r_1! \cdots r_m!} \\
&= \sum_{k=0}^{\infty} K(MR(m+1))^k. \qquad (3.5)
\end{aligned}
$$

Therefore, if $R < 1/M(m+1)$, i.e., if

$$
\max\{\|u\|_1, T\} < \frac{1}{M(m+1)}, \qquad (3.6)
$$

then the series (3.3) converges absolutely and uniformly on $[0,T]$ for each $u \in B_1(R)[0,T]$. ∎

The above proof demonstrates in conjunction with Corollary 3.1 that if $c \in \mathbb{R}^{\ell}_{LC}\langle\langle X \rangle\rangle$ then the series (3.3) defines a Fliess operator from $B^m_{\mathfrak{p}}(R)[t_0,\, t_0+T]$ to a bounded subset of $C[t_0,\, t_0+T]$ for every $\mathfrak{p} \in [1,\infty]$, provided that R and T are sufficiently small. The following theorem provides an even more precise description of F_c.

Theorem 3.2 *Suppose $c \in \mathbb{R}^{\ell}_{LC}\langle\langle X \rangle\rangle$ with growth constants $K, M > 0$. Select any pair of conjugate exponents $\mathfrak{p}, \mathfrak{q} \in [1,\infty]$ and real numbers $R, T > 0$ such that $R_1 := \max\{RT^{1/\mathfrak{q}}, T\} < 1/M(m+1)$. (Let $T^{1/\mathfrak{q}} = 1$ when $\mathfrak{p} = 1$.) Then*

$$
F_c : B^m_{\mathfrak{p}}(R)[t_0,\, t_0+T] \to B^{\ell}_{\mathfrak{q}}(S)[t_0, t_0+T],
$$

where $S = KT^{1/\mathfrak{q}}/(1 - MR_1(m+1))$.

Proof: For any $u \in B^m_{\mathfrak{p}}(R)[t_0, t_0+T]$ observe that

$$
\|u\|_1 \leq \|u\|_{\mathfrak{p}}\, T^{1/\mathfrak{q}} \leq RT^{1/\mathfrak{q}}.
$$

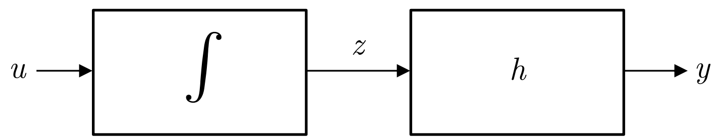


Fig. 3.3. Wiener system in Example 3.3.

Therefore, under the stated assumptions,

$$\begin{aligned}\max\{\|u\|_1, T\} &\leq \max\{RT^{1/\mathfrak{q}}, T\} = R_1 \\ &< \frac{1}{M(m+1)}.\end{aligned}$$

In which case, (3.5) produces the upper bound

$$|y(t)| \leq \frac{K}{1 - MR_1(m+1)}, \quad t_0 \leq t \leq t_0 + T.$$

Thus, $\|y\|_{\mathfrak{q}} \leq KT^{1/\mathfrak{q}}/(1 - MR_1(m+1))$, or equivalently, $y \in B_{\mathfrak{q}}^{\ell}(S)[t_0, t_0 + T]$. ■

Example 3.3 A *Wiener system* is an input-output system consisting of a linear operator whose output is filtered by a function h. This class of systems arises naturally, for example, in control systems where the control law is realized by a linear state space model, and the actuators, which are driven by the controller, exhibit saturation or some other type of static nonlinearity. Certain classes of neural networks also exhibit this type of structure. As an example, consider the single-input, single-output Wiener system shown in Figure 3.3, where the dynamical system is simply an integrator initialized so that $z(0) = 0$ and $h(z) = 1/(1 - z)$. Direct substitution for z in h gives

$$\begin{aligned}y(t) = h(z(t)) &= \sum_{k=0}^{\infty} (z(t))^k \\ &= \sum_{k=0}^{\infty} \left(\int_0^t u(\tau)\, d\tau\right)^k = \sum_{k=0}^{\infty} (E_{x_1}[u](t))^k \\ &= \sum_{k=0}^{\infty} E_{x_1^{\sqcup\!\sqcup\, k}}[u](t) = \sum_{k=0}^{\infty} E_{k!\, x_1^k}[u](t)\end{aligned}$$

$$= \sum_{k=0}^{\infty} k!\, E_{x_1^k}[u](t)$$
$$= F_c[u](t)$$

for all $t \geq 0$ provided that $|z(t)| < 1$ (see Problem 2.4.5(a)). Note here that the generating series $c = \sum_{k\geq 0} k!\, x_1^k$ is locally convergent with growth constants $K = M = 1$. Thus, for any $\mathfrak{p} \in [1, \infty]$ there must exist $R, T > 0$ such that if $\|u\|_{\mathfrak{p}} \leq R$ then $F_c[u]$ is well defined on $[0, T]$. For example, when $\mathfrak{p} = 1$ it follows from (3.6) that a sufficient condition for convergence is

$$\max\{R, T\} < \frac{1}{M(m+1)} = \frac{1}{2}. \tag{3.7}$$

However, this bound on R and T is conservative because not every coefficient of c is growing at the maximum rate $KM^{|\eta|}\,|\eta|!$. Observe that the de facto alphabet in this case has only a single letter, x_1, so one should really set $m+1 = 1$ in (3.7). For an even less conservative bound select a fixed $T' > 0$ and consider all inputs satisfying $\left\|u_{[0,T']}\right\|_1 < 1$. It follows then that

$$\begin{aligned} |F_c[u](t)| &= \left| \sum_{k=0}^{\infty} \left(\int_0^t u(\tau)\, d\tau \right)^k \right| \\ &\leq \sum_{k=0}^{\infty} \left(\int_0^t |u(\tau)|\, d\tau \right)^k \\ &\leq \sum_{k=0}^{\infty} \left\|u_{[0,T']}\right\|_1^k \\ &= \frac{1}{1 - \left\|u_{[0,T']}\right\|_1}. \end{aligned}$$

Thus, F_c is well defined on $B_1(R)[0, T']$ for *any* finite $T' > 0$ and $R < 1$. But this more detailed type of analysis is often not possible when c has a complex structure. Under such conditions, a simple condition like (3.6) is useful and easy to estimate. □

Example 3.4 An important observation is that the set of locally convergent formal power series, $\mathbb{R}_{LC}\langle\langle X\rangle\rangle$, is not a *closed* subset of $\mathbb{R}\langle\langle X\rangle\rangle$

in the ultrametric topology. For example, let $X = \{x_0, x_1\}$ and consider the sequence of polynomials

$$c_i = x_1 + (2!)^2\, x_1^2 + (3!)^2\, x_1^3 + \cdots + (i!)^2\, x_1^i, \quad i \geq 1.$$

Clearly, each polynomial c_i is locally convergent, but the series $c = \lim_{i\to\infty} c_i$ is not. Furthermore, each Fliess operator F_{c_i} is well defined on some closed ball of input functions, but the present theory does not guarantee that the operator F_c is well defined in any sense. □

Now that conditions have been established under which F_c is well defined, several fundamental properties of this class of operators are considered: absolute continuity and differentiability of the output function, uniqueness of the generating series, and preservation of analyticity from input to output. Recall that in general differentiability of a function at a point implies continuity at that same point, but not conversely. The usual counterexample of the latter is the absolute value function $f(z) = |z|$. It is continuous at $z = 0$ but not differentiable there. A more dramatic example is the function defined by the series

$$f(z) = \sum_{n=0}^{\infty} \frac{1}{2^n} \cos(3^n z).$$

Weierstrass demonstrated in 1872 that this function is continuous at every point but *nowhere* differentiable. To more precisely describe differentiability properties, the following stronger notion of continuity is useful.

Definition 3.4 *Let J be a compact interval of $\mathbb{R}$.*[4] *A function $f : \mathbb{R} \to \mathbb{R}$ is called* ***absolutely continuous on J*** *if for every $\epsilon > 0$ there exists a $\delta > 0$ such that whenever $J_i = [a_i, b_i]$ are nonoverlapping subintervals of J with $\sum_{i=1}^n |b_i - a_i| \leq \delta$, it follows that $\sum_{i=1}^n |f(b_i) - f(a_i)| \leq \epsilon$. If I is an arbitrary interval of $\mathbb{R}$ then f is said to be* ***absolutely continuous on I*** *if it is absolutely continuous on every compact subinterval of I.*

It is easily shown that absolute continuity implies continuity in the usual sense (see Problem 3.2.3), but more importantly, consider the following theorem from real analysis.

[4] This is equivalent to requiring J to be closed and bounded.

Theorem 3.3 *If $f : \mathbb{R} \to \mathbb{R}$ is absolutely continuous on the interval I, then it is differentiable a.e. on I.*

The proof of this theorem will not be pursued here, but the result will be utilized momentarily to prove the next theorem concerning differentiability of the output function of a Fliess operator. It is noted in passing, however, that it can be shown that a function f is absolutely continuous on $J = [a, b]$ if and only if there exists a function $g \in L_1[a, b]$ such that $f(z) = f(a) + \int_a^z g(\tau)\, d\tau$ everywhere on J. Readers familiar with measure theory will recognize this result as a special case of the Radon-Nikodym Theorem. Thus, g, which is equivalent to df/dt a.e., is usually called the Radon-Nikodym derivative. On the other hand, if g happens to be continuous, then all this analysis reduces to the Fundamental Theorem of Integral Calculus, namely, $g = df/dt$ everywhere on J. Now consider the differentiability of the output function of a Fliess operator.

Theorem 3.4 *If $c \in \mathbb{R}^{\ell}_{LC}\langle\langle X \rangle\rangle$ and $u \in B^m_{\mathfrak{p}}(R)[t_0, t_0 + T]$, then $y = F_c[u]$ is differentiable a.e. on $[t_0, t_0 + T]$ provided $R, T > 0$ are sufficiently small. In particular,*

$$\frac{d}{dt} F_c[u] = \sum_{i=0}^{m} u_i\, F_{x_i^{-1}(c)}[u]. \tag{3.8}$$

Proof: In light of Theorem 3.3, it is sufficient to show that $y = F_c[u]$ is absolutely continuous. This is accomplished by first showing that $E_\eta[u]$ is absolutely continuous on $[0, T]$ for any $\eta \in X^*$ and $u \in B_1(R)[0, T]$ with R, T satisfying (3.6). For any $J_i = [a_i, b_i]$ it follows from Lemma 3.1 that

$$|E_\eta[u](b_i, a_i)| \leq \frac{R_i^{|\eta|}}{|\eta|!}, \quad \eta \in X^*,$$

where $R_i := \max\{\|u_{[a_i,b_i]}\|_1, |b_i - a_i|\}$. In addition, R_i can be made arbitrarily small by selecting $|b_i - a_i|$ sufficiently small. So fix $\eta \in X^k$ and pick an $\epsilon > 0$. Consider a sequence

$$0 \leq a_1 < b_1 < a_2 < \cdots < b_n \leq T$$

such that

$$R_i \leq \sqrt[k]{\frac{k!}{n}\epsilon}, \quad i = 1, 2, \ldots, n.$$

This defines implicitly a $\delta > 0$ such that if $\sum_{i=1}^n |b_i - a_i| \leq \delta$ then

$$\begin{aligned}\sum_{i=1}^n |E_\eta[u](b_i) - E_\eta[u](a_i)| &= \sum_{i=1}^n |E_\eta[u](b_i, a_i)| \\ &\leq \sum_{i=1}^n \frac{R_i^k}{k!} \\ &\leq \epsilon.\end{aligned}$$

Therefore, the absolute continuity of $E_\eta[u]$ is established. To show that $F_c[u]$ is absolutely continuous, again select an $\epsilon > 0$ and choose an integer $N > 0$ such that

$$\sum_{k=N}^{\infty} K(MR(m+1))^k \leq \frac{\epsilon}{2}.$$

This is always possible since $M, R > 0$ satisfy $MR(m+1) < 1$. Now because each $E_\eta[u]$ is absolutely continuous, there exists a $\delta > 0$ such that

$$\sum_{i=1}^n \sum_{k=0}^{N-1} \sum_{\eta \in X^k} |(c,\eta)|\, |E_\eta[u](b_i, a_i)| \leq \frac{\epsilon}{2}$$

when $\sum_{i=1}^n |b_i - a_i| \leq \delta$. Hence, for this choice of δ,

$$\begin{aligned}\sum_{i=1}^n |F_c[u](b_i) - F_c[u](a_i)| &\leq \sum_{i=1}^n \sum_{k=0}^{\infty} \sum_{\eta \in X^k} |(c,\eta)|\, |E_\eta[u](b_i, a_i)| \\ &= \sum_{i=1}^n \sum_{k=0}^{N-1} \sum_{\eta \in X^k} |(c,\eta)|\, |E_\eta[u](b_i, a_i)| + \\ &\quad \sum_{i=1}^n \sum_{k=N}^{\infty} \sum_{\eta \in X^k} |(c,\eta)|\, |E_\eta[u](b_i, a_i)| \\ &\leq \frac{\epsilon}{2} + \sum_{k=N}^{\infty} K(MR(m+1))^k \\ &\leq \epsilon,\end{aligned}$$

and the differentiability claim is proved for the $\mathfrak{p} = 1$ case. But the claim in fact holds for any $\mathfrak{p} \in [1, \infty]$ since the $L_\mathfrak{p}$ spaces under consideration are nested.

To verify the formula for $\frac{d}{dt}F_c[u]$ observe that

$$\begin{aligned} F_c[u](t) &= \sum_{\eta\in X^*} (c,\eta)E_\eta[u](t) \\ &= \sum_{i=0}^m \sum_{\eta\in X^*} (c,x_i\eta)E_{x_i\eta}[u](t) \\ &= \sum_{i=0}^m \sum_{\eta\in X^*} (x_i^{-1}(c),\eta)\int_0^t u_i(\tau)E_\eta[u](\tau)\,d\tau. \end{aligned}$$

Hence,

$$\begin{aligned} &\frac{F_c[u](t+\Delta t)-F_c[u](t)}{\Delta t} \\ &\qquad = \sum_{i=0}^m \sum_{\eta\in X^*} (x_i^{-1}(c),\eta)\frac{1}{\Delta t}\int_t^{t+\Delta t} u_i(\tau)E_\eta[u](\tau)\,d\tau. \end{aligned}$$

Again $|E_\eta[u](\tau)| \leq R^{|\eta|}/\,|\eta|!$ for all $\tau \in [0,T]$ and $\eta \in X^*$ when $u \in B_1(R)[0,T]$. So the series above will converge when Δt and R are small enough. Taking the limit $\Delta t \to 0$ gives almost everywhere that

$$\begin{aligned} \frac{d}{dt}F_c[u](t) &= \sum_{i=0}^m \sum_{\eta\in X^*} (x_i^{-1}(c),\eta)\,u_i(t)E_\eta[u](t) \\ &= \sum_{i=0}^m u_i(t)\,F_{x_i^{-1}(c)}[u](t). \end{aligned}$$

■

If it is assumed that u has k continuous derivatives on $[t_0, t_0 + T]$, i.e., $u \in C^k[t_0, t_0+T]$, then it can be shown that $y = F_c[u] \in C^{k+1}[t_0, t_0+T]$, $k \geq 0$ (including the smooth case, $k = \infty$.). This is useful for computing higher-order derivatives that appear in the context of differential equations as well as in the following theorem addressing the uniqueness of the generating series.

Theorem 3.5 *Suppose* $c, d \in \mathbb{R}^\ell_{LC}\langle\langle X\rangle\rangle$. *If* $F_c = F_d$ *on* $B^m_\infty(R)[t_0,\, t_0+T]$ *for some real numbers* $R, T > 0$, *then* $c = d$.

Proof: In light of the fact that $F_c - F_d = F_{c-d}$, it is sufficient to show that if $F_c = 0$ on some $B^m_\infty(R)[t_0,\, t_0+T]$ then $c = 0$. The approach is

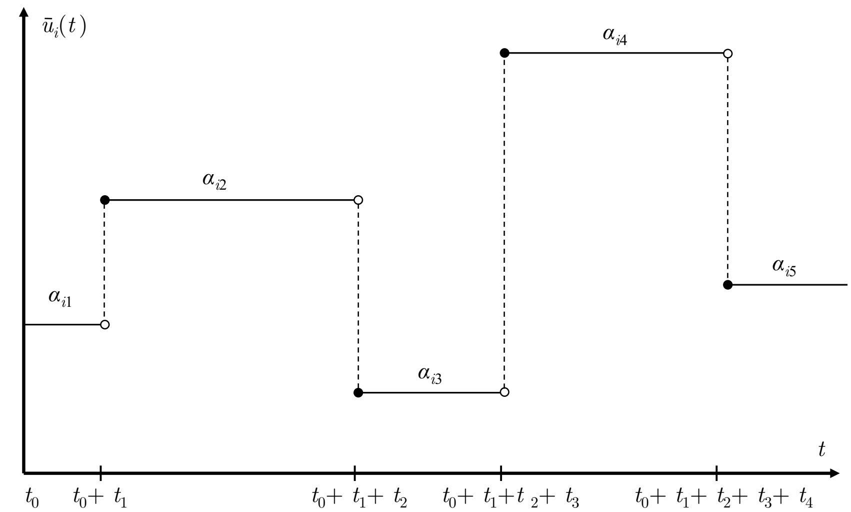


Fig. 3.4. The i-th component of the input function $\bar{u}$ utilized in the proof of Theorem 3.5 ($i \neq 0$).

based on the following simple observation. Suppose the step function input $\bar{u} = \alpha\mathbb{U}$ is applied to an integrator multiplied by a constant (c, x_1), i.e.,

$$F_c[\bar{u}](t) = (c, x_1)\int_0^t \alpha\mathbb{U}(\tau)\,d\tau = \begin{cases} \alpha t(c, x_1) & : \quad t \geq 0^+ \\ 0 & : \quad \text{otherwise}, \end{cases}$$

where $c = (c, x_1)x_1$. Then clearly the only nonzero coefficient of the generating series c can be extracted from the output $y = F_c[\bar{u}]$ by computing $\partial^2 y(t)/\partial\alpha\partial t = (c, x_1)$ (see Problem 3.2.6 for another simple example). The idea is to extend this approach to the case where c is arbitrary. Consider applying a piecewise constant input in $B_\infty^m(R)[t_0,\, t_0 + T]$ defined by

$$\bar{u}(t) = \alpha_j \leq R, \quad t \in \left[\sum_{l=0}^{j-1} t_l, \sum_{l=0}^{j} t_l\right)$$

with $\alpha_j = [\alpha_{1j}\ \alpha_{2j} \cdots \alpha_{mj}]^T \in \mathbb{R}^m$ and $t_l > 0$, $j, l = 1, 2, \ldots, k$; and $\sum_{l=1}^{k} t_l < T$ (see Figure 3.4). Since $\bar{u}_0 = 1$, define $\alpha_{0j} = 1$. The claim to be verified inductively is that for any $k \geq 1$

$$\left.\frac{\partial^k}{\partial t_1 \partial t_2 \cdots \partial t_k} F_c[\bar{u}](t_0 + t_1 + \cdots + t_k)\right|_{\substack{t_j = 0^+,\\ j=1,2,\ldots,k}} = \sum_{\xi \in X^k} \alpha_{\xi k}\,(c, \xi), \tag{3.9}$$

where

$$\alpha_{\xi k} := \alpha_{i_k k} \alpha_{i_{k-1} k-1} \cdots \alpha_{i_1 1}$$

when $\xi = x_{i_k} x_{i_{k-1}} \cdots x_{i_1} \in X^k$.[5] Observe that when $k = 1$

$$\begin{aligned}
\left.\frac{\partial}{\partial t_1} F_c[\bar{u}](t_0 + t_1)\right|_{t_1 = 0^+} &= \left.\sum_{i=0}^{m} \bar{u}_i(t_0 + t_1) F_{x_i^{-1}(c)}[\bar{u}](t_0 + t_1)\right|_{t_1 = 0^+} \\
&= \sum_{\substack{x_i \in X\\ \xi \in X^*}} \bar{u}_i(t_0^+)\,(c, x_i \xi)\, E_\xi[\bar{u}](t_0) \\
&= \sum_{x_i \in X} \alpha_{x_i 1}\,(c, x_i).
\end{aligned}$$

Note that in the last step, the identity

$$E_\xi[\bar{u}](t_0) = \begin{cases} 1 & : \ \xi = \emptyset \\ 0 & : \ \xi \neq \emptyset \end{cases}$$

was employed. Now assume that the identity (3.9) holds up to some fixed $k \geq 1$. Then

$$\begin{aligned}
&\left.\frac{\partial^{k+1}}{\partial t_1 \partial t_2 \cdots \partial t_{k+1}} F_c[\bar{u}](t_0 + t_1 + \cdots + t_{k+1})\right|_{\substack{t_j = 0^+,\\ j=1,2,\ldots,k+1}} \\
&= \left.\frac{\partial^k}{\partial t_1 \partial t_2 \cdots \partial t_k} \left[\left.\frac{\partial}{\partial t_{k+1}} F_c[\bar{u}](t_0 + t_1 + \cdots + t_{k+1})\right|_{t_{k+1} = 0^+} \right]\right|_{\substack{t_j = 0^+,\\ j=1,2,\ldots,k}} \\
&= \frac{\partial^k}{\partial t_1 \partial t_2 \cdots \partial t_k} \left[\sum_{i_{k+1}=0}^{m} \bar{u}_{i_{k+1}}(t_0 + t_1 + \cdots + t_{k+1}) \cdot \right. \\
&\qquad \left.\left. \left. F_{x_{i_{k+1}}^{-1}(c)}[\bar{u}](t_0 + t_1 + \cdots + t_{k+1})\right|_{t_{k+1} = 0^+} \right]\right|_{\substack{t_j = 0^+,\\ j=1,2,\ldots,k}} \\
&= \sum_{x_{i_{k+1}} \in X} \alpha_{x_{i_{k+1}} k+1} \frac{\partial^k}{\partial t_1 \partial t_2 \cdots \partial t_k} \cdot
\end{aligned}$$

---

[5] Since $\bar{u}$ is smooth almost everywhere on $[t_0, t_0 + T]$, the same is true of $F_c[\bar{u}]$.

$$F_{x_{i_{k+1}}^{-1}(c)}[\bar{u}](t_0+t_1+\cdots+t_k)\Big|_{\substack{t_j=0^+,\\ j=1,2,\ldots,k}}.$$

The key fact used above is that $\bar{u}_{i_{k+1}}(t_0+t_1+\cdots+t_k^+) = \alpha_{i_{k+1}k+1}$. Now use the induction hypothesis by applying (3.9):

$$\begin{aligned}
&\frac{\partial^{k+1}}{\partial t_1 \partial t_2 \cdots \partial t_{k+1}} F_c[\bar{u}](t_0+t_1+\cdots+t_{k+1})\Big|_{\substack{t_j=0^+,\\ j=1,2,\ldots,k+1}} \\
&\quad = \sum_{x_{i_{k+1}}\in X} \alpha_{x_{i_{k+1}}k+1}\left[\sum_{\xi\in X^k} \alpha_{\xi k}\,(x_{i_{k+1}}^{-1}(c),\xi)\right] \\
&\quad = \sum_{\xi\in X^{k+1}} \alpha_{\xi k+1}\,(c,\xi).
\end{aligned}$$

Hence, by induction (3.9) holds for all $k \geq 1$. Now by assumption $F_c[\bar{u}] = 0$ on $[t_0, t_0+T]$. Therefore,

$$\sum_{\xi\in X^k} \alpha_{\xi k}\,(c,\xi) = 0, \quad k \geq 1. \tag{3.10}$$

Furthermore, observe that the left-hand side of the above expression is a polynomial function of the input parameters α_{ij}. Thus, one can compute partial derivatives of this expression with respect to α_{ij}. Let $l_1, l_2, \ldots, l_k$ be any k-tuple where $l_r \in \{0, 1, \ldots, m\}$. Then

$$\begin{aligned}
&\frac{\partial^k}{\partial \alpha_{l_k k}\partial \alpha_{l_{k-1}k-1}\cdots \partial \alpha_{l_1 1}}\, \alpha_{i_k k}\alpha_{i_{k-1}k-1}\cdots \alpha_{i_1 1} \\
&\qquad = \begin{cases} 1 & : \quad l_j = i_j, \quad j = 1, 2, \ldots, k \\ 0 & : \quad \text{otherwise.} \end{cases}
\end{aligned}$$

Taking such partial derivatives of both sides of equation (3.10) yields

$$\begin{aligned}
&\frac{\partial^k}{\partial \alpha_{l_k k}\partial \alpha_{l_{k-1}k-1}\cdots \partial \alpha_{l_1 1}} \cdot \\
&\qquad \sum_{x_{i_k}x_{i_{k-1}}\cdots x_{i_1}\in X^k} \alpha_{i_k k}\alpha_{i_{k-1}k-1}\cdots \alpha_{i_1 1}\,(c, x_{i_k}x_{i_{k-1}}\cdots x_{i_1}) \\
&= (c, x_{l_k}x_{l_{k-1}}\cdots x_{l_1}) \\
&= 0.
\end{aligned}$$

That is, for any $\eta = x_{l_k}x_{l_{k-1}}\cdots x_{l_1} \in X^k$, $k \geq 1$, it follows that $(c,\eta) = 0$. One caveat in this argument, however, is the case where one or more $i_j = 0$. Recall that α_{0j} has been fixed to one, so it is not really a free variable in the setup. A straightforward check shows that the above argument will give the desired result when all partials of the form $\partial/\partial\alpha_{0,j}$ are simply omitted. Finally, setting $u = 0$ gives $(c,\emptyset) = 0$. Thus, $c = 0$ as desired. ■

Next, the main analyticity theorem is given. The basic claim is that real analytic inputs produce real analytic outputs. The proof is accomplished by extending the setup to the complex field, using tools from the theory of complex variables, and then restricting the results back to the real field.

Theorem 3.6 *If $c \in \mathbb{R}^{\ell}_{LC}\langle\langle X\rangle\rangle$ and $u \in B^m_{\mathfrak{p}}(R)[t_0, t_0+T]$ is real analytic on $[t_0, t_0+T]$, then $y = F_c[u]$ is real analytic on $[t_0, t_0+T]$ provided that $R, T > 0$ are sufficiently small.*

Proof: Assume $t_0 = 0$ and $R, T > 0$ are selected so that a given $u \in B^m_{\mathfrak{p}}(R)[0,T]$ is real analytic on $[0,T]$ and $y = F_c[u]$ is well defined. Let $\tilde{u}$ be the complex extension of u which is analytic on a neighborhood $W_{\mathbb{C}}$ of $[0,T]$ in the complex plane. Without loss of generality, one can assume that $W_{\mathbb{C}}$ is simply connected, that the closure $\overline{W}_{\mathbb{C}}$ of $W_{\mathbb{C}}$ is compact, and that $\tilde{u}$ is analytic on $\overline{W}_{\mathbb{C}}$. Thus, for any fixed path in $\overline{W}_{\mathbb{C}}$ there exists some $\tilde{R} > 0$ such that $\|\tilde{u}\|_{\mathfrak{p}} \leq \tilde{R}$. Here the norm is extended in a natural way using the moduli of the components of $\tilde{u}(w)$ as w follows the given path. Now for such a $\tilde{u}$ define the iterated integrals $E_\eta[\tilde{u}] : W_{\mathbb{C}} \to \mathbb{C}$ by

$$E_{x_i\eta}[\tilde{u}](w) = \int_0^w \tilde{u}_i(\zeta)E_\eta[\tilde{u}](\zeta)\,d\zeta,\ \ x_i \in X,\ \ \eta \in X^*,$$

where $E_\emptyset = 1$ and $\tilde{u}_0 = 1$. By induction, the integrand is analytic on $W_{\mathbb{C}}$, so the value of the integral is independent of the path (chosen inside $W_{\mathbb{C}}$), and the resulting function is analytic on $W_{\mathbb{C}}$ as well. As in the proof of Theorem 3.1, it can be shown from the assumed growth condition that the series

$$\tilde{y}(w) = \sum_{\eta\in X^*} (c,\eta)E_\eta[\tilde{u}](w)$$

converges uniformly on $W_{\mathbb{C}}$. Since each $E_\eta[\tilde{u}](w)$ is analytic on $W_{\mathbb{C}}$, it follows that $\tilde{y}(w)$ is also analytic on $W_{\mathbb{C}}$ (see Problem 3.2.7). Clearly

$\tilde{y}(w)$ is an analytic complex extension of the real-valued function

$$y(t) = \sum_{\eta \in X^*} (c,\eta) E_\eta[u](t).$$

In which case, the restriction of $\tilde{y}$ to $[0,T]$, namely y, is real analytic on $[0,T]$. ■

It is important to point out what is *not* being claimed in the theorem above. Namely, that if u can be represented by a *single* power series over $[0,T]$ then so can $y = F_c[u]$. This claim is generally false as shown in the next example.

Example 3.5 Reconsider the Wiener system in Example 3.3 with $\mathfrak{p} = 1$. Take the input to be the entire function $u(t) = -\sin \pi t$. Then clearly $\left\|u_{[0,T]}\right\|_1 \leq T$ so that $F_c[u]$ converges at least on $[0,T]$ for any $T < 1$. However, observe more specifically that

$$\begin{aligned} z(t) = \int_0^t u(\tau)\, d\tau &= \frac{1}{\pi}\, \cos(\pi t) - \frac{1}{\pi} \\ &= -\frac{2}{\pi} \sin^2\left(\frac{\pi}{2} t\right), \end{aligned}$$

and hence,

$$y(t) = \frac{1}{1 - z(t)} = \frac{\frac{\pi}{2}}{\frac{\pi}{2} + \sin^2\left(\frac{\pi}{2} t\right)}$$

for any $t \in [0,\infty)$. So while u has a Taylor series at $t = 0$ which converges everywhere on $[0,\infty)$, it will be shown that y does not. Define the function

$$f(w) = \frac{\pi}{2} + \sin^2\left(\frac{\pi}{2} w\right),$$

where $w = a + ib$. It is easy to verify, using the identity

$$\sin\theta = \frac{\mathrm{e}^{i\theta} - \mathrm{e}^{-i\theta}}{2i},$$

that

$$\begin{aligned} \mathrm{Re}(f(w)) &= \frac{1}{2}(1 + \pi - \cos(\pi a)\cosh(\pi b)) \\ \mathrm{Im}(f(w)) &= \frac{1}{2}\sin(\pi a)\sinh(\pi b). \end{aligned}$$

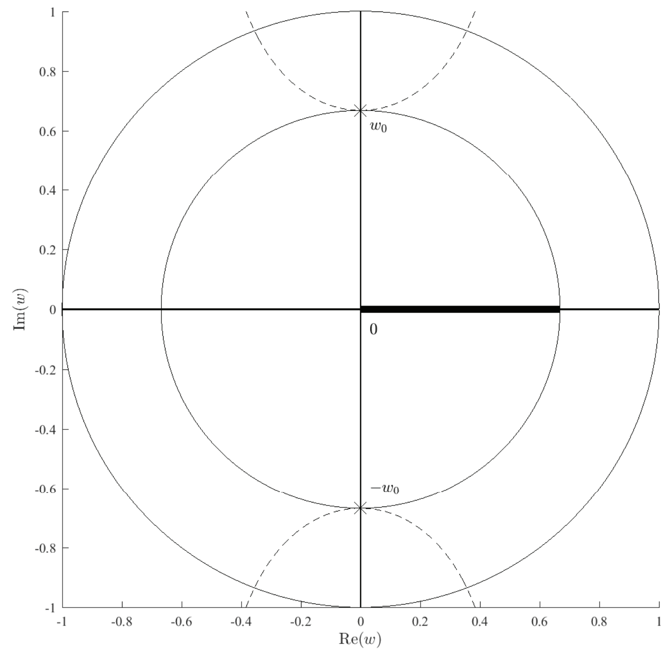


Fig. 3.5. Values of w for which $\mathrm{Re}(f(w)) = 0$ (dashed lines), $\mathrm{Im}(f(w)) = 0$ (real and imaginary axes), and the region of convergence for the Taylor series of y at $t = 0$ in Example 3.5 (inside inner circle).

In which case, $\mathrm{Im}(f(w)) = 0$ within the unit circle only along both the real and imaginary axes. The values of w for which $\mathrm{Re}(f(w)) = 0$ are shown in Figure 3.5. Thus, inside the unit circle, $f(w) = 0$ only when $w = \pm i w_0$, where

$$w_0 = \frac{1}{\pi}\cosh^{-1}(\pi + 1) = 0.6682353705.$$

So while y is real analytic at every point within $[0, \infty)$, the series representation of y at $t = 0$ only converges on $[0, w_0)$. Any shift of this representation to another point within $[0, \infty)$ has similar limitations. Thus, at least two series are needed to represent y on $[0, \infty)$. With the help of symbol manipulation software, the first ten terms of the Taylor series at $w = 0$ are found to be

$$\begin{aligned} y(w) =& 1 - \frac{1}{2}\,\pi\, w^2 + \frac{1}{24}\,\pi^2\,(\pi + 6)\, w^4 - \frac{1}{720}\,\pi^3\left(30\,\pi + \pi^2 + 90\right) w^6 + \\ & \frac{1}{40320}\,\pi^4\left(126\,\pi^2 + \pi^3 + 1260\,\pi + 2520\right) w^8 - \frac{1}{3628800}\,\pi^5 \end{aligned}$$

$$\left(510\,\pi^3+\pi^4+13230\,\pi^2+75600\,\pi+113400\right)w^{10}+O\left(w^{12}\right)$$
$$=1-1.570796327\,w^2+3.759329296\,w^4-8.359524051\,w^6+$$
$$18.73043265\,w^8-41.94529728\,w^{10}+O\left(w^{12}\right).$$

The square root of the ratio of the magnitudes of the last two nonzero coefficients of this series,

$$\sqrt{18.73043265/41.94529728}=0.668239688,$$

must give an estimate of w_0 since the radius of convergence of such a series, R, and the smallest geometric growth constant, M, are related as $R = 1/M$ (see Section 1.1). It is worth restating the very subtle point that while this series representation of y diverges at $t = w_0$, the function y is perfectly well defined at $t = w_0$, namely $y(w_0) = 0.676214167$ (recall Problem 1.1.1). □

As Theorem 3.2 demonstrated, the input-output system F_c can be viewed as a mapping between to closed balls in the normed linear spaces $L_{\mathfrak{p}}^m[t_0, t_0+T]$ and $L_{\mathfrak{q}}^{\ell}[t_0, t_0+T]$ when its generating series is locally convergent. Thus, it makes sense to consider whether the mapping is continuous as an operator between these spaces. The following theorem answers this question to the affirmative.

Theorem 3.7 *Suppose $c \in \mathbb{R}_{LC}^{\ell}\langle\langle X\rangle\rangle$ and select any pair of conjugate exponents $\mathfrak{p}, \mathfrak{q} \in [0,\infty]$. If the real numbers $R, T > 0$ are sufficiently small, then the operator*

$$F_c : B_{\mathfrak{p}}^m(R)[t_0,\, t_0+T] \to B_{\mathfrak{q}}^{\ell}(S)[t_0, t_0+T]$$

for some $S > 0$ is continuous with respect to the $L_{\mathfrak{p}}$ and $L_{\mathfrak{q}}$ norms. That is, for any $\epsilon > 0$ there exists a $\delta > 0$ such that for any $u, v \in B_{\mathfrak{p}}^m(R)[t_0,\, t_0+T]$ satisfying $\|v-u\|_{\mathfrak{p}} < \delta$ it follows that $\|F_c[v]-F_c[u]\|_{\mathfrak{q}} < \epsilon$.

Proof: It is first proved by induction on the length of the word $\eta \in X^*$ that the mapping

$$E_\eta : B_{\mathfrak{p}}^m(R)[t_0,\, t_0+T] \to B_{\mathfrak{q}}^{\ell}(S)[t_0, t_0+T]$$

has the desired continuity property. The focus is on the case where $\mathfrak{p}, \mathfrak{q} \in (1,\infty)$ (the remaining case is handled similarly and left to the reader). Without loss of generality, assume $t_0 = 0$. The claim is trivial when η is the empty word. If $\eta = x_i$, then

$$\begin{aligned}\|E_{x_i}[v] - E_{x_i}[u]\|_{\mathfrak{q}} &= \left(\int_0^T |E_{x_i}[v](t) - E_{x_i}[u](t)|^{\mathfrak{q}}\,dt\right)^{\frac{1}{\mathfrak{q}}} \\ &\leq \left(\int_0^T \left(\int_0^T |v_i(\tau) - u_i(\tau)|\;d\tau\right)^{\mathfrak{q}} dt\right)^{\frac{1}{\mathfrak{q}}} \\ &= \int_0^T |v_i(\tau) - u_i(\tau)|\;d\tau\, T^{\frac{1}{\mathfrak{q}}} \\ &\leq \|v_i - u_i\|_p\; T^{\frac{2}{\mathfrak{q}}} \\ &\leq \|v - u\|_p\; T^{\frac{2}{\mathfrak{q}}},\end{aligned}$$

where Hölder's inequality has been used in the second to the last step above. Thus, if $\|v - u\|_p < \delta := \epsilon/T^{2/\mathfrak{q}}$, then clearly

$$\|E_{x_i}[v] - E_{x_i}[u]\|_{\mathfrak{q}} < \epsilon.$$

Now suppose the claim holds for all words up to some fixed length $k \geq 0$. Then for any $x_i \in X$ and $\eta \in X^k$ observe

$$\begin{aligned}&\|E_{x_i\eta}[v] - E_{x_i\eta}[u]\|_{\mathfrak{q}} \\ &= \left\|\left(E_{x_i\eta}[v] - \int_0^{\cdot} u_i(\tau)E_\eta[v](\tau)\,d\tau\right) + \right. \\ &\qquad \left.\left(\int_0^{\cdot} u_i(\tau)E_\eta[v](\tau)\,d\tau - E_{x_i\eta}[u]\right)\right\|_{\mathfrak{q}} \\ &\leq \left\|E_{x_i\eta}[v] - \int_0^{\cdot} u_i(\tau)E_\eta[v](\tau)\,d\tau\right\|_{\mathfrak{q}} + \\ &\qquad \left\|\int_0^{\cdot} u_i(\tau)E_\eta[v](\tau)\,d\tau - E_{x_i\eta}[u]\right\|_{\mathfrak{q}} \\ &\leq \left(\int_0^T\left(\int_0^T |v_i(\tau) - u_i(\tau)|\,|E_\eta[v](\tau)|\;d\tau\right)^{\mathfrak{q}} dt\right)^{\frac{1}{\mathfrak{q}}} + \\ &\qquad \left(\int_0^T\left(\int_0^T |u_i(\tau)|\,|E_\eta[v](\tau) - E_\eta[u](\tau)|\;d\tau\right)^{\mathfrak{q}} dt\right)^{\frac{1}{\mathfrak{q}}} \\ &\leq \int_0^T |v_i(\tau) - u_i(\tau)|\,|E_\eta[v](\tau)|\;d\tau\, T^{\frac{1}{\mathfrak{q}}} +\end{aligned}$$

$$\int_0^T |u_i(\tau)|\,|E_\eta[v](\tau) - E_\eta[u](\tau)|\; d\tau\, T^{\frac{1}{\mathfrak{q}}}$$
$$\leq \|v-u\|_{\mathfrak{p}} \left\|E_\eta[v]\right\|_{\mathfrak{q}} T^{\frac{1}{\mathfrak{q}}} + \|u\|_{\mathfrak{p}} \left\|E_\eta[v] - E_\eta[u]\right\|_{\mathfrak{q}}\; T^{\frac{1}{\mathfrak{q}}}.$$

From the induction hypothesis E_η is continuous in the desired sense. Thus, it follows that for any $\epsilon > 0$, there exists a $\delta' > 0$ such that

$$\left\|E_\eta[v]\right\|_{\mathfrak{q}} \leq \left\|E_\eta[u]\right\|_{\mathfrak{q}} + 1$$

and

$$\|u\|_{\mathfrak{p}} \left\|E_\eta[v] - E_\eta[u]\right\|_{\mathfrak{q}} T^{\frac{1}{\mathfrak{q}}} < \epsilon/2$$

for all v in a ball centered at u of radius $\delta' > 0$.[6] In which case, choose

$$\delta = \min\left\{\delta', \frac{\epsilon/2}{(\left\|E_\eta[u]\right\|_{\mathfrak{q}} + 1)T^{\frac{1}{\mathfrak{q}}}}\right\}$$

so that if $\|u-v\|_{\mathfrak{p}} < \delta$, then

$$\left\|E_{x_i\eta}[v] - E_{x_i\eta}[u]\right\|_{\mathfrak{q}} < \epsilon.$$

Hence, by induction, E_η is continuous with respect to the $L_{\mathfrak{p}}$ and $L_{\mathfrak{q}}$ norms for every $\eta \in X^*$.

To show that F_c is also continuous in the desired sense, observe that for any integer $N > 0$

$$\begin{aligned}
\left\|F_c[v] - F_c[u]\right\|_{\mathfrak{q}} &= \left\|\sum_{k=0}^{\infty}\sum_{\eta\in X^k}(c,\eta)(E_\eta[v] - E_\eta[u])\right\|_{\mathfrak{q}} \\
&\leq \left\|\sum_{k=0}^{N-1}\sum_{\eta\in X^k}(c,\eta)(E_\eta[v] - E_\eta[u])\right\|_{\mathfrak{q}} + \\
&\quad \left\|\sum_{k=N}^{\infty}\sum_{\eta\in X^k}(c,\eta)(E_\eta[v] - E_\eta[u])\right\|_{\mathfrak{q}} \\
&\leq \left\|\sum_{k=0}^{N-1}\sum_{\eta\in X^k}(c,\eta)(E_\eta[v] - E_\eta[u])\right\|_{\mathfrak{q}} +
\end{aligned}$$

[6] Of course, δ' must be selected so that this ball is contained inside $B_{\mathfrak{p}}^m(R)[0,T]$. It is also being tacitly assumed that u is not on the boundary of $B_{\mathfrak{p}}^m(R)[0,T]$. Otherwise, this argument needs a few minor adjustments.

$$2\sum_{k=N}^{\infty} K(MR(m+1))^k,$$

where $K, M > 0$ are defined as in the proof of the previous theorem. Clearly, the second term above can be bounded by $\epsilon/2$ by selecting N to be sufficiently large. For this fixed N, it is now possible to bound the first term by $\epsilon/2$ since each E_η in this finite sum is continuous as shown above. This proves the theorem. ∎

3.3 Global Input-Output Properties

In this section, a general condition is described under which a Fliess operator is globally convergent. In addition, a global counterpart for the analyticity of the output function is presented. The first theorem introduces a growth condition for the coefficients of c under which F_c is well defined for every input function from $L_{1,e}^m(t_0)$.

Theorem 3.8 *Suppose $c \in \mathbb{R}^\ell\langle\langle X\rangle\rangle$ is a series with coefficients that satisfy*

$$|(c,\eta)| \leq KM^{|\eta|}(|\eta|!)^s, \quad \forall \eta \in X^* \tag{3.11}$$

for some real numbers $K, M > 0$ and $s \in [0,1)$. Then for any $u \in L_{1,e}^m(t_0)$, the series

$$y(t) = F_c[u](t) = \sum_{\eta\in X^*} (c,\eta)E_\eta[u](t) \tag{3.12}$$

converges absolutely and uniformly on $[t_0, t_0+T]$ for any $T > 0$.

Proof: Without loss of generality assume that $t_0 = 0$. Choose any $T > 0$ and pick any $u \in L_{1,e}^m(0)$. Let

$$R = \max\left\{\left\|u_{[0,T]}\right\|_1, T\right\}.$$

If the coefficients of c satisfy the global growth condition (3.11) then the upperbound (3.5) in the proof of Theorem 3.1 can be strengthened to

$$\sum_{\eta\in X^*} |(c,\eta)E_\eta[u](t)| \leq K\sum_{k=0}^{\infty}\frac{(MR(m+1))^k}{(k!)^{1-s}}.$$

Defining the sequence $a_k := (MR(m+1))^k/(k!)^{1-s}$, it is clear that

$$\lim_{k\to\infty} \frac{a_{k+1}}{a_k} = (MR(m+1)) \lim_{k\to\infty} \frac{1}{(k+1)^{1-s}} = 0.$$

Thus, from the ratio test the series in (3.12) must converge absolutely and uniformly on $[0,T]$. ■

The set of all $c \in \mathbb{R}^\ell\langle\langle X\rangle\rangle$ which satisfy the global growth condition (3.11) will be denoted by $\mathbb{R}^\ell_{GC}\langle\langle X\rangle\rangle$. For any given c, a constant s for which there exists $K, M > 0$ satisfying (3.11) is called a *Gevrey order*. Clearly if $s' > s$ then c will also have Gevrey order s'. The minimum of all Gevrey orders of c is written as s^*. The following corollary is a direct consequence of the fact that the extended spaces $L^m_{\mathfrak{p},e}(t_0)$, $\mathfrak{p} \in [1,\infty]$ are nested.

Corollary 3.2 *If* $c \in \mathbb{R}^\ell_{GC}\langle\langle X\rangle\rangle$ *and* $u \in L^m_{\mathfrak{p},e}(t_0)$ *with* $\mathfrak{p} \in [1,\infty]$*, then the series (3.12) converges on* $[t_0, t_0+T)$ *for any* $T > 0$.

Example 3.6 Let $X = \{x_0, x_1\}$ and consider a linear series

$$c = \sum_{k=0}^{\infty} CA^kB\, x_0^k x_1$$

with $A \in \mathbb{R}^{n\times n}$ nonzero and $B, C^T \in \mathbb{R}^{n\times 1}$. Observe that

$$|(c, x_0^k x_1)| \le \|C\|\; \|A\|^k\; \|B\| = (\|C\|\,\|B\|\,\|A\|^{-1})\,\|A\|^{|x_0^k x_1|},\;\; k \ge 0,$$

where $\|\cdot\|$ denotes any (sub-multiplicative) matrix/vector norm. In which case, $c \in \mathbb{R}_{GC}\langle\langle X\rangle\rangle$ with global growth constants $K = \|C\|\,\|B\| \cdot \|A\|^{-1}$, $M = \|A\|$ and $s^* = 0$. □

Example 3.7 Consider the single-input, single-output Wiener system as shown in Figure 3.6. Observe

$$\begin{aligned} y(t) = \mathrm{e}^{z(t)} &= \sum_{k=0}^{\infty} \frac{(z(t))^k}{k!} \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} E^k_{x_1}[u](t) = \sum_{k=0}^{\infty} E_{\frac{1}{k!}x_1^{\sqcup\!\sqcup\, k}}[u](t) \end{aligned}$$

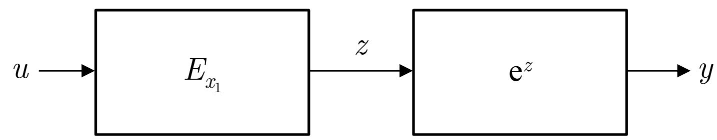


Fig. 3.6. The Wiener system in Example 3.7.

$$= \sum_{k=0}^{\infty} E_{x_1^k}[u](t)$$
$$= F_c[u](t),$$

where $c = \sum_{k\geq 0} x_1^k$. Clearly, $|(c,\eta)| \leq 1$ for all $\eta \in X^*$ with $X = \{x_0, x_1\}$. So c is globally convergent with $K = M = 1$ and $s^* = 0$. □

Example 3.8 This example suggests that the growth condition (3.11) may only be a sufficient condition for global convergence, i.e., it is *not* necessary. Suppose the system in the previous example is cascaded with a copy of itself to produce the new system

$$y(t) = \exp(E_{x_1}[\exp(E_{x_1}[u(t)])].$$

It is not immediately evident that this new system has a Fliess operator representation, this fact will be established in Section 3.6. But observe that for any $u \in L_{1,e}(0)$, the output is well defined for every finite t. Thus, any corresponding Fliess operator would have to be globally convergent. However, consider the special case where $u(t) = 1$ on any finite interval $[0,T]$ so that

$$y(t) = \mathrm{e}^{\mathrm{e}^t - 1} = \sum_{n=0}^{\infty} B_n \frac{t^n}{n!}, \quad t \geq 0$$

as shown in Figure 3.7, where B_n, $n \geq 0$ is the integer sequence known as the *Bell numbers.* The first few Bell numbers are: 1, 2, 5, 15, 52, 203, 877, 4140. Their asymptotic behavior is known to simultaneously satisfy the following three limits:

$$\lim_{n\to\infty} \frac{B_n}{M^n} = \infty, \quad \forall M > 0$$
$$\lim_{n\to\infty} \frac{B_n}{(n!)^s} = \infty, \quad 0 \leq s < 1$$

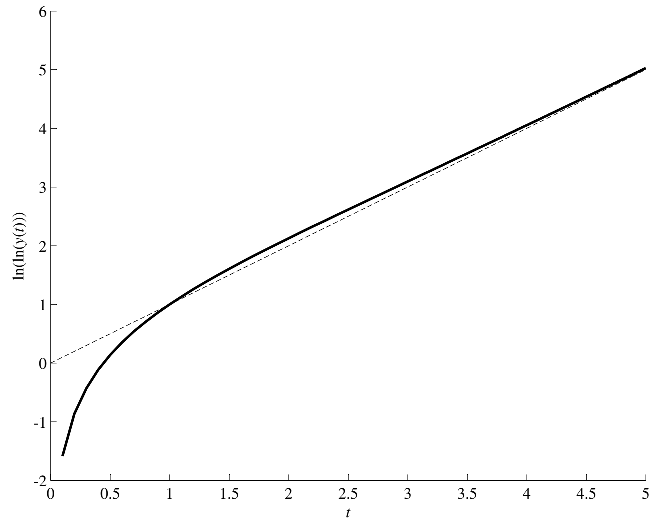


Fig. 3.7. The function $y(t) = \mathrm{e}^{\mathrm{e}^t - 1}$ in Example 3.8 plotted on a double logarithmic scale (solid line) and the function $\tilde{y}(t) = t$ (dashed line).

$$\lim_{n\to\infty} \frac{B_n}{n!} = 0.$$

The first two limits imply that the Bell numbers are growing faster than the global growth rate (3.11), where it is always assumed that $s \in [0, 1)$. The third limit indicates a growth rate corresponding to $s^* = 1$ and $0 < M < 1$. In which case, there exist globally convergent Fliess operators which do *not* satisfy global growth condition (3.11). This issue will be revisited in Section 3.6 where cascade interconnections are considered in detail. It is in this context that such systems can naturally appear. □

Finally, the global analogue of Theorem 3.6 is considered. This result will be used in Chapter 6, where state space realizations of F_c are considered. The proof is very similar to its local counterpart except that the global growth condition now permits an arbitrary $T > 0$ to be handled in much the same way as it was in the previous theorem. The details are thus left to the reader.

Theorem 3.9 *If* $c \in \mathbb{R}^\ell_{GC}\langle\langle X\rangle\rangle$ *and* $u \in L^m_{\mathfrak{p},e}(t_0)$ *with* $\mathfrak{p} \in [1, \infty]$ *is real analytic, then* $y = F_c[u]$ *is real analytic on* $[t_0, t_0 + T]$ *for any* $T > 0$.

3.4 Volterra Operators

Volterra operators date back to as early as the 1880s and are arguably the most widely used class of nonlinear operators in science and engineering. They can be viewed as a natural generalization of a linear integral operator. Recall from Chapter 1 that a causal linear input-output system taking m inputs to ℓ outputs can be expressed in terms of the linear integral operator

$$\begin{aligned} y_1(t) &= \int_{t_0}^{t} w(t,\tau)u(\tau)\,d\tau \\ &= \sum_{i=1}^{m} \int_{t_0}^{t} w_i(t,\tau)u_i(\tau)\,d\tau, \quad t \geq t_0. \end{aligned}$$

One could define a *second-order integral operator* as

$$y_2(t) = \sum_{i_1,i_2=1}^{m} \int_{t_0}^{t} \int_{t_0}^{\tau_2} w_{i_2 i_1}(t,\tau_2,\tau_1)u_{i_2}(\tau_2)u_{i_1}(\tau_1)\,d\tau_1\,d\tau_2$$

or a *k-th order integral operator* as

$$\begin{aligned} y_k(t) = \sum_{i_1,\ldots,i_k=1}^{m} \int_{t_0}^{t} \int_{t_0}^{\tau_k} \cdots \int_{t_0}^{\tau_2} & w_{i_k \cdots i_1}(t,\tau_k,\ldots,\tau_1)\cdot \\ & u_{i_k}(\tau_k)\cdots u_{i_1}(\tau_1)\,d\tau_1\cdots d\tau_k. \end{aligned}$$

This motivates the following definition.[7]

Definition 3.5 *A **Volterra operator** is any mapping of the form*

$$V : u \mapsto y(t) = \sum_{k=0}^{\infty} y_k(t),$$

where the zero-order (nonhomogeneous) term is formally defined as $y_0(t) = w_\emptyset(t)$.

Now in the event that the *kernel functions* $w_{i_k \cdots i_1}$ are real analytic over a common domain, the following theorem states that V can have a Fliess operator representation.

[7] Caution, the subscripts on output y are *not* indicating component functions in this section

Theorem 3.10 *Suppose V is a Volterra operator, where each kernel function has a series representation*

$$w_{i_k\cdots i_1}(t,\tau_k,\ldots,\tau_1) = \sum_{n_0,\ldots,n_k=0}^{\infty} (c, x_0^{n_k} x_{i_k} x_0^{n_{k-1}} \ldots x_{i_1} x_0^{n_0}) \cdot \frac{(t-\tau_k)^{n_k}(\tau_k-\tau_{k-1})^{n_{k-1}}\cdots(\tau_1-t_0)^{n_0}}{n_k!\, n_{k-1}! \cdots n_0!}$$

on some domain

$$\mathcal{D}_k = \{(t,\tau_k,\ldots,\tau_1) \in \mathbb{R}^{k+1} : t_0 + T_0 \geq t \geq \tau_k \geq \cdots \geq \tau_1 \geq t_0\}.$$

If $c \in \mathbb{R}^{\ell}_{LC}\langle\langle X \rangle\rangle$ and $\mathfrak{p} \in [1,\infty]$, then there exists $R, T > 0$ such that $V = F_c$ on $B_{\mathfrak{p}}^m(R)[t_0, t_0+T]$. If $c \in \mathbb{R}^{\ell}_{GC}\langle\langle X \rangle\rangle$, then $V = F_c$ on $L^m_{\mathfrak{p},e}(t_0)$.

Proof: For $k \geq 1$ observe

$$\begin{aligned} y_k(t) &= \sum_{i_1,\ldots,i_k=1}^{m} \sum_{n_0,\ldots,n_k=0}^{\infty} (c, x_0^{n_k} x_{i_k} x_0^{n_{k-1}} \ldots x_{i_1} x_0^{n_0}) \cdot \\ &\quad \int_{t_0}^{t}\int_{t_0}^{\tau_k}\cdots\int_{t_0}^{\tau_2} \frac{(t-\tau_k)^{n_k}}{n_k!} u_{i_k}(\tau_k) \frac{(\tau_k-\tau_{k-1})^{n_{k-1}}}{n_{k-1}!} \cdots \\ &\quad u_{i_1}(\tau_1) \frac{(\tau_1-t_0)^{n_0}}{n_0!}\, d\tau_1 \cdots d\tau_k \\ &= \sum_{i_1,\ldots,i_k=1}^{m} \sum_{n_0,\ldots,n_k=0}^{\infty} (c, x_0^{n_k} x_{i_k} x_0^{n_{k-1}} \cdots x_{i_1} x_0^{n_0}) \cdot \\ &\quad E_{x_0^{n_k} x_{i_k} x_0^{n_{k-1}} \cdots x_{i_1} x_0^{n_0}}[u](t,t_0). \end{aligned}$$

When $k = 0$

$$\begin{aligned} y_0(t) &= (c,\emptyset) + (c,x_0)\frac{t-t_0}{1!} + (c,x_0^2)\frac{(t-t_0)^2}{2!} + \cdots \\ &= \sum_{n_0=0}^{\infty} (c, x_0^{n_0}) E_{x_0^{n_0}}[u](t,t_0). \end{aligned}$$

Therefore,

$$V[u] = \sum_{k=0}^{\infty} y_k = \sum_{\eta \in X^*} (c,\eta) E_\eta[u] = F_c[u],$$

which is well defined for every u on some closed-ball $B_{\mathfrak{p}}^m(R)[t_0, t_0 + T]$ if $c \in \mathbb{R}^{\ell}_{LC}\langle\langle X\rangle\rangle$ (Theorem 3.1) and on $L^m_{\mathfrak{p},e}(t_0)$ if $c \in \mathbb{R}^{\ell}_{GC}\langle\langle X\rangle\rangle$ (Theorem 3.8). ■

Example 3.9 Reconsider the Wiener system in Example 3.3. It was shown earlier that

$$\begin{aligned} y(t) &= \sum_{k=0}^{\infty} k!\, E_{x_1^k}[u](t,t_0) \\ &= 1 + \sum_{k=1}^{\infty} \int_{t_0}^{t}\int_{t_0}^{\tau_k}\cdots\int_{t_0}^{\tau_2} k!\; u_1(\tau_k)\cdots u_1(\tau_1)\, d\tau_1\cdots d\tau_k. \end{aligned}$$

Thus, by direct observation, the kernel functions of the corresponding Volterra operator are

$$\begin{aligned} w_{\underbrace{11\cdots 1}_{k\text{ times}}}(t,\tau_k,\ldots,\tau_1) &= k!, \quad k\geq 1 \\ w_{\emptyset}(t) &= 1. \end{aligned}$$

It is clear that the generating series c is locally convergent with growth constants $K = M = 1$. So the Volterra operator with $t_0 = 0$ is well defined on $B_1(R)[0,T]$ for sufficiently small $R, T > 0$. □

Example 3.10 Consider a series

$$c = \sum_{k=0}^{\infty}\sum_{i_1,\ldots,i_k=0}^{m} (\lambda N_{i_k}\cdots N_{i_1}\gamma)\, x_{i_k}\cdots x_{i_1},$$

where $N_{i_j} \in \mathbb{R}^{n\times n}$, $i_j \in \{0,1,\ldots,m\}$ and $\gamma, \lambda^T \in \mathbb{R}^{n\times 1}$. Each word $x_{i_k}\cdots x_{i_1}$ can be rewritten uniquely in the form

$$x_0^{n_k} x_{i_k} x_0^{n_{k-1}} x_{i_{k-1}} \cdots x_0^{n_1} x_{i_1} x_0^{n_0},$$

where now $i_j \in \{1,2,\ldots,m\}$ (see equation (2.43)). In which case, c has the equivalent form

$$c = \sum_{k=0}^{\infty}\sum_{i_1,\ldots,i_k=1}^{m}\sum_{n_0,\ldots,n_k=0}^{\infty} (\lambda N_0^{n_k} N_{i_k} N_0^{n_{k-1}}\cdots N_{i_1} N_0^{n_0}\gamma)\cdot$$

$$x_0^{n_k} x_{i_k} x_0^{n_{k-1}} \cdots x_{i_1} x_0^{n_0}.$$

By inspection, the corresponding Volterra kernel function for any $k \geq 1$ is

$$\begin{aligned} w_{i_k \cdots i_1}(t, \tau_k, \ldots, \tau_1) &= \sum_{n_0,\ldots,n_k=0}^{\infty} (\lambda N_0^{n_k} N_{i_k} N_0^{n_{k-1}} \cdots N_{i_1} N_0^{n_0} \gamma) \cdot \\ &\qquad \frac{(t-\tau_k)^{n_k} (\tau_k - \tau_{k-1})^{n_{k-1}} \cdots (\tau_1 - t_0)^{n_0}}{n_k!\, n_{k-1}! \cdots n_0!} \\ &= \lambda \mathrm{e}^{N_0(t-\tau_k)} N_{i_k} \mathrm{e}^{N_0(\tau_k - \tau_{k-1})} \cdots N_{i_1} \mathrm{e}^{N_0(\tau_1 - t_0)} \gamma, \end{aligned}$$

and similarly,

$$w_\emptyset(t) = \lambda \mathrm{e}^{N_0(t-t_0)} \gamma.$$

Therefore, the associated Volterra operator is

$$\begin{aligned} y(t) = \lambda \mathrm{e}^{N_0(t-t_0)}\gamma + &\\ &\sum_{k=1}^{\infty} \sum_{i_1,\ldots,i_k=1}^{m} \int_{t_0}^{t} \int_{t_0}^{\tau_k} \cdots \int_{t_0}^{\tau_2} \lambda \mathrm{e}^{N_0(t-\tau_k)} N_{i_k} \mathrm{e}^{N_0(\tau_k - \tau_{k-1})} \cdots \\ &N_{i_1} \mathrm{e}^{N_0(\tau_1 - t_0)} \gamma \, u_{i_k}(\tau_k) \cdots u_{i_1}(\tau_1)\, d\tau_1 \cdots d\tau_k. \end{aligned}$$

It is easily verified that $c \in \mathbb{R}_{GC}\langle\langle X \rangle\rangle$ (see Theorem 4.1), hence the operator is well defined on any $L_{\mathfrak{p},e}^m(t_0)$ space. Truncating this series to first order gives the linear input-output mapping

$$\begin{aligned} \hat{y}(t) &= y_0(t) + y_1(t) \\ &= \lambda \mathrm{e}^{N_0(t-t_0)}\gamma + \sum_{i=1}^{m} \int_{t_0}^{t} \lambda \mathrm{e}^{N_0(t-\tau)} N_i \mathrm{e}^{N_0(\tau - t_0)} \gamma \, u_i(\tau)\, d\tau. \end{aligned}$$

If, in addition, N_0 and N_i commute for each $i = 1, 2, \ldots, m$ then a corresponding linear state space system is

$$\begin{aligned} \dot{z} &= N_0 z + \sum_{i=1}^{m} N_i \gamma\, u_i, \quad z(t_0) = \gamma \\ \hat{y} &= \lambda z. \end{aligned}$$

□

3.5 Parallel Connections

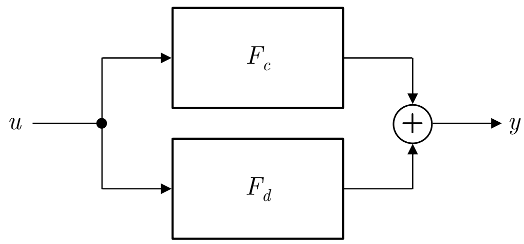
Given two input-output systems with Fliess operator representations F_c and F_d, it is natural in applications to interconnect them to form new and more complex systems. In this section and the two subsequent sections, three elementary types of connections are analyzed in detail: the parallel connection, the cascade connection, and the feedback connection. In each case, the following questions need to be addressed:

1. Is the composite system well-posed? That is, are the applied inputs to each subsystem well defined and admissible?
2. Does the composite system have a Chen-Fliess series representation?
3. If so, how exactly is the generating series of the composite system computed?
4. What is the nature of the convergence of the Chen-Fliess series representing the composite system? Is it divergent, locally convergent, globally convergent?
5. Finally, what can be said about the *radius of convergence* of the composite system if it is known to be only locally convergent? This concept will require a precise definition, but roughly speaking this will provide a more detailed characterization of the composite system's local convergence properties.

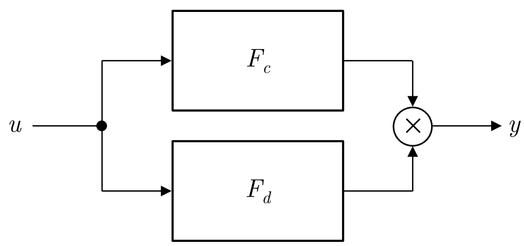
The starting point is the two parallel connections shown in Figure 3.8. Both F_c and F_d are driven with the same input function, u, and their respective outputs are either added or multiplied pointwise in time. In each case, the outputs are also combined componentwise, therefore, it is assumed throughout that both systems have the same number of outputs. These interconnections are clearly well-posed when both generating series are locally convergent in light of applying Theorem 3.1 using the growth constants $K = \max(K_c, K_d)$ and $M = \max(M_c, M_d)$. The first theorem states that both parallel connections have Chen-Fliess series representations, and the generating series is provided in each case.

Theorem 3.11 *Given Fliess operators F_c and F_d, where $c, d \in \mathbb{R}^{\ell}_{LC}\langle\langle X \rangle\rangle$, the parallel connections $F_c + F_d$ and F_cF_d have generating series $c + d$ and $c \sqcup\!\sqcup\, d$, respectively. That is, $F_c + F_d = F_{c+d}$ and $F_cF_d = F_{c \sqcup\!\sqcup\, d}$.*

Proof: For the parallel sum connection, observe that



(a) parallel connection with an adder (parallel sum)



(b) parallel connection with a multiplier (parallel product)

Fig. 3.8. Parallel system connections under consideration

$$
\begin{aligned}
F_c[u] + F_d[u] &= \sum_{\eta \in X^*} (c,\eta) E_\eta[u] + \sum_{\eta \in X^*} (d,\eta) E_\eta[u] \\
&= \sum_{\eta \in X^*} (c + d,\eta) E_\eta[u] \\
&= F_{c+d}[u].
\end{aligned}
$$

For the parallel product connection, in light of the componentwise definition of the shuffle product and Lemma 2.3, it follows that

$$
\begin{aligned}
F_c[u](t) F_d[u] &= \sum_{\eta \in X^*} (c,\eta) E_\eta[u] \sum_{\xi \in X^*} (d,\xi) E_\xi[u] \\
&= \sum_{\eta,\xi \in X^*} (c,\eta)(d,\xi)\, E_{\eta \sqcup\!\sqcup\, \xi}[u] \\
&= F_{c \sqcup\!\sqcup\, d}[u].
\end{aligned}
$$

■

Example 3.11 Reconsider Example 3.3, where

$$z(t) = \int_0^t u(\tau)\, d\tau, \quad y(t) = \frac{1}{1 - z(t)}.$$

The corresponding input-output equation can be shown by direct differentiation of y to be

$$\dot{y} - y^2 u = 0, \;\; y(0) = 1.$$

But since $y = F_c[u]$ with

$$c = \sum_{k=0}^{\infty} k!\, x_1^k,$$

this equation can also be verified starting with F_c by computing its square, which can be viewed as a parallel product connection, and then comparing the result against the time derivative F_c found via Theorem 3.4. Specifically, $y^2 = (F_c)^2 = F_{c \sqcup\!\sqcup\, c}$, where

$$\begin{aligned} c \sqcup\!\sqcup\, c &= \sum_{k,l=0}^{\infty} k!\, l!\, x_1^k \sqcup\!\sqcup\, x_1^l \\ &= \sum_{k,l=0}^{\infty} k!\, l! \binom{k+l}{k} x_1^{l+l} \\ &= \sum_{k,l=0}^{\infty} (k+l)!\, x_1^{k+l} \\ &= \sum_{k=0}^{\infty} \sum_{l=0}^{k} k!\, x_1^k \\ &= \sum_{k=0}^{\infty} (k+1)!\, x_1^k. \end{aligned}$$

In which case,

$$\begin{aligned} uy^2 &= u \sum_{k=0}^{\infty} (k+1)!\, E_{x_1^k} \\ &= uF_{x_1^{-1}(c)}[u] \\ &= \frac{d}{dt} F_c[u] \\ &= \dot{y} \end{aligned}$$

as expected. □

The next two theorems show that both $\mathbb{R}^\ell_{LC}\langle\langle X\rangle\rangle$ and $\mathbb{R}^\ell_{GC}\langle\langle X\rangle\rangle$ are closed under addition and the shuffle product.

Theorem 3.12 *Suppose $c, d \in \mathbb{R}^\ell_{LC}\langle\langle X\rangle\rangle$ with growth constants K_c, $M_c > 0$ and $K_d, M_d > 0$, respectively. Then $c + d \in \mathbb{R}^\ell_{LC}\langle\langle X\rangle\rangle$ and $c \sqcup\!\sqcup\, d \in \mathbb{R}^\ell_{LC}\langle\langle X\rangle\rangle$. Specifically,*

$$|(c + d, \nu)| \le (K_c + K_d) M^{|\nu|} |\nu|!, \quad \forall \nu \in X^*, \tag{3.13}$$

and

$$|(c \sqcup\!\sqcup\, d, \nu)| \le K_c K_d M^{|\nu|} (|\nu| + 1)!, \quad \forall \nu \in X^*, \tag{3.14}$$

where $M = \max\{M_c, M_d\}$.

Note that since $n + 1 \le 2^n$ for all $n \ge 0$, equation (3.14) implies the more conventional local convergence upper bound

$$|(c \sqcup\!\sqcup\, d, \nu)| \le K_c K_d (2M)^{|\nu|} |\nu|!, \quad \forall \nu \in X^*.$$

Proof: The upper bound regarding $c + d$ is trivial to produce. For the shuffle product, observe that

$$\begin{aligned}
|(c \sqcup\!\sqcup\, d, \nu)| &= \left| \sum_{k=0}^{|\nu|} \sum_{\substack{\eta \in X^k \\ \xi \in X^{|\nu|-k}}} (c, \eta)(d, \xi)(\eta \sqcup\!\sqcup\, \xi, \nu) \right| \\
&\le \sum_{k=0}^{|\nu|} \sum_{\substack{\eta \in X^k \\ \xi \in X^{|\nu|-k}}} K_c M_c^k k!\; K_d M_d^{|\nu|-k} (|\nu| - k)!\; (\eta \sqcup\!\sqcup\, \xi, \nu) \\
&\le K_c K_d M^{|\nu|} \sum_{k=0}^{|\nu|} k!\; (|\nu| - k)! \binom{|\nu|}{k} \\
&= K_c K_d M^{|\nu|} \sum_{k=0}^{|\nu|} |\nu|! \\
&= K_c K_d M^{|\nu|} (|\nu| + 1)!
\end{aligned}$$

(see Problem 2.4.3). ■

Theorem 3.13 *Let $c, d \in \mathbb{R}^{\ell}_{GC}\langle\langle X\rangle\rangle$ with the minimum of their Gevrey orders being s^*_c and s^*_d, respectively. Then $c+d \in \mathbb{R}^{\ell}_{GC}\langle\langle X\rangle\rangle$ and $c \sqcup\!\sqcup\, d \in \mathbb{R}^{\ell}_{GC}\langle\langle X\rangle\rangle$, and in particular, $s^*_{c+d} \leq \max(s^*_c, s^*_d)$ and $s^*_{c \sqcup\!\sqcup\, d} \leq \max(s^*_c, s^*_d)$.*

The proof for the theorem above can be found in the literature. A special case is addressed in Problem 3.5.1.

Example 3.12 Let $X = \{x_0, x_1\}$ and consider two linear series $c, d \in \mathbb{R}_{GC}\langle\langle X\rangle\rangle$ with corresponding coefficients

$$(c, x_0^k x_1) = C_c A_c^k B_c, \quad (d, x_0^k x_1) = C_d A_d^k B_d, \quad k \geq 0$$

(see Example 3.6). Then the generating series for the parallel sum connection has coefficients

$$\begin{aligned}(c+d, x_0^k x_1) &= C_c A_c^k B_c + C_d A_d^k B_d \\ &= [C_c \;\; C_d](\operatorname{diag}(A_c, A_d))^k [B_c^T \;\; B_d^T]^T, \quad k \geq 0.\end{aligned}$$

Since this connection produces another linear series in the same class as c and d, the composite system is also globally convergent with $s^*_{c+d} = 0$.

□

Example 3.13 Let $X_0 = \{x_0\}$ and consider two series $c, d \in \mathbb{R}[[X_0]]$ with coefficients

$$(c, x_0^k) = C_c A_c^k z_c, \quad (d, x_0^k) = C_d A_d^k z_d, \quad k \geq 0,$$

where $A_c \in \mathbb{R}^{n_c \times n_c}$, $A_d \in \mathbb{R}^{n_d \times n_d}$, $C_c^T, z_c \in \mathbb{R}^{n_c \times 1}$. $C_d^T, z_d \in \mathbb{R}^{n_d \times 1}$. Then

$$\begin{aligned}(c \sqcup\!\sqcup\, d, x_0^k) &= \sum_{i,j=0}^{\infty} (C_c A_c^i z_c)(C_d A_d^j z_d)(x_0^i \sqcup\!\sqcup\, x_0^j, x_0^k) \\ &= \sum_{j=0}^{k} (C_c A_c^{k-j} z_c)(C_d A_d^j z_d)\binom{k}{j} \\ &= (C_c \otimes C_d)\left[\sum_{j=0}^{k} (A_c^{k-j} \otimes A_d^j)\binom{k}{j}\right](z_c \otimes z_d),\end{aligned}$$

using Problem 2.4.5(b) and Kronecker product identity

$$(A \otimes B)(C \otimes D) = AB \otimes CD.$$

From the definition of the Kronecker sum,

$$A \oplus B = (A \otimes I_{n_d}) + (I_{n_c} \otimes B),$$

it follows that

$$(A \oplus B)^k = \sum_{j=0}^{k} (A^{k-j} \otimes B^j) \binom{k}{j}$$

(see Problem 3.5.2). Therefore, the series $c \sqcup\!\sqcup\, d$ has coefficients

$$(c \sqcup\!\sqcup\, d, x_0^k) = (C_c \otimes C_d)(A_c \oplus A_d)^k (z_c \otimes z_d), \quad k \geq 0,$$

which is clearly globally convergent. The series c and d are associated with autonomous linear state space realizations (A_c, C_c, z_c) and (A_d, C_d, z_d), respectively. In which case, the triple $(A_c \oplus A_d, C_c \otimes C_d, z_c \otimes z_d)$ is a linear state space realization of the input-output system $F_{c \sqcup\!\sqcup\, d}[u]$ when $u = 0$. □

Now that it has been shown that local and global convergence are preserved for the generating series of parallel connections, a finer convergence analysis is pursued for the local case, that is when the minimum Gevrey order is $s^* = 1$. In Theorem 3.1 it was established that the series defining $y = F_c[u]$ converges on $B_1^m(R)[t_0, t_0 + T]$ if R and T satisfy

$$\max\{R, T\} < \frac{1}{M_c(m+1)}.$$

Clearly, the smaller the geometric growth constant M_c, the larger R and T can be while maintaining a well defined output function y. It is natural to ask, therefore, whether the growth constant $M = \max\{M_c, M_d\}$ found in Theorem 3.12 for the parallel connections is the *smallest* possible geometric growth constant for $c + d$ and $c \sqcup\!\sqcup\, d$, assuming of course that M_c and M_d are the smallest possible geometric growth constants for c and d, respectively. If not, then the domains of F_{c+d} and $F_{c \sqcup\!\sqcup\, d}$ are being unnecessarily restricted. To answer this question, some additional concepts and tools are needed. Let $\pi : \mathbb{R}^{\ell}_{LC}\langle\langle X \rangle\rangle \to \mathbb{R}^+ \cup \{0\}$ be the mapping which takes each series c to the infimum of all geometric growth constants satisfying (3.2). Note that since $\mathbb{R}^{\ell}_{GC}\langle\langle X\rangle\rangle \subset \mathbb{R}^{\ell}_{LC}\langle\langle X\rangle\rangle$, it is possible for $\pi(c) = 0$. Now partition $\mathbb{R}^{\ell}_{LC}\langle\langle X\rangle\rangle$ into equivalence classes and define $1/(M_c(m+1))$ to be the *radius of convergence* for the class $\pi^{-1}(M_c)$ when $M_c > 0$. This is in contrast to the usual situation where a radius of convergence is assigned to individual series. In practice, it is not difficult to estimate the minimal M_c for many series either analytically or numerically, in which case, the radius of convergence for $\pi^{-1}(M_c)$ provides an easily computed *lower bound* for the radius of convergence of c in the usual sense as not every coefficient of c necessarily grows at the maximum rate. The specific goal here is to determine analytically this lower bound for $c + d$ and $c \sqcup\!\sqcup d$. The following definition is useful.

Definition 3.6 *Given an alphabet X and a fixed $s \in \mathbb{R}$, the* ***maximal series*** *having growth constants $K, M > 0$ is the element in $\mathbb{R}^{\ell}\langle\langle X\rangle\rangle$ where each component series has the form $\bar{c}_i = \sum_{\eta\in X^*} KM^{|\eta|}(|\eta|!)^s\,\eta$. When $s = 1$ the series is called* ***locally maximal****.*

In the context of convergence analysis, locally maximal series represent worst case series in that every coefficient in the series is growing at its maximum possible rate.[8] The following theorem from complex analysis is the main technical tool needed for the calculation.

Theorem 3.14 *Let $f(z) = \sum_{n\geq 0} a_n z^n/n!$ be an analytic function on some neighborhood of the origin in the complex plane. Suppose $z_0 \neq 0$ is a singularity of f of smallest modulus. Then for any $\epsilon > 0$ there exists an integer $N \geq 0$ such that for all $n > N$, $|a_n| < ((1/|z_0|) + \epsilon)^n\, n!$. Furthermore, for infinitely many n, $|a_n| > ((1/|z_0|) - \epsilon)^n\, n!$.*

The essence of this theorem is that the real number $1/|z_0|$ is an *infimum* on the set of geometric growth constants bounding the magnitude of coefficients a_n, $n \geq 0$ of the exponential generating series for the function f. That is, select any $\epsilon > 0$ and as shown in Figure 3.9, the coefficients will eventually be bounded by $((1/|z_0|) + \epsilon)^n n!$ when n is sufficiently large. Furthermore, no number smaller than $1/|z_0|$ will have this property. No claim is made regarding the case where $\epsilon = 0$. Finally, one can always introduce a $K > 1$, if necessary, so that $|a_n| \leq K((1/|z_0|) + \epsilon)^n n!$, $n \geq 0$. The next lemma applies this theorem and provides the crucial insight into determining the radius of convergence for the parallel sum connection.

[8] See Appendix B for additional information.

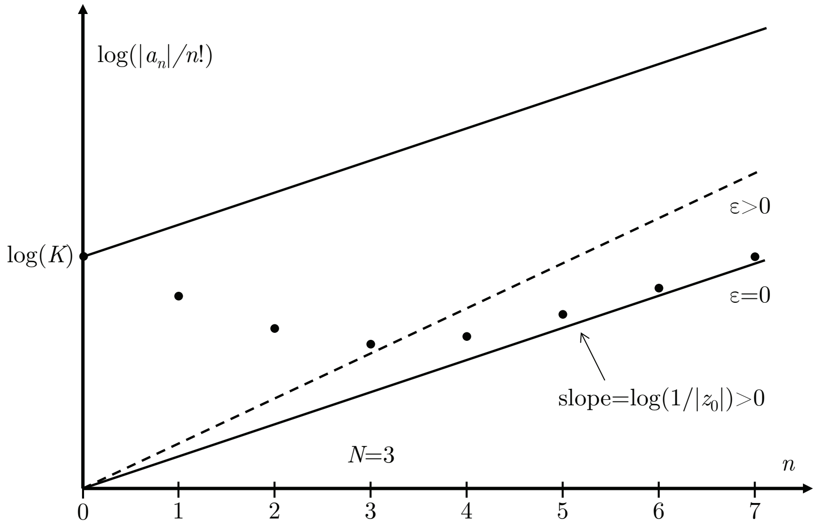


Fig. 3.9. A typical growth bound for the coefficients of an exponential generating series of a function which is real analytic at $z = 0$ as described in Theorem 3.14.

Lemma 3.2 *Suppose* $X = \{x_0, x_1, \ldots, x_m\}$*. Let* $\bar{c}$ *and* $\bar{d}$ *be locally maximal series with growth constants* $K_c, M_c > 0$ *and* $K_d, M_d > 0$*, respectively. If* $\bar{b} = \bar{c} + \bar{d}$*, then the sequence* $(\bar{b}_i, x_0^k)$*,* $k \geq 0$ *has the exponential generating function*

$$f(x_0) := \sum_{k=0}^{\infty} (\bar{b}_i, x_0^k) \frac{x_0^k}{k!} = \frac{K_c}{1 - M_c x_0} + \frac{K_d}{1 - M_d x_0}$$

for every $i = 1, 2, \ldots, \ell$*. Moreover, the infimum of all possible geometric growth constants for* $\bar{b}$ *is*

$$M_b = \max\{M_c, M_d\}.$$

Proof: There is no loss of generality in assuming $\ell = 1$. Observe for any $\nu \in X^n$, $n \geq 0$ that

$$(\bar{b}, \nu) = (\bar{c}, \nu) + (\bar{d}, \nu) = (K_c M_c^n + K_d M_d^n)\, n!.$$

Furthermore, $(\bar{b}, \nu) = (\bar{b}, x_0^n)$, $n \geq 0$. The key idea is that $f(t)$ is the zero-input response of $F_{\bar{b}}$. Specifically,

$$f(t) = \sum_{k=0}^{\infty} (\bar{b}, x_0^k) \frac{t^k}{k!} = F_{\bar{b}}[0] = F_{\bar{c}}[0] + F_{\bar{d}}[0]$$

$$= \sum_{k=0}^{\infty} K_c M_c^k t^k + \sum_{k=0}^{\infty} K_d M_d^k t^k = \frac{K_c}{1 - M_c t} + \frac{K_d}{1 - M_d t}. \quad (3.15)$$

Since f is analytic at the origin, by Theorem 3.14 the infimum of all geometric growth constants for the sequence $(\bar{b}, x_0^n)$, $n \geq 0$, and thus for the entire formal power series $\bar{b}$, is determined by the location of any singularity nearest to the origin in the complex plane, say z_0. Specifically, $M_b = 1/|z_0|$, where it is easily verified from (3.15) that z_0 is the positive real number

$$z_0 = \frac{1}{\max\{M_c, M_d\}}.$$

This proves the lemma. ■

Now the main result is given below. It confirms what was not at all obvious in Theorem 3.12, i.e., that $M_b = \max\{M_c, M_d\}$ is the *minimum* geometric growth constant for all parallel sum connections given only the geometric growth constants for the generating series of the component systems.

Theorem 3.15 *Suppose $X = \{x_0, x_1, \ldots, x_m\}$. Let $c, d \in \mathbb{R}^{\ell}_{LC}\langle\langle X \rangle\rangle / \mathbb{R}^{\ell}_{GC}\langle\langle X \rangle\rangle$ with growth constants $K_c, M_c > 0$ and $K_d, M_d > 0$, respectively. If $b = c + d$, then*

$$|(b, \nu)| \leq K_b M_b^{|\nu|} |\nu|!, \quad \nu \in X^* \quad (3.16)$$

for some $K_b > 0$, where $M_b = \max\{M_c, M_d\}$. Furthermore, if M_c and M_d are minimal, then no smaller geometric growth constant can satisfy (3.16). Thus, the radius of convergence of $c + d$ is

$$\frac{1}{M_b(m+1)}.$$

Proof: First observe that for all $\nu \in X^*$ and $i = 1, 2, \ldots, \ell$

$$\begin{aligned} |(c + d, \nu)| &\leq |(c, \nu)| + |(d, \nu)| \\ &\leq (\bar{c}_i, \nu) + (\bar{d}_i, \nu) \\ &= (\bar{b}_i, \nu), \end{aligned}$$

where $\bar{c}, \bar{d}$ and $\bar{b}$ are defined as in Lemma 3.2. Applying this lemma, it is clear that for any $\epsilon > 0$ and some $K_b > 0$

$$(\bar{b}_i,\nu) \leq K_b(M_b+\epsilon)^{|\nu|}|\nu|!,\ \nu \in X^*.$$

Furthermore, there is no geometric growth constant smaller than M_b since b and $\bar{b}$ are in the same growth equivalence class. In this specific case, it was shown in Theorem 3.12 that the bound *does* apply even when $\epsilon = 0$. ■

A similar analysis is now undertaken for the parallel product connection. The following lemma is a prerequisite.

Lemma 3.3 *Suppose* $X = \{x_0, x_1, \ldots, x_m\}$. *Let* $\bar{c}, \bar{d} \in \mathbb{R}^{\ell}_{LC}\langle\langle X\rangle\rangle$ *be locally maximal series with growth constants* $K_c, M_c > 0$ *and* $K_d, M_d > 0$*, respectively. If* $\bar{b} = \bar{c} \sqcup\!\sqcup \bar{d}$*, then the sequence* $(\bar{b}_i, x_0^k)$*,* $k \geq 0$ *has the exponential generating function*

$$f(x_0) = \frac{K_c K_d}{(1-M_c x_0)(1-M_d x_0)}$$

for every $i = 1, 2, \ldots, \ell$*. Moreover, the infimum of all possible geometric growth constants for* $\bar{b}$ *is*

$$M_b = \max\{M_c, M_d\}.$$

Proof: There is no loss of generality in assuming $\ell = 1$. Observe for any $\nu \in X^n$, $n \geq 0$ that

$$\begin{aligned}
(\bar{b},\nu) &= \sum_{j=0}^{n} \sum_{\substack{\eta\in X^j \\ \xi \in X^{n-j}}} (\bar{c},\eta)(\bar{d},\xi)(\eta \sqcup\!\sqcup \xi,\nu) \\
&= \sum_{j=0}^{n} K_c M_c^j j!\, K_d M_d^{n-j}(n-j)! \sum_{\substack{\eta\in X^j \\ \xi \in X^{n-j}}} (\eta \sqcup\!\sqcup \xi,\nu) \\
&= \sum_{j=0}^{n} K_c M_c^j j!\, K_d M_d^{n-j}(n-j)! \binom{n}{j} \\
&= K_c K_d \left[\sum_{j=0}^{n} M_c^j M_d^{n-j}\right] n!. \qquad (3.17)
\end{aligned}$$

Therefore, $\bar{b}$ and the sequence $(\bar{b}, x_0^n)$, $n \geq 0$ will have the same infimum on their geometric growth constants. Next note that $f(t)$ is the zero-input response of $F_{\bar{b}}$. Specifically,

$$f(t) = \sum_{k=0}^{\infty} (\bar{b}_i, x_0^k)\frac{t^k}{k!} = F_{\bar{b}}[0] = F_{\bar{c}}[0]F_{\bar{d}}[0]$$
$$= \sum_{k=0}^{\infty} K_c M_c^k t^k \sum_{k=0}^{\infty} K_d M_d^k t^k = \frac{K_c K_d}{(1-M_c t)(1-M_d t)}.$$

Since f is analytic at the origin, Theorem 3.14 is applied to compute the infimum on the geometric growth constants, namely, $M_b = 1/|z_0|$, where
$$z_0 = \frac{1}{\max\{M_c, M_d\}}.$$
This proves the theorem. ■

Now the main convergence result for this interconnection is presented. There is one significant difference from the previous case, namely, M_b is only an infimum on the geometric growth constants, not a minimum.

Theorem 3.16 *Suppose $X = \{x_0, x_1, \ldots, x_m\}$. Let $c, d \in \mathbb{R}^{\ell}_{LC}\langle\langle X\rangle\rangle / \mathbb{R}^{\ell}_{GC}\langle\langle X\rangle\rangle$ with growth constants $K_c, M_c > 0$ and $K_d, M_d > 0$, respectively. If $b = c \sqcup\!\sqcup d$, then for every $\epsilon > 0$*
$$|(b,\nu)| \le K_b(M_b+\epsilon)^{|\nu|}|\nu|!, \quad \nu \in X^* \tag{3.18}$$
for some $K_b > 0$, where $M_b = \max\{M_c, M_d\}$. Furthermore, if M_c and M_d are minimal, then no smaller geometric growth constant can satisfy (3.18), and thus the radius of convergence of $c \sqcup\!\sqcup d$ is
$$\frac{1}{M_b(m+1)}.$$

Proof: Assume $\ell = 1$ and observe for all $\nu \in X^*$ that
$$|(c \sqcup\!\sqcup d,\nu)| \le \sum_{j=0}^{n} \sum_{\substack{\eta\in X^j\\ \xi\in X^{n-j}}} |(c,\eta)||(d,\xi)|(\eta \sqcup\!\sqcup \xi,\nu)$$
$$\le \sum_{j=0}^{n} \sum_{\substack{\eta\in X^j\\ \xi\in X^{n-j}}} (\bar{c},\eta)(\bar{d},\xi)(\eta \sqcup\!\sqcup \xi,\nu)$$
$$= (\bar{b},\nu),$$

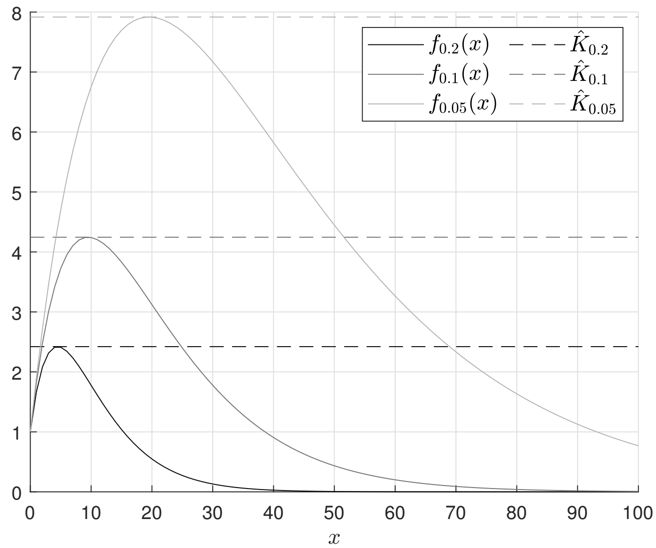


Fig. 3.10. Sample plots of $f_\epsilon(x)$ and $\hat{K}_\epsilon$.

where $\bar{b}$, $\bar{c}$ and $\bar{d}$ are defined as in Lemma 3.3. A direct application of this lemma gives for any $\epsilon > 0$

$$(\bar{b}, \nu) \leq K_b(M_b + \epsilon)^{|\nu|}|\nu|!,\ \nu \in X^*$$

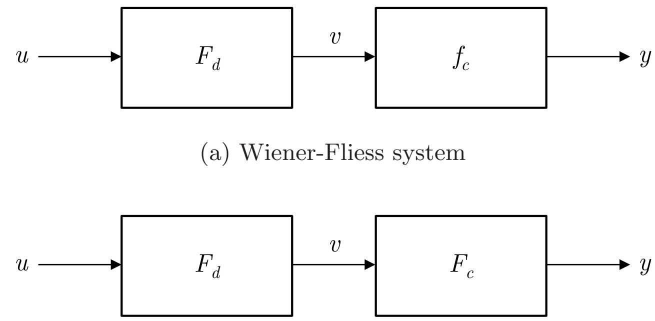
for some $K_b > 0$. Furthermore, $(\bar{b}, x_0^n)$, $n \geq 0$ can not be bounded by any smaller geometric growth constant. Thus, the theorem is proved.

■

Since the exponential generating functions in Lemma 3.2 and Lemma 3.3 have identical sets of singularities, the parallel sum and parallel product connections must have the same radii of convergence. But as indicated early, their behavior exactly at the boundary of the region of convergence is different. Specifically, the parallel sum connection is well defined at the boundary, while the parallel product connection is not. To see this, set $M = M_c = M_d$ in the proof of Lemma 3.3 and define $M_\epsilon = M(1+\epsilon)$ with $\epsilon > 0$. If there exists a K_b such that $(\bar{b}, \nu) \leq K_b M^{|\nu|}\, |\nu|!$ for all $\nu \in X^*$, then necessarily

$$\sup_{\nu \in X^*} \frac{(\bar{b}, \nu)}{M\, |\nu|!} \leq K_b < \infty.$$

From (3.17) it follows directly that



(b) cascade connection of two Fliess operators

Fig. 3.11. Cascade system connections

$$\sup_{\nu \in X^*} \frac{(\bar{b}, \nu)}{M_\epsilon \, |\nu|!} = K_c K_d \sup_{\nu \in X^*} \frac{|\nu| + 1}{(1+\epsilon)^{|\nu|}} = K_c K_\epsilon,$$

where $K_\epsilon := \sup_{\nu \in X^*} (|\nu| + 1)/(1+\epsilon)^{|\nu|}$. An upper bound for K_ϵ is found by showing that $f_\epsilon(x) = (x+1)/(1+\epsilon)^x$ has a single maximum at $x^*_\epsilon = (1/\log(1+\epsilon)) - 1 > 0$ when $0 < \epsilon \leq \mathrm{e} - 1$. Therefore,

$$K_\epsilon \leq \hat{K}_\epsilon := f_\epsilon(x^*_\epsilon) = \mathrm{e}^{-1} \frac{1+\epsilon}{\log(1+\epsilon)}.$$

In this case, the upper bound is tight (see Figure 3.10). For $\epsilon > \mathrm{e} - 1$, $K_\epsilon = 1$ and $\hat{K}_\epsilon > 1$, and thus this upper bound is conservative. In addition, since K_ϵ becomes unbounded as ϵ vanishes, this parallel product connection is not well defined directly on the boundary of the region of convergence.

3.6 Cascade Connections

A cascade or series connection is an interconnection where the output of one system is passed to the input of another system. Two examples of cascade connections are shown in Figure 3.11. The *Wiener-Fliess system* is a generalization of the classical Wiener system, where the linear operator in the first (left-most) system is replaced by a Fliess operator (see Example 3.3). The cascade of two Fliess operators is an interconnection of two dynamical systems. The Wiener-Fliess system is *not* a special case of such an interconnection since in general Fliess operators are not memoryless ($F_{c_0}[u](t) = c_0 \in \mathbb{R}$ for all u and t being the only exception). Therefore, the underlying algebraic structures describing these two types of cascade are distinct.

As with the parallel connections, the same set of five basic questions needs to be addressed. Consider first, as an example, the *well-posedness* of the cascade connection of two Fliess operators with $m = \ell$. The main issue is whether the output of the first system v is an admissible input to the second system F_c. This is most easily handled by applying Theorem 3.2 twice, once for $v = F_d[u]$ and once for $y = F_c[v]$. Specifically, set $M = \max\{M_c, M_d\}$ and select R_c, R_d, and T such that

$$R_{1,c} := \max\{R_c T^{1/p}, T\} < \frac{1}{(m+1)M}$$
$$R_{1,d} := \max\{R_d T^{1/q}, T\} < \frac{1}{(m+1)M}.$$

In which case, operators F_c and F_d converge on $[0,T]$ provided $u \in B_p(R_d)[0,T]$ and $v \in B_q(R_c)[0,T]$. Their corresponding outputs must reside in $B_p(S_c)[0,T]$ and $B_q(S_d)[0,T]$, respectively, where

$$S_c = \frac{K_c T^{1/p}}{1-(m+1)MR_{1,c}}$$
$$S_d = \frac{K_d T^{1/q}}{1-(m+1)MR_{1,d}}.$$

Therefore, the output of F_d will be an admissible input for F_c whenever $B_q(S_d)[0,T] \subseteq B_q(R_c)[0,T]$. That is, when

$$\frac{K_d T^{1/q}}{1-(m+1)MR_{1,d}} \leq R_c.$$

Observe that if this is not the case, then T can always be decreased to produce this condition. Hence, this cascade connection can always well-posed. A similar argument can be given for Wiener-Fliess systems.

Determining whether a cascade system has a Chen-Fliess series representation and what are the convergence properties of the composite system requires much more work. The first two theorems below state that for each case shown in Figure 3.11 the cascade system *does* have a Chen-Fliess series representation, and an explicit expression is given for its generating series using composition products as described in Section 2.7.

Theorem 3.17 *Let* $X = \{x_0, x_1, \ldots, x_m\}$ *and* $\tilde{X} = \{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_{\tilde{m}}\}$. *Given a Fliess operator* F_d, $d \in \mathbb{R}^{\tilde{m}}_{LC}\langle\langle X\rangle\rangle$ *and a function* $f_c : \mathbb{R}^{\tilde{m}} \to \mathbb{R}^{\ell}$ *with generating series* $c \in \mathbb{R}^{\ell}_{LC}[[\tilde{X}]]$ *at* $z = (d, \emptyset)$*, namely,*

$$f_c(z) = \sum_{\tilde{\eta}\in\tilde{X}^*} (c, \tilde{\eta}) \frac{(z - (d, \emptyset))^{\tilde{\eta}}}{\tilde{\eta}!},$$

the cascade connection $f_c \circ F_d$ *has the generating series in* $\mathbb{R}^{\ell}\langle\langle X\rangle\rangle$

$$c \circ d := \sum_{\tilde{\eta}\in\tilde{X}^*} (c, \tilde{\eta}) \frac{(d - (d, \emptyset))^{\sqcup\!\sqcup\, \tilde{\eta}}}{\tilde{\eta}!}.$$

That is, $f_c \circ F_d = F_{c\circ d}$.

Proof: The proof follows from elementary properties of the shuffle product. Defining the proper series $\tilde{d} := d - (d, \emptyset)$, observe that

$$\begin{aligned} f_c \circ F_d[u] &= \sum_{\tilde{\eta}\in\tilde{X}^*} (c, \tilde{\eta}) \frac{(z - (d, \emptyset))^{\tilde{\eta}}}{\tilde{\eta}!}\bigg|_{z=F_d[u]} \\ &= \sum_{\tilde{\eta}\in\tilde{X}^*} (c, \tilde{\eta}) \frac{(F_d[u] - F_{(d,\emptyset)}[u])^{\tilde{\eta}}}{\tilde{\eta}!} \\ &= \sum_{\tilde{\eta}\in\tilde{X}^*} \frac{(c, \tilde{\eta})}{\tilde{\eta}!} (F_{\tilde{d}}[u])^{\tilde{\eta}}. \end{aligned}$$

If $\tilde{\eta} = \tilde{x}_{i_k} \cdots \tilde{x}_{i_1}$ then

$$\begin{aligned} f_c \circ F_d[u] &= \sum_{\tilde{\eta}\in\tilde{X}^*} \frac{(c, \tilde{\eta})}{\tilde{\eta}!} F_{\tilde{d}_{i_k}}[u] F_{\tilde{d}_{i_{k-1}}}[u] \cdots F_{\tilde{d}_{i_1}}[u] \\ &= \sum_{\tilde{\eta}\in\tilde{X}^*} \frac{(c, \tilde{\eta})}{\tilde{\eta}!} F_{\tilde{d}_{i_k} \sqcup\!\sqcup\, \tilde{d}_{i_{k-1}} \cdots \sqcup\!\sqcup\, \tilde{d}_{i_1}}[u] \\ &= \sum_{\tilde{\eta}\in\tilde{X}^*} \frac{(c, \tilde{\eta})}{\tilde{\eta}!} F_{\tilde{d}^{\sqcup\!\sqcup\, \tilde{\eta}}}[u] \\ &= \sum_{\tilde{\eta}\in\tilde{X}^*} \frac{(c, \tilde{\eta})}{\tilde{\eta}!} \left[\sum_{\eta\in X^*} (\tilde{d}^{\sqcup\!\sqcup\, \tilde{\eta}}, \eta) E_\eta[u] \right] \end{aligned}$$

$$= \sum_{\eta \in X^*} \left[\sum_{\tilde{\eta} \in \tilde{X}^*} (c, \tilde{\eta}) \frac{(\tilde{d} \sqcup\!\sqcup \tilde{\eta}, \eta)}{\tilde{\eta}!} \right] E_\eta[u]$$
$$= \sum_{\eta \in X^*} (c \circ d, \eta) E_\eta[u]$$
$$= F_{c \circ d}[u].$$

Recall that the properness of $\tilde{d}$ ensures that $c \circ d$ is a well defined series in $\mathbb{R}^\ell\langle\langle X \rangle\rangle$ (see Theorem 2.12). ■

Example 3.14 Consider a Wiener system where $f_c(z) = K/(1 - Mz)$ for some $K, M > 0$ and $F_d[u] = \sum_{i=0}^m E_{x_i}[u]$. This is a modest generalization of the system in Example 3.3. Here $\tilde{X} = \{\tilde{x}_1\}$ and $X = \{x_0, x_1, \ldots, x_m\}$ such that $c = \sum_{k \geq 0} KM^k k!\, \tilde{x}_1^k$, and d is the proper polynomial char(X). Then $f_c \circ F_d = F_{c \circ d}$ with

$$c \circ d = \sum_{\tilde{\eta} \in \tilde{X}^*} (c, \tilde{\eta}) \, \frac{d^{\sqcup\!\sqcup \tilde{\eta}}}{\tilde{\eta}!} = \sum_{k=0}^{\infty} KM^k k! \, \frac{(\mathrm{char}(X))^{\sqcup\!\sqcup k}}{k!}$$
$$= \sum_{k=0}^{\infty} KM^k k! \, \mathrm{char}(X^k) = \sum_{\eta \in X^*} KM^{|\eta|} \, |\eta|! \, \eta,$$

where the identity in Problem 2.4.6(b) has been used. Therefore, a locally maximal series is the generating series for the Wiener system $y = K/(1 - ME_{\mathrm{char}(X)}[u])$. □

Next, the generating series for the cascade connection of two Fliess operators is described.

Theorem 3.18 *Let $X = \{x_0, x_1, \ldots, x_m\}$ and $\tilde{X} = \{\tilde{x}_0, \tilde{x}_1, \ldots, \tilde{x}_{\tilde{m}}\}$. Given Fliess operators F_c and F_d, where $c \in \mathbb{R}^{\tilde{\ell}}_{LC}\langle\langle \tilde{X} \rangle\rangle$ and $d \in \mathbb{R}^{\tilde{m}}_{LC}\langle\langle X \rangle\rangle$, the cascade connection $F_c \circ F_d$ has the generating series in $\mathbb{R}^{\tilde{\ell}}\langle\langle X \rangle\rangle$*

$$c \circ d = \sum_{\tilde{\eta} \in \tilde{X}^*} (c, \tilde{\eta}) \, \psi_d(\tilde{\eta})(\mathbf{1}),$$

where ψ_d is the continuous (in the ultrametric sense) algebra homomorphism from $\mathbb{R}\langle\langle \tilde{X} \rangle\rangle$ into End$(\mathbb{R}\langle\langle X \rangle\rangle)$ *uniquely specified by*

$$\psi_d(\tilde{x}_i \tilde{\eta}) = \psi_d(\tilde{x}_i) \bullet \psi_d(\tilde{\eta}), \;\; \tilde{x}_i \in \tilde{X}, \;\; \tilde{\eta} \in \tilde{X}^*$$

using the family of mappings

$$\psi_d(\tilde{x}_i) : \mathbb{R}\langle\langle X\rangle\rangle \to \mathbb{R}\langle\langle X\rangle\rangle$$
$$: e \mapsto x_0(d_i \sqcup\!\sqcup\, e),$$

$i = 0, 1, \ldots, \tilde{m}$. Here $d_0 := \mathbf{1}$, and $\psi_d(\emptyset)$ denotes the identity map on $\mathbb{R}\langle\langle X\rangle\rangle$. That is, $F_c \circ F_d = F_{c\circ d}$.

Proof: It is first shown by induction on the length of the word $\tilde{\eta} \in \tilde{X}^*$ that $E_{\tilde{\eta}} \circ F_d = F_{\tilde{\eta}\circ d}$ for any $d \in \mathbb{R}^{\tilde{m}}\langle\langle X\rangle\rangle$. Trivially,

$$(E_\emptyset \circ F_d)[u] = E_\emptyset[F_d[u]] = E_\emptyset[u] = F_{\psi_d(\emptyset)(\mathbf{1})}[u].$$

Now assume that the claim holds for words $\tilde{\eta}$ up to length k. Then for any $\tilde{x}_i \in \tilde{X}$ observe that

$$\begin{aligned}
E_{\tilde{x}_i\tilde{\eta}}[F_d[u]](t, t_0) &= \int_{t_0}^{t} F_{d_i}[u](\tau)\, E_{\tilde{\eta}}[F_d[u]](\tau, t_0)\, d\tau \\
&= F_{x_0(d_i \sqcup\!\sqcup\, \psi_d(\tilde{\eta})(\mathbf{1}))}[u](t) \\
&= F_{\psi_d(\tilde{x}_i\tilde{\eta})(\mathbf{1})}[u](t).
\end{aligned}$$

Thus, the identity in question holds for every $\tilde{\eta} \in \tilde{X}^*$. Finally,

$$\begin{aligned}
(F_c \circ F_d)[u] &= \sum_{\tilde{\eta}\in\tilde{X}^*} (c, \tilde{\eta}) E_{\tilde{\eta}}[F_d[u]] = \sum_{\tilde{\eta}\in\tilde{X}^*} (c, \tilde{\eta}) F_{\psi_d(\tilde{\eta})(\mathbf{1})}[u] \\
&= \sum_{\tilde{\eta}\in\tilde{X}^*} (c, \tilde{\eta}) \left[\sum_{\nu\in X^*} (\psi_d(\tilde{\eta})(\mathbf{1}), \nu) E_\nu[u]\right] \\
&= \sum_{\nu\in X^*} \left[\sum_{\tilde{\eta}\in\tilde{X}^*} (c, \tilde{\eta})(\psi_d(\tilde{\eta})(\mathbf{1}), \nu)\right] E_\nu[u] \\
&= \sum_{\nu\in X^*} (c \circ d, \nu)\, E_\nu[u] \\
&= F_{c\circ d}[u].
\end{aligned}$$

■

Example 3.15 In the case of linear time-invariant systems, there are two approaches to describing the cascade connection of F_c and F_d. In the context of linear system theory, each system can be uniquely identified in terms of its impulse response, $h_c(t) = \sum_{i\geq 0}(c, x_0^i x_1)t^i/i!$ and $h_d(t) = \sum_{i\geq 0}(d, x_0^i x_1)t^i/i!$, respectively. The cascade connection is then characterized by the convolution product:

$$
\begin{aligned}
&(h_c * h_d)(t) \\
&\quad = \int_0^t h_c(t-\tau)h_d(\tau)\,d\tau \\
&\quad = \int_0^t \sum_{i,j=0}^{\infty} (c, x_0^i x_1)\frac{(t-\tau)^i}{i!}\,(d, x_0^j x_1)\frac{\tau^j}{j!}\,d\tau \\
&\quad = \sum_{i,j=0}^{\infty} (c, x_0^i x_1)(d, x_0^j x_1)\frac{1}{i!\,j!}\int_0^t (t-\tau)^i \tau^j\,d\tau \\
&\quad = \sum_{i,j=0}^{\infty} (c, x_0^i x_1)(d, x_0^j x_1)\frac{1}{i!\,j!}\sum_{k=0}^{i}\binom{i}{k}(-1)^k t^{i-k}\int_0^t \tau^{k+j}\,d\tau \\
&\quad = \sum_{i,j=0}^{\infty} (c, x_0^i x_1)(d, x_0^j x_1)\frac{t^{i+j+1}}{i!\,j!}\left[\sum_{k=0}^{i}\binom{i}{k}(-1)^k\frac{1}{k+j+1}\right] \\
&\quad = \sum_{i,j=0}^{\infty} (c, x_0^i x_1)(d, x_0^j x_1)\frac{t^{i+j+1}}{(i+j+1)!} \\
&\quad = \sum_{k=1}^{\infty}\left[\sum_{j=0}^{k-1}(c, x_0^{k-j-1}x_1)(d, x_0^j x_1)\right]\frac{t^k}{k!} \\
&\quad = \sum_{k=1}^{\infty}(c\circ d, x_0^k x_1)\frac{t^k}{k!} \\
&\quad = h_{c\circ d}(t),
\end{aligned}
$$

where the following identity for integer sequences has been used,

$$
\sum_{k=0}^{i}\binom{i}{k}(-1)^k\frac{1}{k+j+1} = \frac{i!j!}{(i+j+1)!},\quad i\geq 0 \tag{3.19}
$$

(see Problem 3.6.1), as well as the formula for the composition product of two linear series as computed in Example 2.33.

A second approach is to use Theorem 3.18 directly. In which case one first needs to determine the generating series for each subsystem. Following the discussion in Section 1.3, recall that

$$\begin{aligned}F_c[u](t) &= \int_0^t h_c(t-\tau)u(\tau)\,d\tau \\ &= \sum_{k=0}^{\infty}(c, x_0^k x_1)\int_0^t \frac{(t-\tau)^k}{k!}u(\tau)\,d\tau \\ &= \sum_{k=0}^{\infty}(c, x_0^k x_1)E_{x_0^k x_1}[u](t),\end{aligned}$$

and likewise for F_d. Thus, $F_c \circ F_d = F_{c\circ d}$, where the definition of the composition product gives

$$(c\circ d, x_0^k x_1) = \sum_{j=0}^{k-1}(c, x_0^{k-j-1}x_1)(d, x_0^j x_1),\ k \geq 0.$$

as the only nonzero coefficients of $c \circ d$. Therefore, reversing the steps above,

$$F_{c\circ d}[u](t) = \int_0^t h_{c\circ d}(t-\tau)u(\tau)\,d\tau$$

as expected. □

Example 3.16 Consider the cascade of F_c and F_d, where $c = \sum_{k\geq 0} k!\, x_1^k$ and $d = x_1$. Using the identity in Problem 2.7.7(a), it follows that

$$c\circ d = \sum_{k=0}^{\infty} k!\, x_1^k \circ x_1 = \sum_{k=0}^{\infty}(x_0x_1)^{\sqcup\!\sqcup\, k}.$$

Therefore,

$$\begin{aligned}F_{c\circ d}[u] &= \sum_{k=0}^{\infty} E_{(x_0x_1)^{\sqcup\!\sqcup\, k}}[u] = \sum_{k=0}^{\infty} E_{x_0x_1}^k[u] \\ &= \frac{1}{1 - E_{x_0x_1}[u]}.\end{aligned}$$

This can be viewed as another generalization of the Wiener system in Example 3.3, where the single integrator is replaced with a double integrator. (See also Problem 3.6.4.) □

Suppose the operator F_c is given with $c \in \mathbb{R}^{\ell}_{LC}\langle\langle X\rangle\rangle$, and one wants to explicitly compute the output $y = F_c[u]$ on some interval $[t_0, t_0+T]$ corresponding to a specific input u which is real analytic at $t = t_0$. In light of Theorem 3.6, y must be real analytic at $t = t_0$. So it suffices to compute its generating series, $c_y \in \mathbb{R}^{\ell}_{LC}[[X_0]]$, where $X_0 = \{x_0\}$. The theorem below provides an explicit formula for this series.

Theorem 3.19 *Consider an operator F_c with $c \in \mathbb{R}^{\ell}_{LC}\langle\langle X\rangle\rangle$. Select any input u which is real analytic at $t = t_0$ and has generating series $c_u \in \mathbb{R}^{m}_{LC}[[X_0]]$. Then the output function $y = F_c[u]$ is also real analytic at $t = t_0$ and has the generating series $c_y = c \circ c_u \in \mathbb{R}^{\ell}_{LC}[[X_0]]$.*

Proof: Only the claim regarding the generating series for c_y remains to be shown. Observe that for any admissible $v \in B^m_{\mathfrak{p}}(R)[t_0, t_0+T]$

$$F_{c_y}[v] = y = F_c[u] = F_c[F_{c_u}[v]] = F_{c\circ c_u}[v].$$

Applying Theorem 3.5 gives directly that $c_y = c \circ c_u$. Note, however, that the input v is just a *dummy* argument since both generating series c_u and c_y have no input letters x_i, $i \neq 0$, and thus, their corresponding Fliess operators do not depend on v. ■

In Section 3.8, the mapping $c_u \mapsto c \circ c_u$ will serve as the definition of a *formal* Fliess operator with generating series c. Since it has been established that this composition is always well defined (via Theorem 2.11), convergence assumptions play no role in this setting. But when they are available, the formal Fliess operator and the convergent Fliess operator coincide, which explains why the convergence assumptions made at the beginning of this section did not play a direct role in the algebraic analysis.

Example 3.17 Consider the casual linear integral operator

$$y(t) = \int_0^t h(t-\tau)u(\tau)\,d\tau,$$

where the kernel function h is real analytic at $t = 0$. Then $y = F_c[u]$ with $(c, x_0^k x_1) = h^{(k)}(0)$, $k \geq 0$ and zero otherwise. If $u(t) = \sum_{k\geq 0}(c_u, x_0^k)\, t^k/k!$, then it follows that $y(t) = \sum_{n\geq 0}(c_y, x_0^n)\, t^n/n!$, where

$$\begin{aligned} c_y = c \circ c_u &= \sum_{k=0}^{\infty} (c, x_0^k x_1)\, x_0^k x_1 \circ c_u \\ &= \sum_{k=0}^{\infty} (c, x_0^k x_1)\, x_0^{k+1} c_u. \end{aligned}$$

Therefore,

$$(c_y, x_0^n) = \sum_{k=0}^{n-1} (c, x_0^k x_1)\, (c_u, x_0^{n-1-k}), \;\; n \geq 1,$$

which is the same convolution sum produced in Examples 2.5, 2.33, and 3.15.
□

Example 3.18 Reconsider the Wiener system in Example 3.3, where it was shown that the input-output mapping $u \mapsto y = F_c[u]$ has the generating series $c = \sum_{k \geq 0} k!\, x_1^k$. If $u(t) = t^n/n!$, $t \geq 0$ then clearly $c_u = x_0^n$. From Theorem 3.19 and the identities in Problems 2.4.5(d) and 2.7.7(a) it follows that

$$c_y = \sum_{k=0}^{\infty} k!\, x_1^k \circ x_0^n = \sum_{k=0}^{\infty} (x_0^{n+1})^{\,\sqcup\!\sqcup\, k} = \sum_{k=0}^{\infty} \frac{((n+1)k)!}{((n+1)!)^k} x_0^{(n+1)k}.$$

Consequently, the output response is

$$y(t) = \sum_{k=0}^{\infty} \frac{((n+1)k)!}{((n+1)!)^k} \frac{t^{(n+1)k}}{((n+1)k)!} = \sum_{k=0}^{\infty} \frac{t^{(n+1)k}}{((n+1)!)^k} = \frac{1}{1 - \frac{t^{n+1}}{(n+1)!}}$$

on the interval $[0, ((n+1)!)^{1/n+1})$.
□

Next convergence properties are considered. For both cascade connections, it is claimed that local and global convergence are preserved in the sense that if each subsystem has a locally (globally) generating series then the Fliess operator representation of the cascaded system is locally (globally) convergent. The radius of convergence is also given for the case where the generating series for the component systems are only locally convergent. The proofs of these results are quite technical, so they will be left to the literature. In the global case, no claim will be made about the growth rate of the generating series of a composite system. It turns out that the composite system, as suggested by Example 3.8, can be globally convergent *without* its generating series being globally convergent as currently defined. This is not entirely unexpected as coefficient growth rates were never shown to be a *necessary* condition for any notion of operator convergence.

The main results for the Wiener-Fliess system are given first. As discussed in Section 1.1, for any analytic function $f:\mathbb{R}^{\tilde{m}}\rightarrow\mathbb{R}^{\tilde{\ell}}$ with generating series $c\in\mathbb{R}^{\tilde{\ell}}[[\tilde{X}]]$, the multivariable Cauchy integral formula provides that there exist real numbers $K_c, M_c > 0$ such that

$$|(c,\tilde{\eta})|\leq K_c M_c^{|\tilde{\eta}|}\,|\tilde{\eta}|!,\quad \tilde{\eta}\in\tilde{X}^*.$$

Analogous to the noncommutative case, the set of all locally convergent series in $\mathbb{R}^{\tilde{\ell}}[[\tilde{X}]]$ will be denoted by $\mathbb{R}^{\tilde{\ell}}_{LC}[[\tilde{X}]]$ and likewise for globally convergent series.

Theorem 3.20 *Suppose* $c\in\mathbb{R}^{\tilde{\ell}}_{LC}[[\tilde{X}]]/\mathbb{R}^{\tilde{\ell}}_{GC}[[\tilde{X}]]$ *and* $d\in\mathbb{R}^{\tilde{m}}_{LC}\langle\langle X\rangle\rangle/\mathbb{R}^{\tilde{m}}_{GC}\langle\langle X\rangle\rangle$ *with growth constants* $K_c, M_c>0$ *and* $K_d, M_d>0$*, respectively. If* $b=c\circ d$*, then for every* $\epsilon>0$

$$|(b,\nu)|\leq K_b(M_b+\epsilon)^{|\nu|}|\nu|!,\quad \nu\in X^* \tag{3.20}$$

for some $K_b>0$*, where*

$$M_b=(1+\tilde{m}M_cK_d)M_d.$$

Furthermore, if M_c*,* K_d *and* M_d *are minimal, then no geometric growth constant smaller than* M_b *can satisfy (3.20), and thus, the radius of convergence of* $b=c\circ d$ *is*

$$\frac{1}{(1+\tilde{m}M_cK_d)M_d(m+1)}.$$

Example 3.19 Reconsider a Wiener system in Example 3.14, where $f_c(z)=1/(1-z)$ and $F_d[u]=\sum_{i=0}^{m}E_{x_i}[u]$. Clearly, c is a locally maximal series with $K_c=M_c=1$. On the other hand, d is locally convergent, but has no coefficient growth after a certain point. The growth constants K_d, M_d can therefore be selected in any manner such that $K_dM_d=1$. In this case, Theorem 3.20 gives $M_b=1+\epsilon$, $\epsilon>0$, which is consistent with the series $c\circ d$ computed in this example.

□

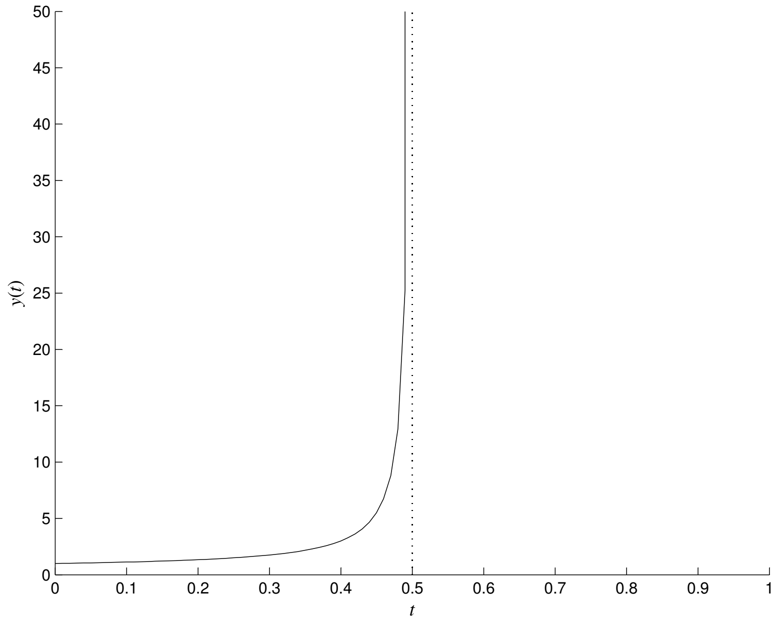


Fig. 3.12. Zero-input response of the cascade system $F_{c \circ d}$ in Example 3.20.

Example 3.20 Consider a Wiener-Fliess system where $\tilde{X} = \{\tilde{x}_1\}$, $X = \{x_0, x_1\}$, $c = \sum_{k \geq 0} k!\, \tilde{x}_1^k$, and $d = \sum_{\eta \in X^*} |\eta|!\, \eta$. Then $f_c(z) = 1/(1-(z-1)) = 1/(2-z)$ (since $(d, \emptyset) = 1$), and from Example 3.14 $F_d[u] = 1/(1 - E_{x_0+x_1}[u])$. The composite system is therefore

$$F_{c \circ d} = f_c(F_d[u] - 1) = \frac{1 - E_{x_0+x_1}[u]}{1 - 2E_{x_0+x_1}[u]}.$$

The zero-input response is clearly $y = (1-t)/(1-2t)$ as shown in Figure 3.12. The presence of a finite escape time at $t_{esc} = 0.5$ implies that $M_b = 1/t_{esc} = 2$. This is consistent with the radius of convergence given in Theorem 3.20 with $m = 0$ since both c and d are locally maximal series for the class of series with the growth rate corresponding to $K_c = M_c = K_d = M_d = 1$.

As a side note, the existence of a finite escape time could have been predicted as a result of what is called *Pringsheim's Theorem*. Let $y(z) = \sum_{n \geq 0} b_n z^n / n!$ be a function that is analytic at the origin of the complex plane with a radius of convergence $R > 0$. The theorem states that if each $b_n \geq 0$, then the point $z = R$ is a *real* singularity of y. Such a condition applies for any cascade of two maximal series. Therefore, when y is restricted to $\mathbb{R}$, it must have a finite escape time. Contrast this behavior to that of $y(t) = 1/(1+t^2)$, which has no real singularities (recall Problem 1.1.1). In this case, y has the Taylor series $y(t) = \sum_{k\geq 0}(-t^2)^k$ at $t = 0$ with a finite radius $R = 1$. But clearly Pringsheim's Theorem does not apply. □

The Wiener-Fliess system comprised of two globally convergent subsystems is considered next.

Theorem 3.21 *A Wiener-Fliess system where each subsystem has a generating series satisfying the global growth rate (3.11) has a radius of convergence equal to infinity. Hence, the output of such a system is always well defined over any finite interval of time when its input* $u \in L_{1,e}^m(t_0)$.

It was pointed out in Example 3.8 that the growth condition (3.11) is not a necessary condition for a Fliess operator to converge globally. The following example illustrates how this issue arises for a Wiener-Fliess system.

Example 3.21 Consider a Wiener-Fliess system where $\tilde{X} = \{\tilde{x}_1\}$, $X = \{x_0, x_1\}$, $c = \sum_{k\geq 0} K_c M_c^k \, \tilde{x}_1^k$, and $d = \sum_{\eta\in X^*} K_d M_d^{|\eta|}\eta$. A calculation analogous to the one given in Example 3.14 yields $f_c(z) = K_c \exp(M_c z)$ and $F_d[u] = K_d \exp\left(M_d E_{x_0+x_1}[u]\right)$. Therefore, the cascaded system is

$$y = F_{c\circ d}[u] = K_c \exp(M_c K_d \exp(M_d E_{x_0+x_1}[u])),$$

so that the zero-input response when $K_c = \mathrm{e}^{-1}$ and $M_c = K_d = M_d = 1$ is $y(t) = \mathrm{e}^{\mathrm{e}^t - 1}$. That is, $(c\circ d, x_0^n) = B_n$, $n \geq 0$, where B_n, $n \geq 0$, are the Bell numbers as described in Example 3.8. Thus, $c \circ d$ can not have Gevrey order less one, i.e., $s^* = 1$, even though by Theorem 3.21 the operator $F_{c\circ d}$ must be globally convergent. □

An interesting fact about the Bell numbers is that their asymptotic behavior is described by

$$B_n \sim n^{-\frac{1}{2}}(\lambda(n))^{n+\frac{1}{2}}\mathrm{e}^{\lambda(n)-n-1},$$

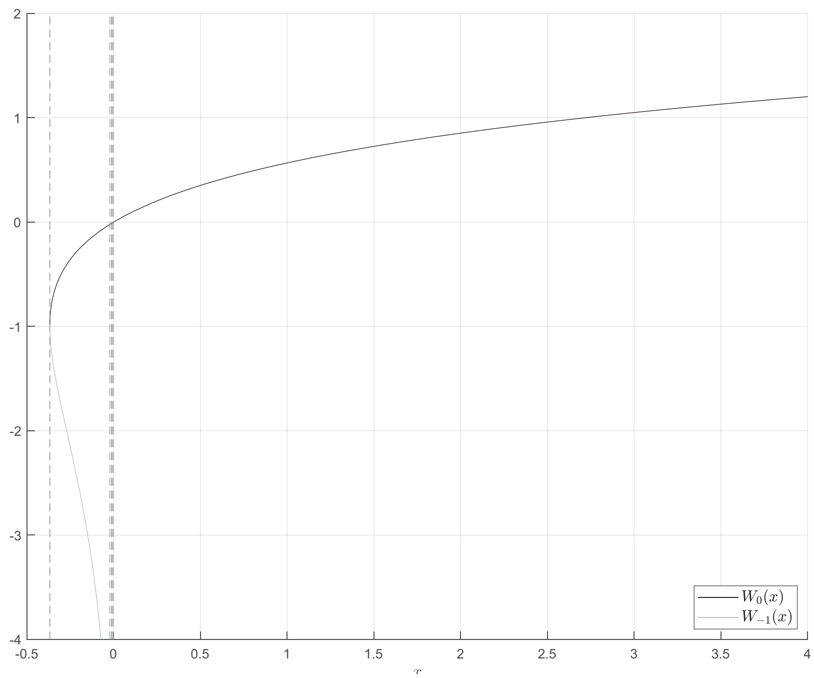


Fig. 3.13. Branches $W_0(x)$ and $W_{-1}(x)$ of the Lambert W-function.

where $\lambda(n) = n/W(n)$, and $W(n)$ denotes the *Lambert W-function*. The Lambert W-function is a multivalued function defined by the branches W_k, $k \in \mathbb{Z}$ of the inverse relation of the function

$$g(z) = z\exp(z) \;\; z \in \mathbb{C}.$$

Here $W(x) := W_0(x)$ denotes the *principal branch* as shown in Figure 3.13 along with the branch $W_{-1}(x)$. This function often appears explicitly whenever cascade structures are present. This will be the case, for example, when two Fliess operators are cascaded as addressed next.

Consider next the interconnection of two Fliess operators whose generating series are only locally convergent. The theorem below states that the cascade system is also only locally convergent, and the radius of convergence is given.

Theorem 3.22 *Suppose* $c \in \mathbb{R}^{\tilde{\ell}}_{LC}\langle\langle \tilde{X}\rangle\rangle/\mathbb{R}^{\tilde{\ell}}_{GC}\langle\langle \tilde{X}\rangle\rangle$ *and* $d \in \mathbb{R}^{\tilde{m}}_{LC}\langle\langle X\rangle\rangle/\mathbb{R}^{\tilde{m}}_{GC}\langle\langle X\rangle\rangle$ *with growth constants* $K_c, M_c > 0$ *and* $K_d, M_d > 0$*, respec-*

tively. If $b = c \circ d$, then for every $\epsilon > 0$

$$|(b,\nu)| \leq K_b(M_b+\epsilon)^{|\nu|}|\nu|!, \quad \nu \in X^* \tag{3.21}$$

for some $K_b > 0$, where

$$M_b = \frac{M_d}{1-\tilde{m}K_d W\left(\frac{1}{\tilde{m}K_d}\exp\left(\frac{M_c-M_d}{\tilde{m}M_cK_d}\right)\right)}.$$

Furthermore, if M_c, K_d and M_d are minimal, then no smaller geometric growth constant can satisfy (3.21), and thus, the radius of convergence for $b = c \circ d$ is

$$\frac{1}{M_d(m+1)}\left[1-\tilde{m}K_d W\left(\frac{1}{\tilde{m}K_d}\exp\left(\frac{M_c-M_d}{\tilde{m}M_cK_d}\right)\right)\right].$$

Example 3.22 *Let $X = \{x_0, x_1\}$ and $c, d \in \mathbb{R}\langle\langle X\rangle\rangle$ such that $M = M_c = M_d$. Using a series expansion about $K_d = \infty$ it follows that*

$$\begin{aligned} M_b &= \frac{M}{1-K_dW\left(1/K_d\right)} \\ &= \left(\frac{3}{2}+K_d+O\left(\frac{1}{K_d}\right)\right)M \\ &\approx K_dM \end{aligned}$$

when $K_d \gg 1$. On the other hand, if $K_d = 1$ then $M_b = (1 - W(1))^{-1}M = 2.3102M$. □

Example 3.23 Consider the linear series $c = \sum_{n\geq 0}(c, x_0^n x_1)\, x_0^n x_1$ and $d = \sum_{n\geq 0}(d, x_0^n x_1)\, x_0^n x_1$ in $\mathbb{R}_{LC}\langle\langle X\rangle\rangle$ with growth constants $K_c, M_c > 0$ and $K_d, M_d > 0$, respectively. From the calculation in Example 3.15, it is apparent that

$$\begin{aligned} \left|(c\circ d, x_0^k x_1)\right| &= \left|\sum_{j=0}^{k-1}(c, x_0^{k-1-j}x_1)(d, x_0^j x_1)\right| \\ &\leq \sum_{j=0}^{k-1}(K_cM_c^{k-j}(k-j)!)(K_dM_d^{j+1}(j+1)!) \\ &= K_cK_dM^{k+1}\left[\sum_{j=0}^{k-1}\binom{k+1}{j+1}^{-1}\right](k+1)! \end{aligned}$$

$$= K_c K_d M^{k+1} \left[\sum_{j=1}^{k} \binom{k+1}{j}^{-1} \right] (k+1)!,$$

where $M = \max\{M_c, M_d\}$ and assuming the convention $\sum_{i=j}^{k} a_i = 0$ when $k < j$. Applying the combinatorial inequality $\sum_{k=1}^{n-1} \binom{n}{k}^{-1} < 1$, $n \geq 2$ (see Problem 3.6.6), it follows directly that

$$|(c \circ d, \nu)| < K_c K_d M^{|\nu|} |\nu|!, \ \ \nu \in X^*.$$

For the case where $K_c = K_d = 1$ and $M_c = M_d$, it evident that $M_b = M$, which is an improvement over the more general case described in the previous example. That is, using the specific structure of c and d, a smaller geometric growth constant can be determined as compared to the general case where only the growth constants are known. □

Example 3.24 Let $X = \{x_0, x_1\}$, and suppose $c = \sum_{n>0} (n!)^2 \, x_1^n$. Then according to Lemma 2.5, $c \circ 0 = 0$ and $1 \circ c = 1$. That is, it is possible that $c \circ d$ can be locally convergent even when c or d is not.

□

It was shown in Theorem 3.19 that the composition product can be used to determine the coefficients of an output function produced by a Fliess operator with an analytic input. The following corollary of Theorem 3.22 describes a lower bound for the interval of convergence for such an output function.

Corollary 3.3 *Let $X = \{x_0, x_1, \ldots, x_m\}$ and $X_0 = \{x_0\}$. Suppose $c \in \mathbb{R}^{\ell}_{LC}\langle\langle X \rangle\rangle / \mathbb{R}^{\ell}_{GC}\langle\langle X \rangle\rangle$ with growth constants $K_c, M_c > 0$ and $c_u \in \mathbb{R}^{m}_{LC}[[X_0]] / \mathbb{R}^{m}_{GC}[[X_0]]$ with growth constants K_{c_u}, M_{c_u}, respectively. If $c_y = c \circ c_u$, then for every $\epsilon > 0$*

$$|(c_y, x_0^k)| \leq K_{c_y} (M_{c_y} + \epsilon)^k k!, \ k \geq 0 \tag{3.22}$$

for some $K_{c_y} > 0$, where

$$M_{c_y} = \frac{M_{c_u}}{\left[1 - m K_{c_u} W\left(\frac{1}{m K_{c_u}} \exp\left(\frac{M_c - M_{c_u}}{m M_c K_{c_u}} \right) \right) \right]}.$$

Furthermore, if M_c, K_{c_u} and M_{c_u} are minimal, then no smaller geometric growth constant can satisfy (3.22). Thus, the interval of convergence for the output $y = F_{c_y}[u]$ is at least as large as $T = 1/M_{c_y}$.

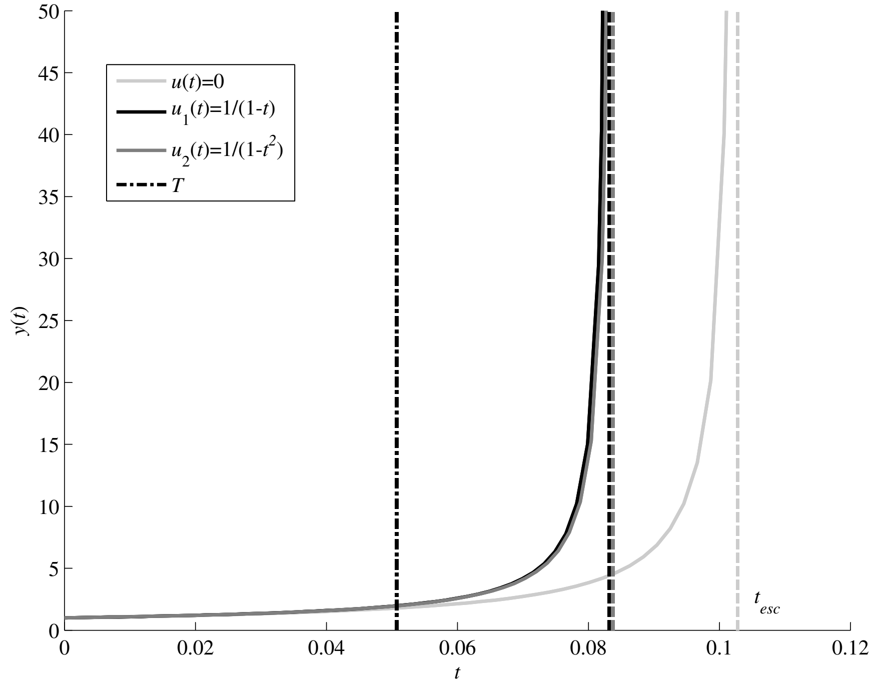


Fig. 3.14. Output responses of the cascaded system $F_{c \circ d}$ to various analytic inputs in Example 3.25

Example 3.25 Suppose $X = \{x_0, x_1\}$ and $b = c \circ d$ with $c = \sum_{\eta \in X^*} K_c M_c^{|\eta|} |\eta|!\, \eta$ and $d = \sum_{\eta \in X^*} K_d M_d^{|\eta|} |\eta|!\, \eta$. The output of the cascade system is $y = F_{c \circ d}[u] = F_c[F_d[u]]$, where $F_c[u] = K_c/(1 - M_c E_{x_0+x_1}[u])$ and $F_d[u] = K_d/(1 - M_d E_{x_0+x_1}[u])$. The zero-input response is therefore,

$$y(t) = \frac{K_c}{1 - M_c E_{x_0+x_1}\left[\frac{K_d}{1-M_d t}\right]}$$

as shown in Figure 3.14 when $K_c = 1$, $M_c = 2$, $K_d = 3$ and $M_d = 4$. Applying Theorem 3.22 with $m = 0$ and $\tilde{m} = 1$ gives $M_b = 9.7284$ so that the finite escape time of the response must be $t_{esc} = 1/M_b = 0.1028$, which is what is observed. The output responses corresponding to the analytic inputs $u_1(t) = 1/1-t$ and $u_2(t) = 1/1-t^2$, each having growth constants $K_{c_u} = M_{c_u} = 1$, are also shown in the figure. Their respective finite escape times are 0.08321 and 0.08377. Here u_1 has the shortest escape time since its generating series

$$c_{u_1} = \sum_{k=0}^{\infty} k!\, x_0^k$$

has all its coefficients growing at the maximum rate. Whereas

$$c_{u_2} = \sum_{k=0}^{\infty} (2k)!\, x_0^{2k}$$

has all its odd coefficients equal to zero. Setting $M_c = M_b = 9.7284$ in Corollary 3.3, any finite escape time for the output corresponding to any analytic input with the given growth constants K_{c_u}, M_{c_u} must be at least as large as $T = 1/M_{c_y} = 0.0507$, which is evident from the simulations. □

The section is concluded by considering the cascade connection of two Fliess operators whose generating series are globally convergent. The following theorem is the main result.

Theorem 3.23 *The cascade connection of Fliess operators each having a generating series satisfying the global growth rate (3.11) has a radius of convergence equal to infinity. Hence, the output of such a system is always well defined over any finite interval of time when its input* $u \in L_{1,e}^m(t_0)$.

Example 3.26 Reconsider the linear series $c = \sum_{n\geq 0}(c, x_0^n x_1)x_0^n x_1$ and $d = \sum_{n\geq 0}(d, x_0^n x_1)x_0^n x_1$ in Example 3.23, except here both $c, d \in \mathbb{R}_{GC}\langle\langle X \rangle\rangle$ with $s^* = 0$. In which case,

$$\begin{aligned} \left|(c \circ d, x_0^k x_1)\right| &= \left|\sum_{j=0}^{k-1} (c, x_0^{k-1-j} x_1)(d, x_0^j x_1)\right| \\ &\leq \sum_{j=0}^{k-1} (K_c M_c^{k-j})(K_d M_d^{j+1}) \\ &= K_c K_d M^{k+1} k \\ &< K_c K_d (2M)^{k+1}. \end{aligned}$$

Therefore, global convergence is preserved, and $s^*_{c\circ d} = 0$. □

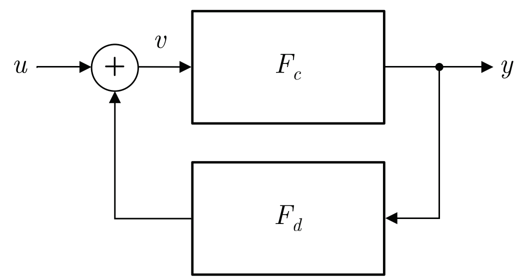


Fig. 3.15. Feedback connection of two Fliess operators

Example 3.27 Suppose $X = \{x_0, x_1\}$ and $c = d = \sum_{k\geq 0} x_1^k$. The output of the cascade system is exactly that of the system considered in Example 3.8, namely,

$$y(t) = F_{c\circ d}[u](t) = \exp(E_{x_1}[\exp(E_{x_1}[u(t)])]).$$

Therefore, when $u(t) = 1$

$$y(t) = \mathrm{e}^{\mathrm{e}^t - 1} = \sum_{n=0}^{\infty} B_n \frac{t^n}{n!}.$$

So the Bell numbers also appear from a simple cascade of two Fliess operators. □

3.7 Feedback Connections

In this section, the feedback connection of two Fliess operators as shown in Figure 3.15 is considered. Such closed-loop systems appear frequently in control engineering. As with cascade connections, one could also replace F_d in the feedback path with a static function f_d. But the focus here will be on the former case. The latter can be found in the literature. Similar to the interconnections in the previous sections, five basic questions must be addressed. But now the analysis is considerably more difficult because feedback is generally described only in implicit terms. That is, given any $c, d \in \mathbb{R}^m_{LC}\langle\langle X\rangle\rangle$ with $X = \{x_0, x_1, \ldots, x_m\}$, the output y of the corresponding closed-loop system must satisfy the feedback equation

$$y = F_c[u + F_d[y]] \tag{3.23}$$

for any admissible input u. The interconnection is well-posed when y is an admissible input for F_d, and $u+F_d[y]$ is an admissible input for F_c. Fortunately, this issue can be handled in much the same manner as it was for cascade connections (see Problem 3.7.1). When there exists a locally generating series $e \in \mathbb{R}^m_{LC}\langle\langle X\rangle\rangle$ so that $y = F_e[u]$, the feedback equation becomes equivalent to

$$F_e[u] = F_c[u + F_{d\circ e}[u]], \tag{3.24}$$

and the *output feedback product* of c and d, denoted by $c@d$, is defined to be e. The first obstacle in the analysis is that F_e is required to be the composition of two operators, F_c and $I + F_{d\circ e}$ as shown in Figure 3.16, where one of the operators is *not* a Fliess operator due to the presence of identity operator I acting as a *direct feed* term. Here $I + F_{d\circ e}$ will be referred to as a *unital* Fliess operator in order to make this distinction. The corresponding set of all unital Chen-Fliess series is denoted by $I + \mathscr{F} = \{I + F_c : c \in \mathbb{R}^m\langle\langle X\rangle\rangle\}$. The central claim is that this mixed composition always renders another Chen-Fliess series. However, none of the composition products introduced so far describe this type of composition. To address the issue, it is first convenient to introduce the symbol δ as the (fictitious) generating series for the identity map. That is, $F_\delta := I$ such that $I + F_c := F_{\delta+c} = F_{c_\delta}$ with $c_\delta := \delta + c$. The set of all such generating series for $I + \mathscr{F}$ will be denoted by $\delta + \mathbb{R}^m\langle\langle X\rangle\rangle$.[9] The following theorem describes the generating series for this new type of composition in terms of what will be called the *mixed composition product*. (See Table 3.1 for a summary of all the series compositions encountered in this section.)

Theorem 3.24 *Let $X = \{x_0, x_1, \ldots, x_m\}$. Given a Fliess operator F_c and unital Fliess operator F_{d_δ}, where $c \in \mathbb{R}^\ell_{LC}\langle\langle X\rangle\rangle$ and $d_\delta \in \delta + \mathbb{R}^m_{LC}\langle\langle X\rangle\rangle$, the cascade connection $F_c \circ F_{d_\delta}$ has the generating series in $\mathbb{R}^\ell\langle\langle X\rangle\rangle$*

$$c \,\tilde{\circ}\, d_\delta = \sum_{\eta\in X^*} (c, \eta)\, \phi_d(\eta)(\mathbf{1}), \tag{3.25}$$

where ϕ_d is the continuous (in the ultrametric sense) algebra homomorphism from $\mathbb{R}\langle\langle X\rangle\rangle$ into $\mathrm{End}(\mathbb{R}\langle\langle X\rangle\rangle)$ *uniquely specified by*

[9] A suitable subscript like 'LC' will be added when the set is restricted to series satisfying a certain growth condition.

Table 3.1. Composition products involving c, d, $c_\delta = \delta + c$, $d_\delta = \delta + d$ when $X = \{x_0, x_1, \ldots, x_m\}$

Name	Symbol	Map
composition	$c \circ d$	$\mathbb{R}^\ell\langle\langle X\rangle\rangle \times \mathbb{R}^m\langle\langle X\rangle\rangle \to \mathbb{R}^\ell\langle\langle X\rangle\rangle$
mixed composition	$c \,\tilde{\circ}\, d_\delta$	$\mathbb{R}^\ell\langle\langle X\rangle\rangle \times \delta + \mathbb{R}^m\langle\langle X\rangle\rangle \to \mathbb{R}^\ell\langle\langle X\rangle\rangle$
group composition	$c \circledcirc d$	$\mathbb{R}^m\langle\langle X\rangle\rangle \times \mathbb{R}^m\langle\langle X\rangle\rangle \to \mathbb{R}^m\langle\langle X\rangle\rangle$
group product	$c_\delta \circ d_\delta$	$\delta + \mathbb{R}^m\langle\langle X\rangle\rangle \times \delta + \mathbb{R}^m\langle\langle X\rangle\rangle \to \delta + \mathbb{R}^m\langle\langle X\rangle\rangle$

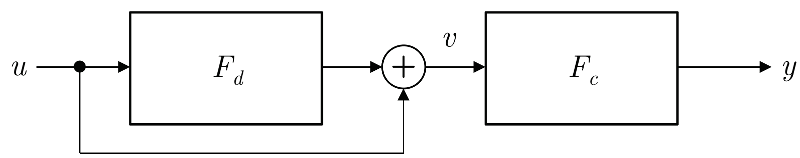


Fig. 3.16. Mixed composition of two Fliess operators

$$\phi_d(x_i\eta) = \phi_d(x_i) \bullet \phi_d(\eta), \quad x_i \in X, \quad \eta \in X^*$$

using the family of mappings

$$\phi_d(x_i)(e) = x_i e + x_0(d_i \sqcup\!\sqcup\, e),$$

$i = 0, 1, \ldots, m$. *Here* $d_0 := 0$, *and* $\phi_d(\emptyset)$ *denotes the identity map on* $\mathbb{R}\langle\langle X\rangle\rangle$. *That is,* $F_c \circ F_{d_\delta} = F_{c \,\tilde{\circ}\, d_\delta}$.

Proof: Observe that the mixed composition product is identical to that used for the usual composition product in Theorem 3.18 except for the extra leading term $x_i e$ in each operator $\phi_d(x_i)$. It is precisely this term that implements the direct feed component I. In which case, the proof is very similar to previous proof modulo an insertion of this extra term in each step (see Problem 3.7.4). ■

It can be verified in a manner completely analogous to the regular composition product for Fliess operators that the mixed composition product is always well defined (summable) and ultrametric continuous in both arguments. More advanced properties of this product will be described shortly, but of particular importance here is the fact that the feedback equation (3.24) can be written in terms of the mixed composition product as

$$F_e[u] = F_{c\tilde{\circ}(d\circ e)_\delta}[u].$$

In light of the uniqueness of generating series (Theorem 3.5 for the locally convergent case and Theorem 3.38 for the formal case) this implies that

$$e = c\tilde{\circ}(d \circ e)_\delta. \tag{3.26}$$

This equation suggests the possibility of describing e as a fixed point of a contractive iterated mapping. Consider the following theorem.

Theorem 3.25 *For any $c \in \mathbb{R}^m\langle\langle X\rangle\rangle$, the mapping $d_\delta \mapsto c\,\tilde{\circ}\,d_\delta$ is an ultrametric contraction on $\delta + \mathbb{R}^m\langle\langle X\rangle\rangle$.*

Proof: The proof is a minor variation of the previous result for the regular composition product of two Fliess operators, i.e., Theorem 2.15. The contraction coefficient, σ, is unaffected by the required modifications (see Problem 3.7.4). ■

The first main result of this section is given below and addresses question 2.

Theorem 3.26 *Let $c, d \in \mathbb{R}^m\langle\langle X\rangle\rangle$. Then the following propositions hold:*

1. *The mapping*

$$\begin{aligned} \mathcal{S} \;&:\; \mathbb{R}^m\langle\langle X\rangle\rangle \to \mathbb{R}^m\langle\langle X\rangle\rangle \\ &:\; e_i \mapsto e_{i+1} = c\,\tilde{\circ}\,(d \circ e_i)_\delta \end{aligned} \tag{3.27}$$

   has a unique fixed point in $\mathbb{R}^m\langle\langle X\rangle\rangle$, $c@d := \lim_{i\to\infty} e_i$, which is independent of e_0.
2. *The generating series $c@d$ satisfies the feedback equation (3.26).*

Proof:
1. The mapping $\mathcal{S}$ is a contraction on $\mathbb{R}^m\langle\langle X\rangle\rangle$ since by Theorems 2.15 and 3.25:

$$\mathrm{dist}(S(e_i), S(e_j)) \leq \sigma\,\mathrm{dist}((d \circ e_i)_\delta, (d \circ e_j)_\delta) \leq \sigma^2\,\mathrm{dist}(e_i, e_j).$$

Therefore, the mapping S has a unique fixed point, $c@d$, that is independent of e_0, i.e.,

$$c@d = c\,\tilde{\circ}\,(d \circ (c@d))_\delta. \tag{3.28}$$

2. The claim follows directly from comparing (3.26) and (3.28). ■

Now that it has been established that the closed-loop system in Figure 3.15 has a Chen-Fliess series representation, the next question is how to actually compute its generating series. Equivalently, how can (3.26) be solved to determine $e = c@d$? Observe that the function v in Figure 3.15 must satisfy the identity

$$v = u + F_{d\circ c}[v].$$

Therefore,

$$(I + F_{-d\circ c})\,[v] = u.$$

Now suppose there exists a series $(-d\circ c)^{-1}$ such that

$$\big(I + F_{(-d\circ c)^{-1}}\big) \circ \big(I + F_{(-d\circ c)}\big) = I.$$

Then it would follow that

$$v = \big(I + F_{(-d\circ c)^{-1}}\big)\,[u],$$

and thus,

$$\begin{aligned} F_{c@d}[u] &= F_c[v] = F_c[\big(I + F_{(-d\circ c)^{-1}}\big)\,[u]] \\ &= F_{c\,\tilde{\circ}\,(-d\circ c)_\delta^{-1}}[u]. \end{aligned}$$

In which case, the feedback product can be written in the form $c@d = c\,\tilde{\circ}\,(-d\circ c)_\delta^{-1}$ provided this inverse series can be determined. This suggests that two issues need to be investigated. First, the presence of an inverse implies that some group is involved in this calculation. What is this underlying group? Second, the group element $(-d\circ c)_\delta^{-1}$ is interacting with the generating series c describing the forward path via the mixed composition product. What is the exact nature of this interaction? This latter question motivates the following definition.

Definition 3.7 *Let G be a group and S a given set. Then G is said to act as a* ***transformation group*** *on the right of S if there exists a mapping $\mathcal{A} : S \times G \to S : (h, g) \mapsto hg$ such that:*

i. $h1 = h$, *where* 1 *is the identity element of* G*;*

ii. $h(g_1g_2) = (hg_1)g_2$ *for all* $g_1, g_2 \in G$.

The action $\mathcal{A}$ is said to be ***free*** *if $hg = h$ implies that $g = 1$.*

Example 3.28 Suppose the two Fliess operators F_c and F_d in Figure 3.15 are linear time-invariant systems with $m \times m$ transfer matrices H and G, respectively. In this case, the corresponding feedback equation

$$H_{cl} = H(I + GH_{cl}) \tag{3.29}$$

can be solved directly by substitution

$$\begin{aligned} H_{cl} &= H(I + G[H(I + GH_{cl})]) \\ &= H(I + GH + GHG[H(I + GH_{cl})]) \\ &\vdots \\ &= H\sum_{k=0}^{\infty}(GH)^k \\ &= H(I - GH)^{-1}. \end{aligned} \tag{3.30}$$

One can verify directly that the set of transfer functions $\{I+G\}$, where G is an $m \times m$ matrix of strictly proper rational functions, is a group under the product

$$(I + G_1)(I + G_2) = I + G_1 + G_2 + G_1G_2$$

with identity element I and inverse $(I+G)^{-1} = \sum_{k\geq 0}(-G)^k$. In which case, (3.30) can be interpreted as this group *acting* on the operator H from the right to produce the transfer function for the closed-loop system, H_{cl} (see Problem 3.7.2). One subtle point is that the inverse of $(I - GH)^{-1}$ is clearly $(I - GH)$. So applying this transformation will remove the feedback, but it is not clear what, if any, system interconnection this second action corresponds to. Nevertheless, there *is* a system interconnection that will remove the feedback loop. This issue will be revisited a bit later in the section. □

Moving on to the more general setting, the key idea is that $(I + \mathscr{F}, \circ, I)$ forms a group under the composition

$$\begin{aligned} F_{c_\delta} \circ F_{d_\delta} &= (I + F_c) \circ (I + F_d) \\ &= I + F_d + F_c \circ (I + F_d) \\ &= I + F_d + F_{c\,\tilde{\circ}\,d_\delta} \\ &= F_{c_\delta \circ d_\delta}, \end{aligned}$$

where

$$c_\delta \circ d_\delta := \delta + d + (c \,\tilde{\circ}\, d_\delta) =: \delta + c \circledcirc d. \tag{3.31}$$

Note that the same symbol will be used for composition on $\mathbb{R}^m\langle\langle X\rangle\rangle$ and $\delta + \mathbb{R}^m\langle\langle X\rangle\rangle$. As elements in these two sets have a distinct notation, i.e., c versus c_δ, respectively, it will always be clear which product is at play. Given the uniqueness of generating series for Chen-Fliess series, $(I + \mathscr{F}, \circ, I)$ is a group if and only if $(\delta + \mathbb{R}^m\langle\langle X\rangle\rangle, \circ, \delta)$ is a group. The main proposition is that this latter group acts as a right transformation group on $\mathbb{R}^m\langle\langle X\rangle\rangle$ via the mixed composition product to give the generating series for the closed-loop system under output feedback. Before any of these claims can be rigorously verified, additional properties of the mixed composition product and the composition product are needed.

Lemma 3.4 *Let $X = \{x_0, x_1, \ldots, x_m\}$. The mixed composition product (3.25) has the following properties:*

1. *left $\mathbb{R}$-linearity;*
2. $c \,\tilde{\circ}\, 0_\delta = c$;
3. $c \,\tilde{\circ}\, d_\delta = k\mathbf{1}$, $k \in \mathbb{R}^\ell$ *for any fixed* d_δ *if and only if* $c = k\mathbf{1}$;
4. $(x_i c) \,\tilde{\circ}\, d_\delta = x_i(c \,\tilde{\circ}\, d_\delta) + x_0(d_i \sqcup\!\sqcup (c \,\tilde{\circ}\, d_\delta))$ *for all* $x_i \in X$;
5. $(c \sqcup\!\sqcup d) \,\tilde{\circ}\, e_\delta = (c \,\tilde{\circ}\, e_\delta) \sqcup\!\sqcup (d \,\tilde{\circ}\, e_\delta)$;
6. $(c \circ d) \,\tilde{\circ}\, e_\delta = c \circ (d \,\tilde{\circ}\, e_\delta)$;
7. $(c \,\tilde{\circ}\, d_\delta) \,\tilde{\circ}\, e_\delta = c \,\tilde{\circ}\, (d \,\tilde{\circ}\, e_\delta + e)_\delta$,

where c, d, and e are suitably compatible formal power series over X.

Proof:
1. This fact follows directly from the definition of the mixed composition product.
2. The claim is immediate since $\phi_0(\eta)(\mathbf{1}) = \eta$ for all $\eta \in X^*$.
3. The only nontrivial assertion is that $c \,\tilde{\circ}\, d_\delta = k$ implies $c = k$. This claim is best handled later once the Hopf algebra context is developed (see page 189).
4. Observe

$$\begin{aligned}
(x_0 c) \,\tilde{\circ}\, d_\delta &= \phi_d(x_0 c)(\mathbf{1}) = \phi_d(x_0) \bullet \phi_d(c)(\mathbf{1}) \\
&= x_0(c \,\tilde{\circ}\, d_\delta) \\
(x_i c) \,\tilde{\circ}\, d_\delta &= \phi_d(x_i c)(\mathbf{1}) = \phi_d(x_i) \bullet \phi_d(c)(\mathbf{1}) \\
&= x_0(c \,\tilde{\circ}\, d_\delta)) + x_0(d_i \sqcup\!\sqcup (c \,\tilde{\circ}\, d_\delta)),
\end{aligned}$$

$i = 1, 2, \ldots, m$.
5. For any $e_\delta \in \delta + \mathbb{R}\langle\langle X \rangle\rangle$, one can define a shuffle product on $\mathrm{End}(\mathbb{R}\langle\langle X \rangle\rangle)$ via

$$\begin{aligned}\phi_e(x_i\eta) \sqcup\!\sqcup\, \phi_e(x_j\xi) = \phi_e(x_i) \bullet [\phi_e(\eta) \sqcup\!\sqcup\, \phi_e(x_j\xi)] + \\ \phi_e(x_j) \bullet [\phi_e(x_i\eta) \sqcup\!\sqcup\, \phi_e(\xi)].\end{aligned}$$

In which case, ϕ_e acts as an algebra map between the shuffle algebra on $\mathbb{R}\langle\langle X \rangle\rangle$ and the shuffle algebra on $\mathrm{End}(\mathbb{R}\langle\langle X \rangle\rangle)$. That is, $\phi_e(c \sqcup\!\sqcup\, d) = \phi_e(c) \sqcup\!\sqcup\, \phi_e(d)$. Hence, $(c \sqcup\!\sqcup\, d) \,\tilde{\circ}\, e_\delta = \phi_e(c \sqcup\!\sqcup\, d)(\mathbf{1}) = \phi_e(c)(\mathbf{1}) \sqcup\!\sqcup\, \phi_e(d)(\mathbf{1})$ $= (c \,\tilde{\circ}\, e_\delta) \sqcup\!\sqcup\, (d \,\tilde{\circ}\, e_\delta)$ (cf. Problem 3.7.5).
6. See Problems 3.7.4 and 3.7.5.
7. This identity has another interpretation, which will be presented in the next lemma (mixed associativity). So the proof is deferred until then. Also, see Problem 3.7.5. ■

Lemma 3.5 *Let $X = \{x_0, x_1, \ldots, x_m\}$. The composition product (3.31) has the following properties:*

1. $0_\delta \circ c_\delta = c_\delta \circ 0_\delta = c_\delta$;
2. $(c \,\tilde{\circ}\, d_\delta) \,\tilde{\circ}\, e_\delta = c \,\tilde{\circ}\, (d_\delta \circ e_\delta)$ *(mixed associativity);*
3. *associativity,*

where c, d, and e are suitably compatible formal power series over X.

Proof:
1. Observe $0_\delta \circ c_\delta = c_\delta + 0 \,\tilde{\circ}\, c_\delta = c_\delta$ using Lemma 3.4 item 1. On the other hand, $c_\delta \circ 0_\delta = 0_\delta + c \,\tilde{\circ}\, 0_\delta = \delta + \sum_{\eta \in X^*} (c, \eta)\phi_0(\eta)(\mathbf{1}) = c_\delta$ using the fact that $\phi_0(\eta)(\mathbf{1}) = \eta$ for all $\eta \in X^*$.
2. In light of item 1 in Lemma 3.4, it is sufficient to prove the claim only for $c = \eta \in X^k$, $k \geq 0$. The cases $k = 0$ and $k = 1$ are trivial. Assume the claim holds up to some fixed $k \geq 0$. Then via Lemma 3.4, item 4, and the induction hypothesis it follows that

$$\begin{aligned}((x_0\eta) \,\tilde{\circ}\, d_\delta) \,\tilde{\circ}\, e_\delta &= (x_0(\eta \,\tilde{\circ}\, d_\delta)) \,\tilde{\circ}\, e_\delta \\ &= x_0((\eta \,\tilde{\circ}\, d_\delta) \,\tilde{\circ}\, e_\delta) \\ &= x_0(\eta \,\tilde{\circ}\, (d_\delta \circ e_\delta)) \\ &= (x_0\eta) \,\tilde{\circ}\, (d_\delta \circ e_\delta).\end{aligned}$$

In a similar fashion, for $i = 1, 2, \ldots, m$ apply the properties in Lemma 3.4, items 1, 4, and 5 to get

$$
\begin{aligned}
&((x_i\eta)\,\tilde\circ\, d_\delta)\,\tilde\circ\, e_\delta \\
&\quad = [x_i(\eta\,\tilde\circ\, d_\delta) + x_0(d_i \sqcup\!\sqcup (\eta\,\tilde\circ\, d_\delta))]\,\tilde\circ\, e_\delta \\
&\quad = [x_i(\eta\,\tilde\circ\, d_\delta)]\,\tilde\circ\, e_\delta + [x_0(d_i \sqcup\!\sqcup (\eta\,\tilde\circ\, d_\delta))]\,\tilde\circ\, e_\delta \\
&\quad = x_i[(\eta\,\tilde\circ\, d_\delta)\,\tilde\circ\, e_\delta] + x_0[e_i \sqcup\!\sqcup ((\eta\,\tilde\circ\, d_\delta)\,\tilde\circ\, e_\delta)] + x_0[(d_i \sqcup\!\sqcup (\eta\,\tilde\circ\, d_\delta))\,\tilde\circ\, e_\delta] \\
&\quad = x_i[(\eta\,\tilde\circ\, d_\delta)\,\tilde\circ\, e_\delta] + x_0[\underbrace{(e_i + d_i\,\tilde\circ\, e_\delta)}_{(d\circledcirc e)_i} \sqcup\!\sqcup ((\eta\,\tilde\circ\, d_\delta)\,\tilde\circ\, e_\delta)].
\end{aligned}
$$

Now employ the induction hypothesis so that

$$
\begin{aligned}
((x_i\eta)\,\tilde\circ\, d_\delta)\,\tilde\circ\, e_\delta &= x_i[\eta\,\tilde\circ\,(d_\delta\circ e_\delta)] + x_0[(d\circledcirc e)_i \sqcup\!\sqcup (\eta\,\tilde\circ\,(d_\delta\circ e_\delta))] \\
&= (x_i\eta)\,\tilde\circ\,(d_\delta\circ e_\delta).
\end{aligned}
$$

Therefore, the claim holds for all $\eta\in X^*$, and the identity is proved. Note that this identity is equivalent to the one given in Lemma 3.4, item 7.

3. First apply (3.31) twice, then Lemma 3.4, item 1, and finally mixed associativity to get

$$
\begin{aligned}
(c_\delta\circ d_\delta)\circ e_\delta &= (d + (c\,\tilde\circ\, d_\delta))_\delta\circ e_\delta \\
&= (e + (d + (c\,\tilde\circ\, d_\delta))\,\tilde\circ\, e_\delta)_\delta \\
&= (e + (d\,\tilde\circ\, e_\delta) + c\,\tilde\circ\,(d_\delta\circ e_\delta))_\delta \\
&= ((d\circledcirc e) + c\,\tilde\circ\,(d_\delta\circ e_\delta))_\delta \\
&= c_\delta\circ(d_\delta\circ e_\delta).
\end{aligned}
$$

Hence, the lemma is proved. ■

The following theorem establishes what will be the underlying group describing output feedback for Fliess operators. Henceforth, it will be called the output feedback group.

Theorem 3.27 *The triple $(\delta+\mathbb{R}^m\langle\langle X\rangle\rangle,\circ,\delta)$ forms a group. This group acts as a right transformation group on $\mathbb{R}^m\langle\langle X\rangle\rangle$.*

Proof: In light of Lemma 3.5, the only open issue in establishing that $\delta+\mathbb{R}^m\langle\langle X\rangle\rangle$ is a group is demonstrating the existence of an inverse. Specifically, for a fixed $c_\delta\in\delta+\mathbb{R}^m\langle\langle X\rangle\rangle$, the composition inverse, $c_\delta^{-1}=\delta+c^{-1}$, must satisfy $c_\delta\circ c_\delta^{-1}=\delta$ and $c_\delta^{-1}\circ c_\delta=\delta$. From the first equation,

$$c_\delta\circ c_\delta^{-1} = \delta + c^{-1}\,\tilde\circ\, c_\delta^{-1} = \delta,$$

which reduces to

$$c^{-1} = (-c)\,\tilde{\circ}\,c_\delta^{-1}. \tag{3.32}$$

Likewise, from the second equation,

$$c = (-c^{-1})\,\tilde{\circ}\,c_\delta. \tag{3.33}$$

Now it was established in Theorem 3.25 that $e \mapsto (-c)\,\tilde{\circ}\,e_\delta$ is a contraction in the ultrametric sense on $\mathbb{R}^m\langle\langle X\rangle\rangle$ as a complete ultrametric space and thus has a unique fixed point. So it follows directly that c_δ^{-1} is a right inverse of c_δ, i.e., satisfies (3.32). To see that this same series is also a left inverse, first observe that (3.32) is equivalent to

$$c^{-1}\,\tilde{\circ}\,0_\delta + c\,\tilde{\circ}\,c_\delta^{-1} = 0, \tag{3.34}$$

using the identity $c^{-1}\,\tilde{\circ}\,0_\delta = c^{-1}$ and the left linearity of the mixed composition product. Substituting (3.34) back into itself where zero appears on the left-hand side and applying Lemma 3.4, item 7 gives

$$\begin{aligned} c^{-1}\,\tilde{\circ}\,(c\,\tilde{\circ}\,c_\delta^{-1} + c^{-1})_\delta + c\,\tilde{\circ}\,c_\delta^{-1} &= 0 \\ (c^{-1}\,\tilde{\circ}\,c_\delta)\,\tilde{\circ}\,c_\delta^{-1} + c\,\tilde{\circ}\,c_\delta^{-1} &= 0. \end{aligned}$$

Again from left linearity of the mixed composition product it follows that

$$(c^{-1}\,\tilde{\circ}\,c_\delta + c)\,\tilde{\circ}\,c_\delta^{-1} = 0.$$

Finally, Lemma 3.4, item 3 implies that $c^{-1}\,\tilde{\circ}\,c_\delta + c = 0$, which is equivalent to (3.33). Therefore, every element of $\delta + \mathbb{R}^m\langle\langle X\rangle\rangle$ has an inverse. Finally, it is clear that $\delta + \mathbb{R}^m\langle\langle X\rangle\rangle$ acts as a right transformation group on $\mathbb{R}^m\langle\langle X\rangle\rangle$ in light of Lemma 3.5, item 2, namely, the mixed associativity property. ■

Example 3.29 If $c, d \in \mathbb{R}^m_{LC}\langle\langle X\rangle\rangle$, and c is a linear series then

$$F_{c\,\tilde{\circ}\,d_\delta}[u] = F_c[u + F_d[u]] = F_c[u] + F_{c\circ d}[u],$$

or equivalently,

$$c\,\tilde{\circ}\,d_\delta = c + c \circ d \tag{3.35}$$

Similarly,

$$c \circ (d_1 + d_2) = c \circ d_1 + c \circ d_2.$$

Therefore, using (3.32) and then (3.35) repeatedly, it follows that

$$\begin{aligned}
c_\delta^{-1} &= \delta + c^{-1} \\
&= \delta - c \,\tilde{\circ}\, c_\delta^{-1} \\
&= \delta - c - c \circ c^{-1} \\
&= \delta - c - c \circ (-c \,\tilde{\circ}\, c_\delta^{-1}) \\
&= \delta - c - c \circ (-c - c \circ c^{-1}) \\
&= \delta - c + c \circ c + c \circ c \circ c^{-1} \\
&\;\vdots \\
&= \delta - c + c^{\circ 2} - c^{\circ 3} + \cdots ,
\end{aligned}$$

where $c^{\circ i}$ denotes the composition product power. This is equivalent to the series expansion of the inverse appearing in (3.30). When $c = x_1$ observe

$$\begin{aligned}
(\delta + x_1)^{-1} &= \delta - x_1 + x_0 x_1 - x_0^2 x_1 + \cdots \\
&= \delta - (-x_0)^* x_1,
\end{aligned}$$

where $d^* := \sum_{i\geq 0} d^i$. In contrast, the series $c = x_0$ is not linear, and in this case

$$\delta - x_0 + x_0^{\circ 2} - x_0^{\circ 3} + \cdots = \delta - x_0 + x_0 - x_0 + \cdots ,$$

which is neither locally finite nor summable. Nevertheless, it can be easily verified directly that $(\delta + x_0)^{-1} = \delta - x_0$. So the element is invertible but does not have a series expansion of the type available for linear series. □

The explicit formula for the output feedback product conjectured earlier is now verified directly using the machinery developed above.

Theorem 3.28 *For any $c, d \in \mathbb{R}^m\langle\langle X\rangle\rangle$, it follows that*

$$c@d = c \,\tilde{\circ}\, (-d \circ c)_\delta^{-1}. \tag{3.36}$$

Proof: Recall that the feedback equation for the system in Figure 3.15 reduces to the fixed point equation

$$e = c \,\tilde{\circ}\, (d \circ e)_\delta .$$

where $e = c@d$. The solution above can be checked by direct substitution with the aid of the identity in Lemma 3.4, item 6, and (3.32):

$$\begin{aligned} c\,\tilde{\circ}\,(d\circ e)_\delta\big|_{e=c\,\tilde{\circ}\,(-d\circ c)_\delta^{-1}} &= c\,\tilde{\circ}\,(d\circ(c\,\tilde{\circ}\,(-d\circ c)_\delta^{-1}))_\delta \\ &= c\,\tilde{\circ}\,((d\circ c)\,\tilde{\circ}\,(-d\circ c)_\delta^{-1})_\delta \\ &= c\,\tilde{\circ}\,(-d\circ c)_\delta^{-1} \\ &= e. \end{aligned}$$

■

The goal now is to describe a Faà di Bruno type Hopf algebra associated with the group $(\delta+\mathbb{R}^m\langle\langle X\rangle\rangle,\circ,\delta)$, whose antipode facilitates the explicit computation of the inverse of the group element appearing in the feedback product above. The coordinate maps for this group have the form

$$a_\eta^i:\delta+\mathbb{R}^m\langle\langle X\rangle\rangle\to\mathbb{R}:c_\delta\mapsto(c_i,\eta),$$

where $\eta\in X^*$ and $i=1,2,\ldots,m$.[10] In addition, a special coordinate function $\mathbf{1}_\delta$ is introduced with the defining property that $c_\delta\in\delta+\mathbb{R}^m\langle\langle X\rangle\rangle$ maps to one in every case.[11] Let V denote the $\mathbb{R}$-vector space spanned by these maps. If the *degree* of a_η^i is defined as $\deg(a_\eta^i)=2\,|\eta|_{x_0}+\sum_{j=1}^m|\eta|_{x_j}+1$, then V is a connected graded vector space. That is, $V=\bigoplus_{n\geq 0}V_n$ with

$$V_n=\mathrm{span}_{\mathbb{R}}\{a_\eta^i:\deg(a_\eta^i)=n\},\;\; n>0,$$

and $V_0=\mathbb{R}\mathbf{1}_\delta$.

Consider next the free unital commutative $\mathbb{R}$-algebra, H, with product

$$\mu:a_\eta^i\otimes a_\xi^j\mapsto a_\eta^i a_\xi^j$$

and unit $\mathbf{1}_\delta$. This product is clearly associative. The grading on V induces a connected grading on H with $\deg(a_\eta^i a_\xi^j)=\deg(a_\eta^i)+\deg(a_\xi^j)$ and $\deg(\mathbf{1}_\delta)=0$. Specifically, $H=\bigoplus_{n\geq 0}H_n$, where

$$H_n=\mathrm{span}_{\mathbb{R}}\{a_{\eta_1}^{i_1}a_{\eta_2}^{i_2}\cdots a_{\eta_\ell}^{i_\ell}:\sum_{j=1}^{\ell}\deg(a_{\eta_j}^{i_j})=n\},\;\; n>0,$$

and $H_0=\mathbb{R}\mathbf{1}_\delta$.

---

[10] Given the bijection between $\delta+\mathbb{R}^m\langle\langle X\rangle\rangle$ and $\mathbb{R}^m\langle\langle X\rangle\rangle$, $a_\eta^i(c_\delta)$ will often be abbreviated by $a_\eta^i(c)$.

[11] The subscript δ is added here to distinguish between this coordinate function and the monomial $\mathbf{1}=1\emptyset$, namely, the unit for the catenation and shuffle algebras.

Three coproducts are now introduced. The first coproduct is used to define the Hopf algebra on H. The remaining two coproducts provide a recursive manner in which to compute it. Recalling that $c_\delta \circ d_\delta = \delta + c \circledcirc d$, define Δ for any $a_\eta^i \in V^+ := \bigoplus_{n>0} V_n$ such that

$$\Delta a_\eta^i(c,d) = a_\eta^i(c \circledcirc d) = (c_i \circledcirc d, \eta).$$

The coassociativity of Δ follows from the associativity of the group product. Specifically, for any $c, d, e \in \mathbb{R}^m\langle\langle X\rangle\rangle$:

$$\begin{aligned}(\mathrm{id} \otimes \Delta) \circ \Delta a_\eta^i(c,d,e) &= (c_i \circledcirc (d \circledcirc e), \eta) \\ &= ((c \circledcirc d)_i \circledcirc e, \eta) \\ &= (\Delta \otimes \mathrm{id}) \circ \Delta a_\eta^i(c,d,e).\end{aligned}$$

Therefore, $(\mathrm{id} \otimes \Delta) \circ \Delta = (\Delta \otimes \mathrm{id}) \circ \Delta$ as required.

The second coproduct is $\Delta_{\sqcup\sqcup}^j(V^+) \subset V^+ \otimes V^+$, which is isomorphic to sh* via the coordinate maps. That is,

$$\Delta_{\sqcup\sqcup}^j a_\emptyset^i = a_\emptyset^i \otimes a_\emptyset^j \tag{3.37a}$$

$$\Delta_{\sqcup\sqcup}^j \circ \theta_k = (\theta_k \otimes \mathrm{id} + \mathrm{id} \otimes \theta_k) \circ \Delta_{\sqcup\sqcup}^j, \tag{3.37b}$$

where id is the identity map on V^+, and θ_k denotes the endomorphism on V^+ specified by $\theta_k a_\eta^i = a_{x_k\eta}^i$ for $k = 0, 1, \ldots, m$ and $i, j = 1, 2, \ldots, m$.

Example 3.30 *The first few terms of $\Delta_{\sqcup\sqcup}^j$ are:*

$$\begin{aligned}
\Delta_{\sqcup\sqcup}^j a_\emptyset^i &= a_\emptyset^i \otimes a_\emptyset^j \\
\Delta_{\sqcup\sqcup}^j a_{x_{i_1}}^i &= a_{x_{i_1}}^i \otimes a_\emptyset^j + a_\emptyset^i \otimes a_{x_{i_1}}^j \\
\Delta_{\sqcup\sqcup}^j a_{x_{i_2}x_{i_1}}^i &= a_{x_{i_2}x_{i_1}}^i \otimes a_\emptyset^j + a_{x_{i_2}}^i \otimes a_{x_{i_1}}^j + a_{x_{i_1}}^i \otimes a_{x_{i_2}}^j + \\
&\quad a_\emptyset^i \otimes a_{x_{i_2}x_{i_1}}^j \\
\Delta_{\sqcup\sqcup}^j a_{x_{i_3}x_{i_2}x_{i_1}}^i &= a_{x_{i_3}x_{i_2}x_{i_1}}^i \otimes a_\emptyset^j + a_{x_{i_3}x_{i_2}}^i \otimes a_{x_{i_1}}^j + \\
&\quad a_{x_{i_3}x_{i_1}}^i \otimes a_{x_{i_2}}^j + a_{x_{i_3}}^i \otimes a_{x_{i_2}x_{i_1}}^j + \\
&\quad a_{x_{i_2}x_{i_1}}^i \otimes a_{x_{i_3}}^j + a_{x_{i_2}}^i \otimes a_{x_{i_3}x_{i_1}}^j + \\
&\quad a_{x_{i_1}}^i \otimes a_{x_{i_3}x_{i_2}}^j + a_\emptyset^i \otimes a_{x_{i_3}x_{i_2}x_{i_1}}^j.
\end{aligned}$$

□

The third coproduct is $\tilde{\Delta} a_\eta^i = \Delta a_\eta^i - \mathbf{1}_\delta \otimes a_\eta^i$ or, equivalently, the coproduct induced by the identity

$$\tilde{\Delta} a_\eta^i(c,d) = (c_i \,\tilde{\circ}\, d_\delta, \eta) = \sum a_{\eta(1)}^i(c) a_{\eta(2)}^i(d).$$

A key observation is that this coproduct can be computed recursively as described in the next lemma. It is not difficult to show using items 2 and 3 of this lemma that $a_{\eta(1)}^i \in V^+$ and $a_{\eta(2)}^i \in H$, and thus, $\tilde{\Delta} V^+ \subseteq V^+ \otimes H$.

Lemma 3.6 *The following identities hold:*

1. $\tilde{\Delta} a_\emptyset^i = a_\emptyset^i \otimes \mathbf{1}_\delta$
2. $\tilde{\Delta} \circ \theta_i = (\theta_i \otimes \mathrm{id}) \circ \tilde{\Delta}$
3. $\tilde{\Delta} \circ \theta_0 = (\theta_0 \otimes \mathrm{id}) \circ \tilde{\Delta} + (\theta_i \otimes \mu) \circ (\tilde{\Delta} \otimes \mathrm{id}) \circ \Delta_{\sqcup\!\sqcup}^i$,

$i = 1, 2, \ldots, m$, *where* id *denotes the identity map on* H.[12]

Proof:
1. First note that any series c can be uniquely decomposed as $c = (c,\emptyset)\emptyset + x_i c^i$, $i = 0, 1, \ldots, m$, where the series c^i are arbitrary. In which case, using the left linearity of the mixed composition product and Lemma 3.4, item 4, it follows that

$$\begin{aligned}
\tilde{\Delta} a_\emptyset^i(c,d) &= a_\emptyset^i(c \,\tilde{\circ}\, d_\delta) = a_\emptyset^i\left((c,\emptyset)\emptyset + (x_j c^j) \,\tilde{\circ}\, d_\delta\right) \\
&= (c_i, \emptyset) + a_\emptyset^i(x_j(c^j \,\tilde{\circ}\, d_\delta) + x_0(d_j \sqcup\!\sqcup (c^j \,\tilde{\circ}\, d_\delta))) \\
&= (c_i, \emptyset) = (a_\emptyset^i \otimes \mathbf{1}_\delta)(c,d).
\end{aligned}$$

2. For any $\eta \in X^*$ observe

$$\begin{aligned}
(\tilde{\Delta} \circ \theta_i) a_\eta^j(c,d) &= \tilde{\Delta} a_{x_i\eta}^j(c,d) \\
&= a_{x_i\eta}^j(x_k(c^k \,\tilde{\circ}\, d_\delta) + x_0(d_k \sqcup\!\sqcup (c^k \,\tilde{\circ}\, d_\delta))) \\
&= a_\eta^j(c^i \,\tilde{\circ}\, d_\delta) \\
&= \tilde{\Delta} a_\eta^j(c^i, d) \\
&= \sum a_{\eta(1)}^j \otimes a_{\eta(2)}^j(c^i, d)
\end{aligned}$$

[12] The Einstein summation notation is used in item 3 and throughout to indicate summations from either 0 or 1 to m, e.g., $\sum_{i=1}^m a_i b^i = a_i b^i$. It will be clear from the context which lower bound is applicable.

$$= \sum \theta_i(a^j_{\eta(1)}) \otimes a^j_{\eta(2)}(c,d)$$
$$= (\theta_i \otimes \mathrm{id}) \circ \tilde{\Delta} a^j_\eta(c,d).$$

Note that since $a^j_{\eta(1)} \in V^+$, the operation $\theta_i(a^j_{\eta(1)})$ is well defined.
3. Proceeding as in the previous item, it follows that

$$\begin{aligned}
&(\tilde{\Delta} \circ \theta_0) a^i_\eta(c,d) \\
&= a^i_{x_0\eta}(c \,\tilde{\circ}\, d_\delta) \\
&= a^i_{x_0\eta}(x_j(c^j \,\tilde{\circ}\, d_\delta) + x_0(d_j \sqcup\!\sqcup (c^j \,\tilde{\circ}\, d_\delta))) \\
&= a^i_\eta(c^0 \,\tilde{\circ}\, d_\delta + d_j \sqcup\!\sqcup (c^j \,\tilde{\circ}\, d_\delta)) \\
&= a^i_\eta(c^0 \,\tilde{\circ}\, d_\delta) + \sum_{j=1}^m \Delta^j_{\sqcup\!\sqcup}\, a^i_\eta(c^j \,\tilde{\circ}\, d_\delta, d) \\
&= a^i_\eta(c^0 \,\tilde{\circ}\, d_\delta) + \sum_{j=1}^m \sum_{\xi,\nu \in X^*} (\eta, \xi \sqcup\!\sqcup \nu)\, a^i_\xi(c^j \,\tilde{\circ}\, d_\delta) a^j_\nu(d) \\
&= \tilde{\Delta} a^i_\eta(c^0, d) + \sum_{j=1}^m \sum_{\xi,\nu \in X^*} (\eta, \xi \sqcup\!\sqcup \nu)\, (\tilde{\Delta} a^i_\xi \otimes a^j_\nu)(c^j, d, d) \\
&= (\theta_0 \otimes \mathrm{id}) \circ \tilde{\Delta} a^i_\eta(c,d) + (\theta_j \otimes \mathrm{id}) \circ \\
&\qquad \sum_{\xi,\nu \in X^*} (\eta, \xi \sqcup\!\sqcup \nu)\, (\tilde{\Delta} a^i_\xi \otimes a^j_\nu)(c, d, d) \\
&= (\theta_0 \otimes \mathrm{id}) \circ \tilde{\Delta} a^i_\eta(c,d) + (\theta_j \otimes \mu) \circ (\tilde{\Delta} \circ \mathrm{id}) \circ \Delta^j_{\sqcup\!\sqcup}\, a^i_\eta(c,d).
\end{aligned}$$

■

The next theorem is a central result describing the algebraic underpinnings of the feedback connection.

Theorem 3.29 (H, μ, Δ) *is a connected graded commutative noncocommutative unital Hopf algebra.*

Proof: From the development above, it is clear that (H, μ, Δ) is a bialgebra with unit $\mathbf{1}_\delta$ and counit ϵ defined by $\epsilon(a_\eta) = 0$ for all $\eta \in X^*$ and $\epsilon(\mathbf{1}_\delta) = 1$. Here it is shown that this bialgebra is graded and connected. Therefore, H automatically has an antipode, and thus, is a Hopf algebra by Theorem 2.10. Specifically, since the algebra H is graded by H_n, $n \geq 0$ with $H_0 = \mathbb{R}\mathbf{1}_\delta$, it only needs to be shown for any $a^i_\eta \in V^+$ that

$$\tilde{\Delta} a^i_\eta \in (V^+ \otimes H)_n := \bigoplus_{\substack{j+k=n \\ j\geq 1, k\geq 0}} V_j \otimes H_k. \tag{3.38}$$

This fact is evident from the first few terms computed via Lemma 3.6:

$$\begin{aligned}
n=1 &: \tilde{\Delta} a^i_\emptyset = a^i_\emptyset \otimes \mathbf{1}_\delta \\
n=2 &: \tilde{\Delta} a^i_{x_j} = a^i_{x_j} \otimes \mathbf{1}_\delta \\
n=3 &: \tilde{\Delta} a^i_{x_0} = a^i_{x_0} \otimes \mathbf{1}_\delta + a^i_{x_\ell} \otimes a^\ell_\emptyset \\
n=3 &: \tilde{\Delta} a^i_{x_j x_k} = a^i_{x_j x_k} \otimes \mathbf{1}_\delta \\
n=4 &: \tilde{\Delta} a^i_{x_0 x_j} = a^i_{x_0 x_j} \otimes \mathbf{1}_\delta + a^i_{x_\ell} \otimes a^\ell_{x_j} + a^i_{x_\ell x_j} \otimes a^\ell_\emptyset \\
n=4 &: \tilde{\Delta} a^i_{x_j x_0} = a^i_{x_j x_0} \otimes \mathbf{1}_\delta + a^i_{x_j x_\ell} \otimes a^\ell_\emptyset \\
n=4 &: \tilde{\Delta} a^i_{x_j x_k x_l} = a^i_{x_j x_k x_l} \otimes \mathbf{1}_\delta \\
n=5 &: \tilde{\Delta} a^i_{x_0^2} = a^i_{x_0^2} \otimes \mathbf{1}_\delta + a^i_{x_\ell} \otimes a^\ell_{x_0} + a^i_{x_\ell x_0} \otimes a^\ell_\emptyset + \\
&\qquad a^i_{x_0 x_\ell} \otimes a^\ell_\emptyset + a^i_{x_\ell x_\nu} \otimes a^\ell_\emptyset a^\nu_\emptyset,
\end{aligned}$$

where $i,j,k,l = 1,2,\ldots m$. In which case, using the identities $\Delta(a^i_\eta a^j_\xi) = \Delta a^i_\eta \Delta a^j_\xi$ and $\Delta a^i_\eta = \tilde{\Delta} a^i_\eta + \mathbf{1}_\delta \otimes a^i_\eta$, it follows that $\Delta H_n \subseteq (H \otimes H)_n$, and this would complete the proof. To prove (3.38), the following facts are essential:

1. $\deg(\theta_l a^i_\eta) = \deg(a^i_\eta) + 1$, $l = 1,2,\ldots,m$
2. $\deg(\theta_0 a^i_\eta) = \deg(a^i_\eta) + 2$
3. $\Delta^j_{\sqcup\!\sqcup} a^i_\eta \in (V^+ \otimes V^+)_{n+1}$, $n = \deg(a^i_\eta)$.

The proof is via induction on the length of η. When $|\eta| = 0$ then clearly $\tilde{\Delta} a^i_\emptyset = a^i_\emptyset \otimes \mathbf{1}_\delta \in V_1 \otimes H_0$ and $n = 1$. Assume now that (3.38) holds for words up to some fixed length $|\eta| \geq 0$. Let $n = \deg(a^i_\eta)$. There are two ways to increase the length of η. First consider $a^i_{x_l \eta}$ for some $l \neq 0$. From item 1 above $\deg(a^i_{x_l \eta}) = n+1$, and from Lemma 3.6 $\tilde{\Delta} a^i_{x_l \eta} = (\theta_l \otimes \mathrm{id}) \circ \tilde{\Delta} a^i_\eta$. Therefore, using the induction hypothesis, $\tilde{\Delta} a^i_{x_l a_\eta} \in \bigoplus_{j+k=n} V_{j+1} \otimes H_k \subset (V \otimes H)_{n+1}$, which proves the assertion. Consider next $a^i_{x_0 \eta}$. From item 2 above $\deg(a^i_{x_0 \eta}) = n+2$. Lemma 3.6 is employed as in the first case. First note that item 3 above $\Delta^j_{\sqcup\!\sqcup} a^i_\eta \in (V^+ \otimes V^+)_{n+1}$, and so using the induction hypothesis it follows that $(\tilde{\Delta} \otimes \mathrm{id}) \circ \Delta^j_{\sqcup\!\sqcup} a^i_\eta \in (V^+ \otimes H \otimes V^+)_{n+1}$. In which case, $(\theta_i \otimes \mu) \circ (\tilde{\Delta} \otimes \mathrm{id}) \circ \Delta^j_{\sqcup\!\sqcup} a^i_\eta \in (V^+ \otimes H)_{n+2}$. By a similar argument,

$(\theta_0 \otimes \mathrm{id}) \circ \tilde{\Delta} a_\eta^i \in (V^+ \otimes H)_{n+2}$. Thus, $\tilde{\Delta} a_{x_0\eta}^i \in (V^+ \otimes H)_{n+2}$, which again proves the assertion and completes the proof. ■

The deferred proof from Lemma 3.4 is addressed next.

Proof of Lemma 3.4, item 3: Recall the claim is that $c \,\tilde{\circ}\, d_\delta = k\mathbf{1}$ implies $c = k\mathbf{1}$, $k \in \mathbb{R}^\ell$. If $c \,\tilde{\circ}\, d_\delta = k\mathbf{1}$ then clearly $k_i = a_\emptyset^i(c \,\tilde{\circ}\, d_\delta) = \tilde{\Delta} a_\emptyset^i(c,d) = a_\emptyset^i c$, $i = 1,2,\ldots,\ell$. Furthermore, for any $x_j \in X$ with $j \neq 0$, $0 = a_{x_j}^i(c \,\tilde{\circ}\, d_\delta) = \tilde{\Delta} a_{x_j}^i(c,d) = a_{x_j}^i c$, $i = 1,2,\ldots,\ell$. Now suppose $a_\eta^i c = 0$, $i = 1,2,\ldots,\ell$ for all $a_\eta^i \in V_k$ with $k = 1,2,\ldots,n$. Then for any $x_j \in X$

$$0 = \tilde{\Delta} a_{x_j\eta}^i(c,d) = a_{x_j\eta}^i c + \sum_{a_{x_j\eta(2)}^i \neq 1} a_{x_j\eta(1)}^i(c)\, a_{x_j\eta(2)}^i(d),$$

where in general $a_{x_j\eta(1)}^i \neq a_\emptyset^i$. Therefore, $a_{x_j\eta}^i c = 0$, $i = 1,2,\ldots,\ell$. In which, case $c = k\mathbf{1}$. ■

Example 3.31 *Recall that Lemma 3.4, item 3 was used to establish that $\delta + \mathbb{R}^m\langle\langle X\rangle\rangle$ constitutes a transformation group. But once established, this identity becomes trivial to justify. Namely, if $c \,\tilde{\circ}\, d_\delta = k\mathbf{1}$ then*

$$c = c \,\tilde{\circ}\, (d_\delta \circ d_\delta^{-1}) = (c \,\tilde{\circ}\, d_\delta) \,\tilde{\circ}\, d_\delta^{-1} = k\mathbf{1} \,\tilde{\circ}\, d_\delta^{-1} = k\mathbf{1}.$$

□

The following result supports the primary application of the Hopf algebra (H, μ, Δ) in computing the feedback product.

Lemma 3.7 *The Hopf algebra (H, μ, Δ) has an antipode S satisfying $a_\eta^i(c^{-1}) = (Sa_\eta^i)(c)$ for all $\eta \in X^*$ and $c \in \mathbb{R}^m\langle\langle X\rangle\rangle$.*

Proof: The claim follows directly from (2.32). ■

Finally, it was established in Theorem 2.10 that the antipode, S, of any graded connected Hopf algebra (H, μ, Δ) can be computed for any $a \in H^+$ by

$$Sa = -a - \sum (Sa'_{(1)})a'_{(2)}, \tag{3.39}$$

or alternatively,

$$Sa = -a - \sum a'_{(1)} Sa'_{(2)}, \tag{3.40}$$

where the reduced coproduct is $\Delta' a = \Delta a - a \otimes \mathbf{1}_\delta - \mathbf{1}_\delta \otimes a = \sum a'_{(1)} \otimes a'_{(2)}$. This can be viewed as being *partially* recursive in that the coproduct needs to be computed first before the antipode recursion can be applied. The next theorem provides a *fully* recursive algorithm to compute the antipode for the output feedback group.

Theorem 3.30 *The antipode, S, of any $a_\eta^i \in V^+$ in the output feedback Hopf algebra can be computed by the following algorithm:*

i. Recursively compute $\Delta^j_{\sqcup\sqcup}$ via (3.37).
ii. Recursively compute $\tilde{\Delta}$ via Lemma 3.6.
iii. Recursively compute S via (3.39) or (3.40) with

$$\Delta' a_\eta^i = \tilde{\Delta} a_\eta^i - a_\eta^i \otimes \mathbf{1}_\delta.$$

Proof: In light of the previous results, the only detail is the minor observation that S is the antipode of the Hopf algebra with coproduct $\Delta a = \tilde{\Delta} a + \mathbf{1}_\delta \otimes a$. In which case, the corresponding reduced coproduct is as described in step *iii*. ■

Applying the algorithm above via the left antipode formula (3.39) gives the antipode of the first few coordinate maps:

$$H_1 : Sa^i_\emptyset = -a^i_\emptyset \tag{3.41a}$$
$$H_2 : Sa^i_{x_j} = -a^i_{x_j} \tag{3.41b}$$
$$H_3 : Sa^i_{x_0} = -a^i_{x_0} + a^i_{x_\ell} a^\ell_\emptyset \tag{3.41c}$$
$$H_3 : Sa^i_{x_j x_k} = -a^i_{x_j x_k} \tag{3.41d}$$
$$H_4 : Sa^i_{x_0 x_j} = -a^i_{x_0 x_j} + a^i_{x_\ell} a^\ell_{x_j} + a^i_{x_\ell x_j} a^\ell_\emptyset \tag{3.41e}$$
$$H_4 : Sa^i_{x_j x_0} = -a^i_{x_j x_0} + a^i_{x_j x_\ell} a^\ell_\emptyset \tag{3.41f}$$
$$H_4 : Sa^i_{x_j x_k x_l} = -a^i_{x_j x_k x_l} \tag{3.41g}$$

$$\begin{aligned} H_5 : Sa^i_{x_0^2} &= -a^i_{x_0^2} - (Sa^i_{x_\ell}) a^\ell_{x_0} - (Sa^i_{x_\ell x_0}) a^\ell_\emptyset - (Sa^i_{x_0 x_\ell}) a^\ell_\emptyset - \\ &\quad (Sa^i_{x_\ell x_\nu}) a^\ell_\emptyset a^\nu_\emptyset \\ &= -a^i_{x_0^2} - (-a^i_{x_\ell}) a^\ell_{x_0} - (-a^i_{x_\ell x_0} + \cancel{a^i_{x_\ell x_\nu} a^\nu_\emptyset}) a^\ell_\emptyset - \\ &\quad (-a^i_{x_0 x_\ell} + a^i_{x_\nu} a^\nu_{x_\ell} + a^i_{x_\nu x_\ell} a^\nu_\emptyset) a^\ell_\emptyset - \cancel{(-a^i_{x_\ell x_\nu}) a^\ell_\emptyset a^\nu_\emptyset} \\ &= -a^i_{x_0^2} + a^i_{x_\ell} a^\ell_{x_0} + a^i_{x_\ell x_0} a^\ell_\emptyset + a^i_{x_0 x_\ell} a^\ell_\emptyset - a^i_{x_\nu} a^\nu_{x_\ell} a^\ell_\emptyset - \end{aligned}$$

$$a^i_{x_\nu x_\ell} a^\nu_\emptyset a^\ell_\emptyset, \tag{3.41h}$$

where $i, j, k, l = 1, 2, \ldots m$. The explicit calculations for $Sa^i_{x_0^2}$ are shown above to display the inter-term cancellation. This is the same phenomenon observed for the classical Faà di Bruno Hopf algebra presented in Section 2.6. As with that Hopf algebra, the right antipode formula here is also known to be cancellation free and thus is preferred for calculations. Finally, it should be noted when $m = 1$, i.e., the single-input, single-output case, that all the summations above vanish.

Example 3.32 Consider a linear time-invariant system with an $m \times m$ transfer function $H(s)$ and state space realization (A, B, C). The corresponding components of the linear generating series are $c_i = \sum_{k\geq 0} \sum_{j=1}^m (c_i, x_0^k x_j)\, x_0^k x_j$, where $(c_i, x_0^k x_j) = C_i A^k B_j$, $k \geq 0$, and C_i, B_j denote the i-th row of C and the j-th column of B, respectively. The composition inverse of the return difference matrix $I + H(s)$ is computed directly as

$$(I + C(sI - A)^{-1}B)^{-1} = I - C(sI - (A - BC))^{-1}B.$$

Therefore, it follows that

$$(c_i^{-1}, x_0^k x_j) = -C_i(A - BC)^k B_j, \;\; k \geq 0, \;\; i, j = 1, 2, \ldots, m.$$

Expanding this product gives results which are consistent with the antipode formulas (3.41). For example,

$$\begin{aligned}
(c_i^{-1}, x_0 x_j) &= -C_i(A - BC)B_j \\
&= -C_i A B_j + C_i B C B_j \\
&= -C_i A B_j + \sum_{\ell=1}^m C_i B_\ell C_\ell B_j \\
&= -(c_i, x_0 x_j) + \sum_{\ell=1}^m (c_i, x_\ell)(c_\ell, x_j) \\
&= (-a^i_{x_0 x_j} + a^i_{x_\ell} a^\ell_{x_j} + a^i_{x_\ell x_j} a^\ell_\emptyset)c \\
&= (Sa^i_{x_0 x_j})c,
\end{aligned}$$

where the fact that $(c, x_\ell x_j) = (c, \emptyset) = 0$ has been used in the second to the last line. □

A unity feedback system is one where the operator F_d in Figure 3.15 is replaced with the identity map $F_\delta = I$. At first glance, it does not appear that the output feedback formula (3.36) will apply to this situation. However, if the *loop* generating series $-d \circ c$ in this formula is replaced with $-c$ this corresponds exactly to a unity feedback system, and the formula does render the correct closed-loop generating series. So introducing a slight abuse of notation, the generating series for a unity feedback system will be denoted by $c@\delta := c \,\tilde{\circ}\, (-c)_\delta^{-1}$, and it is evident from (3.32) that $c@\delta = (-c)^{-1}$, and therefore, $c^{-1} = (-c)@\delta$. That is, every inverse generating series can be viewed as coming from a unity feedback system.

Example 3.33 Let $c = \sum_{k\geq 0} k!\, x_1^k$. The generating series for the unity feedback system $c@\delta = (-c)^{-1}$ is computed directly from (3.41). For example, the coefficients for all the degree four terms are:

$$\begin{aligned}
((-c)^{-1}, x_0x_1) &= Sa_{x_0x_1}(-c) \\
&= a_{x_1}(-x_1)a_{x_1}(-x_1) + a_{x_1x_1}(-2!\, x_1x_1)a_\emptyset(-1) = 3 \\
((-c)^{-1}, x_1x_0) &= Sa_{x_1x_0}(-c) \\
&= a_{x_1x_1}(-2!\, x_1x_1)a_\emptyset(-1) = 2 \\
((-c)^{-1}, x_1x_1x_1) &= Sa_{x_1x_1x_1}(-c) \\
&= -a_{x_1x_1x_1}(-3!\, x_1x_1x_1) = 6.
\end{aligned}$$

Therefore, the polynomial

$$a_4 = 6x_1^3 + 3x_0x_1 + 2x_1x_0$$

is comprised of all the degree 4 terms appearing in $c@\delta$. Continuing in this way, $c@\delta = \sum_{k\geq 1} a_k$, where

$$\begin{aligned}
a_1 &= 1 \\
a_2 &= x_1 \\
a_3 &= 2x_1^2 + x_0 \\
a_4 &= 6x_1^3 + 3x_0x_1 + 2x_1x_0 \\
a_5 &= 24x_1^4 + 12x_0x_1^2 + 8x_1x_0x_1 + 6x_1^2x_0 + 3x_0^2 \\
a_6 &= 120x_1^5 + 60x_0x_1^3 + 40x_1x_0x_1^2 + 30x_1^2x_0x_1 + 24x_1^3x_0 + \\
&\quad\; 15x_0^2x_1 + 12x_0x_1x_0 + 8x_1x_0^2.
\end{aligned}$$

These are the *Devlin polynomials*. They are known to be related by the simple linear recursion

$$a_n = (n-1)a_{n-1}x_1 + (n-2)a_{n-2}x_0, \;\; n \geq 2, \tag{3.42}$$

where $a_0 = 0$ and $a_1 = 1$. □

In applications, generally the plant, modeled by F_c in Figure 3.15 is fixed and the feedback law F_d is considered variable. In which case, there is an underlying additive feedback transformation group in this setting.

Definition 3.8 *For any fixed $c \in \mathbb{R}\langle\langle X\rangle\rangle$, define*

$$\mathcal{O}_c = \{e_\delta \in \delta + \mathbb{R}\langle\langle X\rangle\rangle : e_\delta = (d\circ c)_\delta,\; d \in \mathbb{R}\langle\langle X\rangle\rangle\}\,.$$

Theorem 3.31 *For any fixed series $c \in \mathbb{R}\langle\langle X\rangle\rangle$, the triple $(\mathcal{O}_c, +, \delta)$ defines an additive group, where*

$$e_\delta + e'_\delta = (d\circ c)_\delta + (d'\circ c)_\delta := ((d+d')\circ c)_\delta$$

for any $e_\delta = (d\circ c)_\delta, e'_\delta = (d'\circ c)_\delta \in \mathcal{O}_c$.

Proof: The claim follows directly from the left linearity of the composition product on the $\mathbb{R}$-vector space $\mathbb{R}\langle\langle X\rangle\rangle$. ■

The group $(\mathcal{O}_c, +, \delta)$ is isomorphic to the additive transformation group described in the following theorem.

Theorem 3.32 *The additive group $(\mathbb{R}\langle\langle X\rangle\rangle, +, 0)$ acts on the set $\mathbb{R}\langle\langle X\rangle\rangle$ as a right transformation group, where the action is given by the output feedback product. That is, $c@0 = c$ and*

$$(c@d_1)@d_2 = c@(d_1+d_2).$$

Proof: The first identity is trivial. For the second, two algebraic facts are needed. First, as described in Lemma 3.7, the composition inverse is defined in terms of a Hopf algebra antipode, S, using the group $(\delta + \mathbb{R}\langle\langle X\rangle\rangle, \circ, \delta)$. Such an S is always an antihomomorphism for both the algebra and the coalgebra structures on H, for example, $S(a_1a_2) = S(a_2)S(a_1), \forall a_1, a_2 \in H$. Therefore, it follows directly that $(c_\delta \circ d_\delta)^{-1} = d_\delta^{-1} \circ c_\delta^{-1}$. Second, from Lemma 3.4, item 6, recall that

$$c \circ (d \,\tilde{\circ}\, e_\delta) = (c \circ d) \,\tilde{\circ}\, e_\delta.$$

Proceeding with the calculation, it follows by definition of the output feedback product and the fact that $(\mathbb{R}\langle\langle X\rangle\rangle, \circ, \delta)$ is known to act as a right transformation on $\mathbb{R}\langle\langle X\rangle\rangle$ via the product $c \,\tilde{\circ}\, d_\delta$ that

$$\begin{aligned}
&(c@d_1)@d_2 \\
&\qquad = (c \,\tilde{\circ}\, (-d_1 \circ c)_\delta^{-1})@d_2 \\
&\qquad = (c \,\tilde{\circ}\, (-d_1 \circ c)_\delta^{-1}) \,\tilde{\circ}\, (-d_2 \circ (c \,\tilde{\circ}\, (-d_1 \circ c)_\delta^{-1}))_\delta^{-1} \\
&\qquad = c \,\tilde{\circ}\, \left[(-d_1 \circ c)_\delta^{-1} \circ (-d_2 \circ (c \,\tilde{\circ}\, (-d_1 \circ c)_\delta^{-1}))_\delta^{-1}\right].
\end{aligned}$$

Now apply the first fact stated above, the definition of the group product on $\delta + \mathbb{R}\langle\langle X\rangle\rangle$, and the second fact in this order to get

$$\begin{aligned}
(c@d_1)@d_2 &= c \,\tilde{\circ}\, \left[(-d_2 \circ (c \,\tilde{\circ}\, (-d_1 \circ c)_\delta^{-1}))_\delta \circ \right. \\
&\qquad \left. (-d_1 \circ c)_\delta\right]^{-1} \\
&= c \,\tilde{\circ}\, \left[(-d_1 \circ c) + (-d_2 \circ (c \,\tilde{\circ}\, (-d_1 \circ c)_\delta^{-1})) \,\tilde{\circ}\, \right. \\
&\qquad \left. (-d_1 \circ c)_\delta\right]_\delta^{-1} \\
&= c \,\tilde{\circ}\, \left[(-d_1 \circ c) + ((-d_2 \circ c) \,\tilde{\circ}\, (-d_1 \circ c)_\delta^{-1}) \,\tilde{\circ}\, \right. \\
&\qquad \left. (-d_1 \circ c)_\delta\right]_\delta^{-1}.
\end{aligned}$$

Finally, just simplify the result using properties already stated so that

$$\begin{aligned}
(c@d_1)@d_2 &= c \,\tilde{\circ}\, \left[(-d_1 \circ c) + (-d_2 \circ c) \,\tilde{\circ}\, ((-d_1 \circ c)_\delta^{-1} \circ \right. \\
&\qquad \left. (-d_1 \circ c)_\delta)\right]_\delta^{-1} \\
&= c \,\tilde{\circ}\, (-(d_1 + d_2) \circ c))_\delta^{-1} \\
&= c@(d_1 + d_2).
\end{aligned}$$

Example 3.34 Returning to the linear time-invariant case in Example 3.28, the above theorem reduces to a simple identity concerning transfer functions. Namely, if H is the transfer function for the plant and feedback G_1 is applied, then the closed-loop system has the transfer function

$$H_{cl,1} = H(I - G_1 H)^{-1}.$$

If a second feedback loop G_2 is then applied the resulting closed-loop transfer function is

$$H_{cl,2} = H_{cl,1}(I - G_2 H_{cl,1})^{-1}$$

$$= H(I - (G_1 + G_2)H)^{-1}.$$

Clearly, the second feedback loop will *cancel* the first feedback loop when $G_2 = -G_1$. It is easy to see that $(I - (G_1 + G_2)H)^{-1} \neq (I - G_1H)^{-1}(I - G_2H)^{-1}$. Therefore, what is essentially a compositional group in Example 3.28 will not describe the output feedback transformation group consider here. Nevertheless, all inverse operations above *are* with respect to this composition group. So it is still an essential concept.

□

Finally, issues connected with convergence are addressed. As with the cascade connection, the main results will be stated along with illustrative examples, but the most difficult proofs will be left to the literature. The first theorem states that local convergence is preserved under the composition inverse. The radius of convergence for this operation is also given.

Theorem 3.33 *Suppose $X = \{x_0, x_1, \ldots, x_m\}$. Let $c \in \mathbb{R}^m_{LC}\langle\langle X\rangle\rangle/\mathbb{R}^m_{GC}\langle\langle X\rangle\rangle$ with growth constants $K_c, M_c > 0$. If $b = c^{-1}$, then for every $\epsilon > 0$*

$$|(b,\eta)| \leq K_b\,(M_b + \epsilon)^{|\eta|}\,|\eta|!,\quad \eta \in X^*, \tag{3.43}$$

for some $K_b > 0$, where

$$M_b = \frac{M_c}{1 - mK_c \ln\left(1 + \frac{1}{mK_c}\right)}.$$

Therefore, $b \in \mathbb{R}^m_{LC}\langle\langle X\rangle\rangle$. Furthermore, if K_c and M_c are minimal, then no geometric growth constant smaller than M_b can satisfy (3.43), and thus, the radius of convergence of the composition inverse operation is

$$\frac{1}{M_c(m+1)}\left[1 - mK_c \ln\left(1 + \frac{1}{mK_c}\right)\right].$$

Example 3.35 In Example 3.33 it was shown that if $c = \sum_{k\geq 0} k!x_1^k$ then $(-c)^{-1} = \sum_{k\geq 0} a_k$, where a_k are the Devlin polynomials. Define the integer sequence

$$b_k^n = \max_{\eta\in X^n}(a_k,\eta)$$

corresponding to the largest coefficient in a_k for a given word length n. A straightforward inductive argument using (3.42) (see Problem 3.7.8) yields the identity

$$b_{2n+1}^{n} = \frac{(2n)!}{2^n n!} = 1 \cdot 3 \cdot 5 \cdots (2n-1), \tag{3.44}$$

which is usually denoted as the double factorial $(2n-1)!!$. Note, in particular, that

$$(2n-1)!! = \deg(\mathbf{1})\deg(x_0)\cdots\deg(x_0^{n-1}).$$

This implies that the fastest growing subsequence of coefficients of $(-c)^{-1}$ with respect to word length corresponds to those coefficients attached to the words x_0^k, $k \geq 0$, since it is known in general that

$$(a_k, x_{i_1}x_{i_2}\cdots x_{i_n}) = \prod_{j=1}^{n-1} \deg(x_{i_1}x_{i_2}\cdots x_{i_j})$$

when $k = \deg(x_{i_1}x_{i_2}\cdots x_{i_n})$ and $n \geq 2$. Therefore, in light of Theorem 3.33 with $K_c = M_c = 1$, it should be true that for some $K > 0$

$$(2n-1)!! \leq K\left(\frac{1}{1-\log 2}\right)^n n! = K(3.25889...)^n n!$$

for all $n \geq 1$. To see this is so, one can employ the well known identity

$$(2n-1)!! = \frac{1}{\pi}2^n \Gamma\left(n+\frac{1}{2}\right),$$

where Γ denotes the gamma function. The claim is now evident since $\Gamma(n+\frac{1}{2}) < \Gamma(n+1) = n!$ for all $n = 1, 2, \ldots$. □

Two immediate consequences of the previous theorem are the following.

Theorem 3.34 *The triple* $(\delta + \mathbb{R}^m_{LC}\langle\langle X\rangle\rangle, \circ, \delta)$ *is a subgroup of the group* $(\delta + \mathbb{R}^m\langle\langle X\rangle\rangle, \circ, \delta)$.

Proof: The set of generating series $\delta + \mathbb{R}^m_{LC}\langle\langle X\rangle\rangle$ is closed under composition since the set $\mathbb{R}^m_{LC}\langle\langle X\rangle\rangle$ is closed under addition and mixed composition (the proof is similar to that of Theorem 3.22). In light of Theorem 3.33, $\delta + \mathbb{R}^m_{LC}\langle\langle X\rangle\rangle$ is also closed under inversion. Hence, the theorem is proved. ■

Theorem 3.35 *If* $c, d \in \mathbb{R}^m_{LC}\langle\langle X\rangle\rangle$, *then* $c@d \in \mathbb{R}^m_{LC}\langle\langle X\rangle\rangle$.

Proof: Since the composition product, the mixed composition product, and the composition inverse all preserve local convergence, the claim follows directly from Theorem 3.28. ■

The radius of convergence of the output feedback product is given in the following theorem.

Theorem 3.36 *Suppose* $X = \{x_0, x_1, \ldots, x_m\}$. *Let* $c, d \in \mathbb{R}^m_{LC}\langle\langle X\rangle\rangle/\mathbb{R}^m_{GC}\langle\langle X\rangle\rangle$ *with growth constants* $K_c, M_c > 0$ *and* $K_d, M_d > 0$, *respectively. If* $e = c@d$, *then for every* $\epsilon > 0$

$$|(e,\eta)| \leq K_e(M_e+\epsilon)^{|\eta|}|\eta|!, \quad \eta \in X^*, \tag{3.45}$$

for some $K_e > 0$, *where*

$$M_e = \frac{1}{\int_0^{1/M_c} \frac{W(\exp(f(z)))}{1+W(\exp(f(z)))}\,dz}$$

and

$$f(z) = \frac{1-M_d z}{mK_d} + \ln\left(\frac{(1-M_c z)^{\frac{K_c M_d}{K_d M_c}}}{mK_d}\right).$$

Furthermore, if K_c, M_c, K_d *and* M_d *are minimal, then no geometric growth constant smaller than* M_e *can satisfy (3.45), and thus, the radius of convergence of* $e = c@d$ *is*

$$\frac{1}{(m+1)}\left[\int_0^{1/M_c} \frac{W(\exp(f(z)))}{1+W(\exp(f(z)))}\,dz\right]. \tag{3.46}$$

.

Example 3.36 Suppose $X = \{x_0, x_1\}$. Recall from Example 3.14 that if c is a locally maximal series with growth constants K_c, M_c then

$$y = F_c[u] = \frac{K_c}{1 - M_c E_{x_0+x_1}[u]}.$$

Setting $z_1 = y$ and computing the derivative gives directly a state space realization for this input-output system, namely,

$$\begin{aligned}\dot{z}_1 &= \frac{M_c}{K_c} z_1^2(1+u), \quad z_1(0) = K_c\\ y &= z_1.\end{aligned}$$

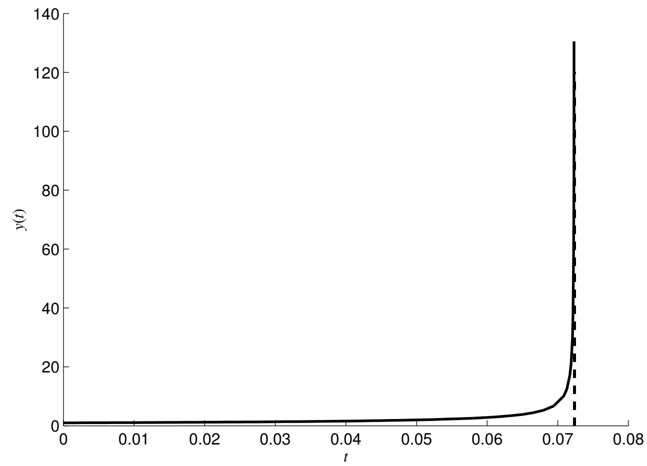


Fig. 3.17. Zero-input response of feedback system in Example 3.36

An analogous realization exists for F_d if d is locally maximal. If these two systems are now interconnected as shown in Figure 3.15, then a realization for the closed-loop system $F_{c@d}$ is

$$\dot{z}_1 = \frac{M_c}{K_c} z_1^2 \left(1 + z_2 + u\right), \quad z_1(0) = K_c \tag{3.47a}$$

$$\dot{z}_2 = \frac{M_d}{K_d} z_2^2 \left(1 + z_1\right), \quad z_2(0) = K_d \tag{3.47b}$$

$$y = z_1. \tag{3.47c}$$

A numerical simulation of (3.47) with $K_c = 1, M_c = 2, K_d = 3$ and $M_d = 4$ and u=0 gives the response shown in Figure 3.17. There is a finite escape time at $t_{esc} \approx 0.0723$. As discussed earlier for the parallel and cascade connections, it is known that the zero input response of locally maximal series defines the radius of convergence for the corresponding connection. The same is true here. Numerically integrating (3.46) for this case gives $t_{esc} = 0.0723$ as expected. □

The following corollary gives the radius of convergence for a unity feedback system.

Corollary 3.4 *Suppose* $X = \{x_0, x_1, \ldots, x_m\}$. *Let* $c \in \mathbb{R}^m_{LC}\langle\langle X\rangle\rangle / \mathbb{R}^m_{GC}\langle\langle X\rangle\rangle$ *with growth constants* $K_c, M_c > 0$. *If* $e = c@\delta$, *then for*

every $\epsilon > 0$
$$|(e,\eta)| \leq K_e(M_e+\epsilon)^{|\eta|}|\eta|!, \quad \eta \in X^*,$$
for some $K_e > 0$*, where*
$$M_e = \frac{M_c}{1 - mK_c \ln\left(1+\frac{1}{mK_c}\right)}. \tag{3.48}$$
Furthermore, if K_c *and* M_c *are minimal, then no geometric growth constant smaller than* M_e *is possible, and thus, the radius of convergence of* $e = c@\delta$ *is*
$$\frac{1}{M_c(m+1)}\left[1 - mK_c \ln\left(1+\frac{1}{mK_c}\right)\right].$$

The next corollary is useful for the convergence analysis of unity feedback systems having analytic inputs.

Corollary 3.5 *Suppose* $X = \{x_0, x_1, \ldots, x_m\}$*. Let* $c \in \mathbb{R}^m_{LC}\langle\langle X\rangle\rangle/\mathbb{R}^m_{GC}\langle\langle X\rangle\rangle$ *with growth constants* $K_c, M_c > 0$*. Assume* $e = c@\delta$ *and let* M_e *be as defined in (3.48). If* $c_u \in \mathbb{R}^m_{LC}[[X_0]]/\mathbb{R}^m_{GC}[[X_0]]$ *with growth constants* $K_{c_u}, M_{c_u} > 0$ *and* $c_y = e \circ c_u$*, then for every* $\epsilon > 0$
$$|(c_y, x_0^k)| \leq K_{c_y}(M_{c_y}+\epsilon)^k k!, \ k \geq 0,$$
for some $K_{c_y} > 0$*, where*
$$M_{c_y} = \frac{M_{c_u}}{1 - mK_{c_u} W\left(\frac{1}{mK_{c_u}} \exp\left(\frac{M_e - M_{c_u}}{mK_{c_u}M_e}\right)\right)}.$$
Thus, the interval of convergence for the output $y = F_{c_y}[u]$ *is at least as large as* $T = 1/M_{c_y}$*.*

Proof: The proof is an immediate consequence of Corollaries 3.3 and 3.4. ■

Example 3.37 Let $X = \{x_0, x_1\}$. Suppose $e = c@\delta$, where c is a locally maximal series with growth constants $K_c, M_c > 0$. The zero-input response of the feedback system is described by the solution of the state space system

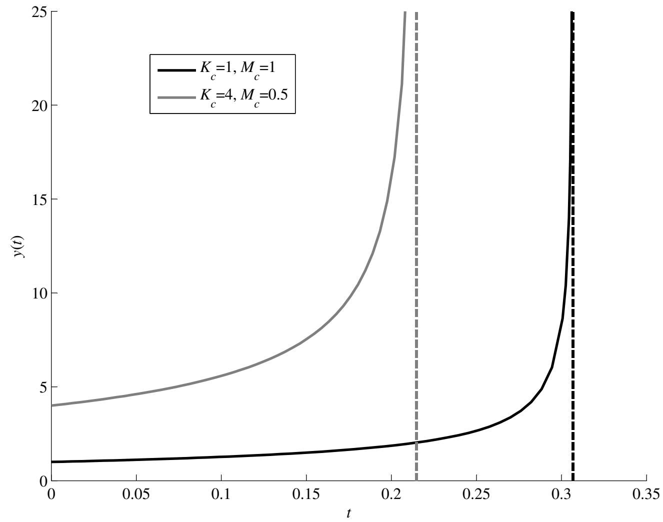


Fig. 3.18. Zero-input response of feedback system in Example 3.37

$$\dot{z} = \frac{M_c}{K_c}(z^2 + z^3), \quad z(0) = K_c$$
$$y = z.$$

Numerical solutions of this system are shown in Figure 3.18 when $K_c = M_c = 1$ and when $K_c = 4$, $M_c = 0.5$. As expected from Corollary 3.4, the respective finite escape times are $t_{esc} = 1 - \ln(2) \approx 0.3069$ and $t_{esc} = 2(1 - 4\log(5/4)) \approx 0.2149$. □

Example 3.38 Let $X = \{x_0, x_1\}$ and consider the case where $e = c@\delta$ with $c = \sum_{n\geq 0} n!\, x_1^n$. In comparison to the previous example, c has most of its coefficients equal to zero. Therefore, it is likely that the output will be finite over a longer interval. The zero-input response of the unity feedback system is described by the solution of

$$\dot{z} = z^3, \quad z(0) = 1$$
$$y = z.$$

Therefore, $y(t) = 1/\sqrt{1-2t}$ is finite up to $t = 0.5$, which is longer than the finite escape time of $t_{esc} = 0.3069$ obtained in the previous example. □

The next example illustrates an important distinction between the feedback connection and all previous interconnections considered, namely, feedback does *not* preserve global convergence.

Example 3.39 Consider a feedback interconnection involving the globally convergent series $c = x_1$ and $d = \sum_{k\geq 0} x_1^k$. Setting $z_1 = F_c[u] = E_{x_1}[u]$ and $z_2 = F_d[u] = \exp[E_{x_1}[u]]$, it is clear that $\dot{z}_1 = u$ and $\dot{z}_2 = z_2 u$. In which case, $F_{c@d}$ has the state space realization

$$\begin{aligned}
\dot{z}_1 &= z_2 + u, \quad z_1(0) = 0 \\
\dot{z}_2 &= z_1 z_2, \quad z_2(0) = 1 \\
y &= z_1.
\end{aligned}$$

Setting $u = 0$, y satisfies the initial value problem $\ddot{y} - \dot{y}y = 0$, $y(0) = 0$, $\dot{y}(0) = 1$, which has the solution

$$\begin{aligned}
y(t) = \sqrt{2}\tan\left(\frac{t}{\sqrt{2}}\right) &= \sum_{k=1}^{\infty}(-1)^{k-1}2^k(2^{2k-1})\frac{B_{2k}}{k}\,\frac{t^{2k-1}}{(2k-1)!} \\
&= t + \frac{t^3}{3!} + 4\,\frac{t^5}{5!} + 34\,\frac{t^7}{7!} + 496\,\frac{t^9}{9!} + \cdots
\end{aligned}$$

for $0 \leq t < \pi/\sqrt{2} \approx 2.2214$, where B_k denotes the k-th Bernoulli number. A numerical simulation of the state space realization confirms a finite escape time at $t_{esc} = 2.2214$. Hence, the closed-loop system is not globally convergent. □

While global convergence is not preserved under feedback, there is no doubt in light of Theorem 3.35 that the closed-loop system still has a locally convergent generating series. It stands to reason that the radius of convergence in this case might be *larger* than that given by Theorem 3.36 since stronger growth bound have been imposed on the generating series of the component systems. The first theorem describes the radius of convergence of the feedback connection of two globally convergent subsystems with Gevrey order $s = 0$. Then the unity feedback case is presented as a corollary. It is easy to directly compare radii of convergence in this latter case as illustrated by an example.

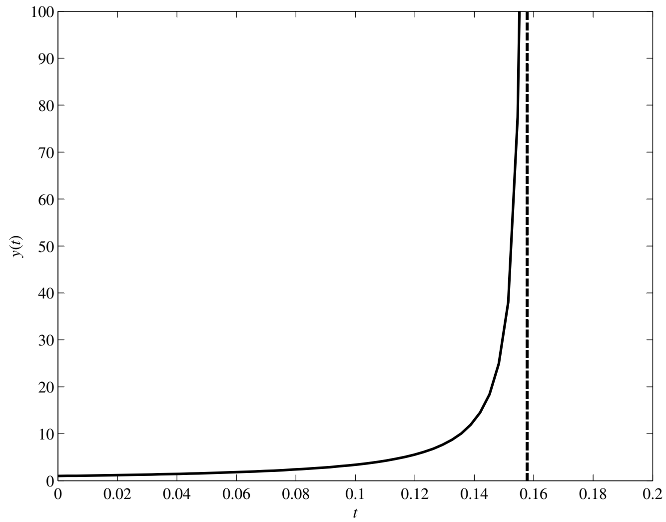


Fig. 3.19. Zero-input response of feedback system in Example 3.40

Theorem 3.37 *Suppose* $X = \{x_0, x_1, \ldots, x_m\}$. *Let* $c, d \in \mathbb{R}^m_{GC}\langle\langle X \rangle\rangle$ *with growth constants* $K_c, M_c > 0$ *and* $K_d, M_d > 0$, *respectively, and Gevrey order* s*=0. If* $e = c@d$, *then for every* $\epsilon > 0$

$$|(e, \eta)| \leq K_e(M_e + \epsilon)^{|\eta|}|\eta|!, \quad \eta \in X^*, \tag{3.49}$$

for some $K_e > 0$, *where*

$$M_e = \frac{1}{\int_0^\infty \frac{1}{1+W(\exp(f(z)))}\, dz}, \tag{3.50}$$

and

$$f(z) = \frac{mK_cM_d}{M_c}(\exp(M_c z) - 1) + M_d z + mK_d + \ln(mK_d).$$

Furthermore, if K_c, M_c, K_d, *and* M_d *are minimal, then no geometric growth constant smaller than* M_e *can satisfy (3.49), and thus, the radius of convergence of* $e = c@d$ *is*

$$\frac{1}{(m+1)}\left[\int_0^\infty \frac{1}{1 + W(\exp(f(z)))}\, dz\right]. \tag{3.51}$$

Example 3.40 Suppose $X = \{x_0, x_1\}$. Recall from Example 3.21 that a maximal series with Gevrey order $s = 0$ and growth constants K_c, M_c yields the Fliess operator

$$y = F_c[u] = K_c \exp\left(M_c E_{x_0+x_1}[u]\right).$$

Setting $z_1 = y$ gives directly the state space realization

$$\begin{aligned}\dot{z}_1 &= M_c z_1(1+u), \quad z_1(0) = K_c \\ y &= z_1.\end{aligned}$$

Therefore, the feedback interconnection of two such systems is realized by

$$\begin{aligned}\dot{z}_1 &= M_c z_1\left(1 + z_2 + u\right), \quad z_1(0) = K_c \\ \dot{z}_2 &= M_d z_2\left(1 + z_1\right), \quad z_2(0) = K_d \\ y_i &= z_1.\end{aligned}$$

A numerical simulation of this system with growth constants $K_c = 1, M_c = 2, K_d = 3$ and $M_d = 4$ gives the zero-input response shown in Figure 3.19. Numerical integration of (3.51) for this case gives $t_{esc} = 0.1570$ as observed in the figure. □

Corollary 3.6 *Suppose* $X = \{x_0, x_1, \ldots, x_m\}$. *Let* $c \in \mathbb{R}^m_{GC}\langle\langle X\rangle\rangle$ *with growth constants* $K_c, M_c > 0$ *and Gevrey order* $s = 0$. *If* $e = c@\delta$, *then for every* $\epsilon > 0$

$$|(e,\eta)| \le K_e(M_e+\epsilon)^{|\eta|}|\eta|!, \quad \eta \in X^*, \tag{3.52}$$

for some $K_e > 0$, *where*

$$M_e = \frac{M_c}{\ln\left(1 + \frac{1}{mK_c}\right)}.$$

Furthermore, if K_c *and* M_c *are minimal, then no geometric growth constant smaller than* M_e *can satisfy (3.52), and thus, the radius of convergence of* $e = c@\delta$ *is*

$$\frac{1}{M_c(m+1)}\left[\ln\left(1+\frac{1}{mK_c}\right)\right].$$

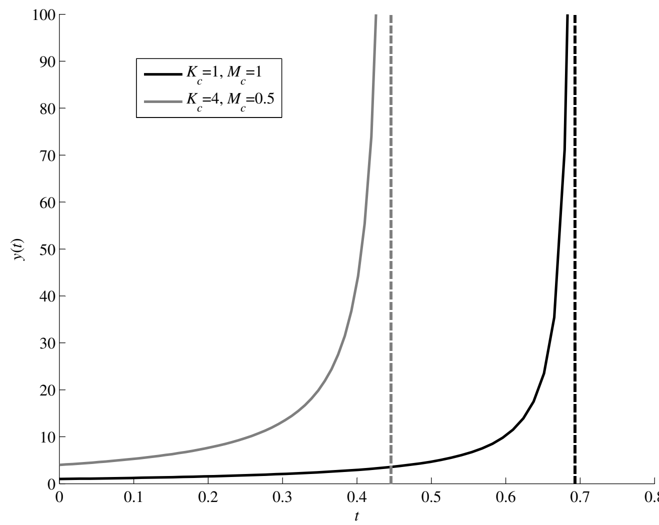


Fig. 3.20. Zero-input responses of the feedback system in Example 3.41

Proof: Set $K_c = K_d$ and $M_c = M_d$ in (3.50) and then evaluate directly. ■

Example 3.41 Suppose $X = \{x_0, x_1\}$. Let $e = c@\delta$ with c being a globally maximal series with growth constants K_c, M_c and $s = 0$. This is the global version of Example 3.37. The zero-input response of the unity feedback system is described by the solution of the state space system

$$\dot{z} = M_c(z + z^2), \quad z(0) = K_c$$
$$y = z.$$

Numerically generated solutions of this system are shown in Figure 3.20 when $K_c = M_c = 1$ and when $K_c = 4$, $M_c = 0.5$. From Corollary 3.6 the respective finite escape times are $t_{esc} = \ln(2) \approx 0.6931$ and $t_{esc} = 2\log(5/4) \approx 0.4463$. Note that these escape times are about twice that of the respective cases in Example 3.37. To see why this happens, observe that the geometric growth constant for the local case in Corollary 3.4 has the form $M_e = \alpha(K_c)M_c$, where

$$\alpha(K_c) = \frac{1}{1 - mK_c \ln\left(1 + \frac{1}{mK_c}\right)},$$

while for the global case in Corollary 3.6, $M_e = \gamma(K_c)M_c$, where

$$\gamma(K_c) = \frac{1}{\ln\left(1 + \frac{1}{mK_c}\right)}.$$

In light of the series expansions about $K_c = \infty$:

$$\alpha(K_c) = \frac{4}{3} + 2K_c + O\left(\frac{1}{K_c}\right)$$
$$\gamma(K_c) = \frac{1}{2} + K_c + O\left(\frac{1}{K_c}\right),$$

the radius of convergence for the global case with $s = 0$ is *always* about twice that for the local case, especially when $K_c \gg 1$. □

Example 3.42 Suppose $X = \{x_0, x_1\}$ and consider the case where $e = c@\delta$ with $c = \sum_{n\geq 0} x_1^n$. The series c has the same growth constants $K_c = M_c = 1$ as in Example 3.41 except most of its coefficients are zero. Thus, the zero-input response is expected to be finite over a longer interval. The zero-input response of the unity feedback system is described by the solution of

$$\dot{z} = z^2, \quad z(0) = 1$$
$$y = z.$$

Therefore, $y(t) = 1/(1-t)$ is finite up to $t = 1$, which exceeds the finite escape time of $t_{esc} = 0.6931$ in the previous example. □

Finally, a summary of the radii of convergence for all four elementary system interconnections is given in Table 3.2 for the case where the subsystems have only locally convergence generating series. The analogous summary for the globally convergent case is given in Table 3.3. Here the distinction is made between the composite system having a globally convergent generating series in the sense of (3.11) (GC) versus having only the corresponding operator being globally convergent.

Table 3.2. Radii of convergence for connections of only locally convergent systems

connection	$c,d \in \mathbb{R}^{\ell}_{LC}\langle\langle X\rangle\rangle/\mathbb{R}^{\ell}_{GC}\langle\langle X\rangle\rangle$	source
parallel	$\frac{1}{\max\{M_c,M_d\}(m+1)}$	Theorem 3.15
product	$\frac{1}{\max\{M_c,M_d\}(m+1)}$	Theorem 3.16
cascade ($\ell=m$ for d)	$\frac{1}{M_d(m+1)}\left[1-mK_dW\left(\frac{1}{mK_d}\exp\left(\frac{M_c-M_d}{mK_dM_c}\right)\right)\right]$	Theorem 3.22
feedback ($\ell=m$)	$\frac{1}{(m+1)}\int_0^{1/M_c}\frac{W(\exp(f(z)))}{1+W(\exp(f(z)))}\,dz$ $f(z)=\frac{1-M_dz}{mK_d}+\ln\left(\frac{(1-M_cz)^{\frac{K_cM_d}{K_dM_c}}}{mK_d}\right)$	Theorem 3.36
unity feedback ($\ell=m$)	$\frac{1}{M_c(m+1)}\left[1-mK_c\ln\left(1+\frac{1}{mK_c}\right)\right]$	Corollary 3.4

Table 3.3. Radii of convergence for connections of globally convergent systems

connection	$c,d \in \mathbb{R}^{\ell}_{GC}\langle\langle X\rangle\rangle$	source
parallel	∞ (GC)	Theorem 3.13
product	∞ (GC)	Theorem 3.13
cascade ($\ell=m$ for d)	∞	Theorem 3.23
feedback ($\ell=m,\ s=0$)	$\frac{1}{(m+1)}\int_0^{\infty}\frac{1}{1+W(\exp(f(z)))}\,dz$ $f(z)=\frac{mK_cM_d}{M_c}(\exp(M_cz)-1)+M_dz$ $+mK_d+\ln(mK_d)$	Theorem 3.37
unity feedback ($\ell=m,\ s=0$)	$\frac{1}{M_c(m+1)}\ln\left(1+\frac{1}{mK_c}\right)$	Corollary 3.6

3.8 Formal Fliess Operators

All the focus up to this point has been on Fliess operators which have at least a locally convergent generating series. This provides for a well defined mapping from a ball of input functions in $L_{\mathfrak{p}}^m[t_0,t_0+T]$ to a ball of output functions in $L_{\mathfrak{q}}^{\ell}[t_0,t_0+T]$. In this section, this requirement will be relaxed and instead, the class of *formal* Fliess operators will be defined without any reference to convergence. Theorem 3.19 motivates the approach, namely, that F_c can be viewed as a mapping from the set of formal inputs, $\mathbb{R}^m[[X_0]]$, to the set of formal outputs, $\mathbb{R}^\ell[[X_0]]$, using the composition product. After all, the composition product of two series as induced by operator composition is well defined (summable) independent of whether its arguments are convergent in any sense. The following definition makes this notion precise.

Definition 3.9 *The set of* ***formal Fliess operators*** *is the collection of mappings*

$$\mathscr{F} := \left\{\mathbb{R}^m[[X_0]] \to \mathbb{R}^\ell[[X_0]] : c_u \mapsto c_y = c \circ c_u,\ c \in \mathbb{R}^\ell\langle\langle X\rangle\rangle\right\}.$$

As was shown in Theorem 3.5 for the locally convergent case, the generating series of a formal Fliess operator is unique. It should be noted from the onset, however, that the method of proof for Theorem 3.5 does not apply here. The piecewise constant test input $\bar{u}$ employed earlier is not in general characterized over $[t_0, t_0+T]$ by any single generating series, $c_{\bar{u}}$. Thus, a completely different approach is needed here.

Theorem 3.38 *Let $c, d \in \mathbb{R}^\ell\langle\langle X\rangle\rangle$. If $c \circ c_u = d \circ c_u$ for all $c_u \in \mathbb{R}^m[[X_0]]$ then $c = d$.*

The following three lemmas are essential to the development of a proof. They recast the composition product in a different light, showing more of its combinatoric nature. (This approach will also be useful in Chapter 5 for defining Chen series.)

Lemma 3.8 *Let $c \in \mathbb{R}^\ell\langle\langle X\rangle\rangle$ and $c_u \in \mathbb{R}^m[[X_0]]$. Then for any $n \geq 0$*

$$(c\circ c_u, x_0^n) = (c, x_0^n) + \sum_{k=1}^{n} \sum_{\substack{i_1,\ldots,i_k=1\\ j_1,\ldots,j_k=0}}^{m,n} \left(c, \bar{P}_{i_1\cdots i_k}^{j_1\cdots j_k}(n)\right)(c_{u_{i_1}}, x_0^{j_1})\cdots(c_{u_{i_k}}, x_0^{j_k}),$$

where c_{u_i} is the i-th component series of c_u and

$$\bar{P}_{i_1\cdots i_k}^{j_1\cdots j_k}(n) = \sum_{n_0,\ldots,n_k=0}^{\infty} \chi_{n_0\cdots n_k}^{j_1\cdots j_k}(n)\, x_0^{n_k} x_{i_k}\cdots x_0^{n_1} x_{i_1} x_0^{n_0}$$

is a polynomial with coefficients

$$\chi_{n_0\cdots n_k}^{j_1\cdots j_k}(n) = \left(x_0^{n_k+1}[x_0^{j_k} \sqcup\!\sqcup [x_0^{n_{k-1}+1}[x_0^{j_{k-1}} \sqcup\!\sqcup \cdots x_0^{n_1+1}[x_0^{j_1} \sqcup\!\sqcup x_0^{n_0}]\cdots]]], x_0^n\right).$$

Clearly, $\chi^{j_1\cdots j_k}_{n_0\cdots n_k}(n)=0$ *whenever* $n_0+n_1+\cdots+n_k+k+j_1+j_2+\cdots j_k\neq n$. *So* $\bar{P}^{j_1\cdots j_k}_{i_1\cdots i_k}(n)$ *is homogeneous of degree* $n-j$, *where* $j:=j_1+j_2+\cdots+j_k$ *when* $k\leq n-j$, *and* $\bar{P}^{j_1\cdots j_k}_{i_1\cdots i_k}(n)=0$ *when* $k>n-j$.

Proof: From the definition of the composition product,

$$\begin{aligned}(c\circ c_u,x_0^n)&=\sum_{\eta\in X^*}(c,\eta)(\eta\circ c_u,x_0^n)\\&=(c,x_0^n)+\sum_{k=1}^{\infty}\sum_{\eta\in\Gamma_k}(c,\eta)(\eta\circ c_u,x_0^n),\end{aligned}\tag{3.53}$$

where

$$\Gamma_k:=\left\{\xi\in X^*:\sum_{j=1}^{m}|\xi|_{x_j}=k\right\}.$$

Let η be any word in Γ_k. Substitute $c_u=\sum_{j\geq 0}(c_u,x_0^j)$ componentwise into the following expression and use the $\mathbb{R}$-linearity of the shuffle product:

$$\begin{aligned}&(\eta\circ c_u,x_0^n)\\&=((x_0^{n_k}x_{i_k}\cdots x_0^{n_1}x_{i_1}x_0^{n_0})\circ c_u,x_0^n)\\&=\left(x_0^{n_k+1}[c_{u_{i_k}}\sqcup\!\sqcup[x_0^{n_{k-1}+1}[c_{u_{i_{k-1}}}\sqcup\!\sqcup\cdots x_0^{n_1+1}[c_{u_{i_1}}\sqcup\!\sqcup x_0^{n_0}]\cdots]]],x_0^n\right)\\&=\sum_{j_1,\ldots,j_k=0}^{\infty}\left(x_0^{n_k+1}[x_0^{j_k}\sqcup\!\sqcup[x_0^{n_{k-1}+1}[x_0^{j_{k-1}}\sqcup\!\sqcup\cdots x_0^{n_1+1}[x_0^{j_1}\sqcup\!\sqcup x_0^{n_0}]\cdots]]],x_0^n\right)\\&\quad(c_{u_{i_1}},x_0^{j_1})\cdots(c_{u_{i_k}},x_0^{j_k})\\&=\sum_{j_1,\ldots,j_k=0}^{\infty}\chi^{j_1\cdots j_k}_{n_0\cdots n_k}(n)\,(c_{u_{i_1}},x_0^{j_1})\cdots(c_{u_{i_k}},x_0^{j_k}).\end{aligned}$$

Finally, substitute the above identity in equation (3.53)

$$\begin{aligned}&(c\circ c_u,x_0^n)\\&\quad=(c,x_0^n)+\sum_{k=1}^{\infty}\sum_{\substack{i_1,\ldots,i_k=1\\j_1,\ldots,j_k\geq 0\\n_0,\ldots,n_k\geq 0}}^{m}(c,x_0^{n_k}x_{i_k}\cdots x_0^{n_1}x_{i_1}x_0^{n_0})\,\chi^{j_1\cdots j_k}_{n_0\cdots n_k}(n)\\&\qquad(c_{u_{i_1}},x_0^{j_1})\cdots(c_{u_{i_k}},x_0^{j_k})\end{aligned}$$

$$= (c, x_0^n) + \sum_{k=1}^{n} \sum_{\substack{i_1,\ldots,i_k=1 \\ j_1,\ldots,j_k=0}}^{m,n} \left(c, \bar{P}_{i_1\cdots i_k}^{j_1\cdots j_k}(n)\right) (c_{u_{i_1}}, x_0^{j_1}) \cdots (c_{u_{i_k}}, x_0^{j_k}),$$

and the lemma is proved. ■

An alternative identity for $(c \circ c_u, x_0^n)$ can be deduced from the one in Lemma 3.8 by introducing an ordering on the coefficients $(c_{u_{i_1}}, x_0^{j_1}) \cdots (c_{u_{i_k}}, x_0^{j_k})$. For each $k \geq 1$ define the set of $2 \times k$ matrices

$$S_k = \left\{ \begin{pmatrix} j_1 & j_2 & \cdots & j_k \\ i_1 & i_2 & \ldots & i_k \end{pmatrix} : 1 \leq i_l \leq m,\ j_l \geq 0, \right.$$
$$\left. (1,0) \leq (i_1, j_1) \leq \cdots \leq (i_k, j_k) \right\},$$

where " $\leq$ " denotes the lexicographic order on the set $\{(i,j) : i,j \in \mathbb{N}_0\}$. Define the positive integers $s_1, \ldots, s_p$ for a given element of S_k by

$$\begin{pmatrix} j_1 & j_2 & \cdots & j_k \\ i_1 & i_2 & \ldots & i_k \end{pmatrix} = \begin{pmatrix} \beta_1 \cdots \beta_1 & \beta_2 \cdots \beta_2 & \cdots & \beta_p \cdots \beta_p \\ \underbrace{\alpha_1 \cdots \alpha_1}_{s_1} & \underbrace{\alpha_2 \cdots \alpha_2}_{s_2} & \cdots & \underbrace{\alpha_p \cdots \alpha_p}_{s_p} \end{pmatrix}.$$

Using this ordering, the lemma below follows naturally.

Lemma 3.9 *Let $c \in \mathbb{R}^\ell\langle\langle X \rangle\rangle$ and $c_u \in \mathbb{R}^m[[X_0]]$. Then*

$$(c \circ c_u, x_0^n) = (c, x_0^n) + \sum_{k=1}^{n} \sum_{S_k} \frac{1}{s_1! \cdots s_p!} \left(c, P_{i_1\cdots i_k}^{j_1\cdots j_k}(n)\right)$$
$$(c_{u_{i_1}}, x_0^{j_1}) \cdots (c_{u_{i_k}}, x_0^{j_k}),$$

where the inner sum is taken over all elements of S_k such that $k+j \leq n$ and

$$P_{i_1\cdots i_k}^{j_1\cdots j_k}(n) := \sum_{\sigma \in \Pi_k} \bar{P}_{i_{\sigma(1)}\cdots i_{\sigma(k)}}^{j_{\sigma(1)}\cdots j_{\sigma(k)}}(n).$$

Here Π_k denotes the permutation group defined on the set $\{1, 2, \ldots, k\}$.

Proof: This new expression follows from Lemma 3.8 by grouping the terms related by permutation. ■

Example 3.43 Consider the single-input, single-output case so that $X = \{x_0, x_1\}$ and $\ell = 1$. Here the lower indices of $\bar{P}$ and P are such that $i_1 \cdots i_k = 1 \cdots 1$ in every case, so they are suppressed in the notation. For $k = 3$, the first few polynomials $P^{j_1 j_2 j_3}(n)$ are written below in terms of the polynomials $\bar{P}^{j_1 j_2 j_3}(n)$ using the definition:

$$\begin{aligned} P^{j_1 j_2 j_3}(n) = \bar{P}^{j_1 j_2 j_3}(n) + \bar{P}^{j_1 j_3 j_2}(n) + \bar{P}^{j_2 j_1 j_3}(n) + \\ \bar{P}^{j_2 j_3 j_1}(n) + \bar{P}^{j_3 j_1 j_2}(n) + \bar{P}^{j_3 j_2 j_1}(n). \end{aligned}$$

For example,

$$\begin{aligned} P^{000}(n) &= 3!\,\bar{P}^{000}(n) \\ P^{001}(n) &= 2!\,(\bar{P}^{001}(n) + \bar{P}^{010}(n) + \bar{P}^{100}(n)) \\ P^{011}(n) &= 2!\,(\bar{P}^{011}(n) + \bar{P}^{101}(n) + \bar{P}^{110}(n)) \\ P^{111}(n) &= 3!\,\bar{P}^{111}(n). \end{aligned}$$

□

A more compact form of the above identity is possible if one associates a family of polynomials in $\mathbb{R}\langle X\rangle$ with each $c_u \in \mathbb{R}^m[[X_0]]$:

$$P_{c_u}(n) := x_0^n + \sum_{k=1}^{n} \sum_{S_k} \frac{1}{s_1! \cdots s_p!} P^{j_1 \cdots j_k}_{i_1 \cdots i_k}(n)\,(c_{u_{i_1}}, x_0^{j_1}) \cdots (c_{u_{i_k}}, x_0^{j_k}), \tag{3.54}$$

$n \geq 0$. Clearly, $\deg(P_{c_u}(n)) = n$ and

$$(c \circ c_u, x_0^n) = (c, P_{c_u}(n)), \quad n \geq 0.$$

Therefore,

$$c \circ c_u = \sum_{n=0}^{\infty} (c, P_{c_u}(n))\, x_0^n.$$

Example 3.44 Continuing the previous example, it follows directly from the definition of $P_{c_u}(n)$ that

$$\begin{aligned} P_{c_u}(0) &= 1 \\ P_{c_u}(1) &= x_0 + x_1 (c_u, \emptyset) \\ P_{c_u}(2) &= x_0^2 + x_1 (c_u, x_0) + (x_0 x_1 + x_1 x_0)(c_u, \emptyset) + x_1^2 (c_u, \emptyset)^2 \\ P_{c_u}(3) &= x_0^3 + x_1 (c_u, x_0^2) + (x_0 x_1 + 2 x_1 x_0)(c_u, x_0) + \end{aligned}$$

$$3x_1^2(c_u,\emptyset)(c_u,x_0)+(x_0^2x_1+x_1x_0^2+x_0x_1x_0)(c_u,\emptyset)+$$
$$(x_0x_1^2+x_1x_0x_1+x_1^2x_0)(c_u,\emptyset)^2+x_1^3(c_u,\emptyset)^3$$
$$\vdots$$

□

The next lemma combined with the previous one provides the core argument for the uniqueness of the generating series of a formal Fliess operator.

Lemma 3.10 *Let $X=\{x_1,x_2,\ldots,x_n\}$ and $p\in\mathbb{R}[X]$. Define the corresponding generating function on $\mathbb{R}^n$*

$$f_p(z)=\sum_{\eta\in X^*}(p,\eta)\frac{z^\eta}{\eta!}.$$

Then $f_p(z)=0$ for all $z\in\mathbb{R}^n$ if and only if $p=0$.

Proof: For any $\eta=x_{i_1}\cdots x_{i_k}\in X^*$ define the partial differentiation operator

$$\frac{\partial^\eta}{\partial z^\eta}=\frac{\partial^k}{\partial z_{i_1}\partial z_{i_2}\cdots\partial z_{i_k}}.$$

Assume the support of p is ordered. If $f_p(z)=0$ everywhere on $\mathbb{R}^n$ then it follows that

$$\left.\frac{\partial^\eta}{\partial z^\eta}f_p(z)\right|_{z=0}=(p,\eta)=0,\ \ \forall\eta\in\mathrm{supp}(p).$$

Thus, $p=0$. The converse claim is trivial. ■

Proof of Theorem 3.38: Since $c\circ c_u=d\circ c_u$ is equivalent to $(c-d)\circ c_u=0$, it is sufficient to prove that if $c\circ c_u=0$ for all $c_u\in\mathbb{R}^m[[X_0]]$ then $c=0$. From $c\circ 0=0$ it follows directly using Lemma 3.9 that $(c,x_0^n)=0$ for all $n\geq 0$. Thus, it is only necessary to show that $(c,\eta)=0$ for every $\eta\in\Gamma_k$, $k\geq 1$. This fact is proved by contradiction. That is, suppose $c\neq 0$ but $c\circ c_u=0$ for all $c_u\in\mathbb{R}^m[[X_0]]$. For any fixed $n\geq 1$, it is immediately evident from Lemma 3.9 that $(c\circ c_u,x_0^n)$ is a polynomial in the ordered variables $z_\ell:=(c_{u_{i_\ell}},x_0^{j_\ell})$, $\ell=1,\ldots,n$ with coefficients proportional to $\left(c,P_{i_1\cdots i_k}^{j_1\cdots j_k}(n)\right)$, $k\leq n-j$. In which case, an ordered alphabet can be introduced so that this polynomial can be represented exactly as p in Lemma 3.10. Since $f_p(z) = (c \circ c_u, x^n) = 0$ for any z (i.e, any c_u), it follows from the lemma that

$$\left(c, P_{i_1\cdots i_k}^{j_1\cdots j_k}(n)\right) = 0 \tag{3.55}$$

for every $n \geq 1$ and any set of indices $\binom{j_1\cdots j_k}{i_1\cdots i_k} \in S_k$. This by itself, however, is not sufficient to conclude that $c = 0$. Suppose there exists for some fixed $\bar{k} \geq 1$ a word $\eta_0 \in \Gamma_{\bar{k}}$ such that $(c, \eta_0) \neq 0$. Define a corresponding language

$$\Omega = \{\eta \in \Gamma_{\bar{k}} : |\eta|_{x_i} = |\eta_0|_{x_i}\,,\ i = 0, 1, \ldots, m\},$$

which is comprised of all words which are permutations of the letters of η_0. With Ω one can associate a nonzero polynomial

$$Q = \sum_{\eta \in \Omega} (c, \eta)\, \eta.$$

Observe for any $n \geq 0$ and $d_u \in \mathbb{R}^m[[X_0]]$ that

$$\begin{aligned}
&(Q \circ d_u, x_0^n) \\
&\quad = \sum_{k=1}^{n} \sum_{S_k} \frac{1}{s_1! \cdots s_p!} \left(Q, P_{i_1\cdots i_k}^{j_1\cdots j_k}(n)\right) (d_{u_{i_1}}, x_0^{j_1}) \cdots (d_{u_{i_k}}, x_0^{j_k}) \\
&\quad = \sum_{k=1}^{\infty} \sum_{S_k} \sum_{\eta \in \Omega} \frac{(c, \eta)}{s_1! \cdots s_p!} \left(\eta, P_{i_1\cdots i_k}^{j_1\cdots j_k}(n)\right) (d_{u_{i_1}}, x_0^{j_1}) \cdots (d_{u_{i_k}}, x_0^{j_k}) \\
&\quad = \sum_{\tilde{S}_{\bar{k}}} \frac{1}{s_1! \cdots s_p!} \left(c, P_{i_1\cdots i_{\bar{k}}}^{j_1\cdots j_{\bar{k}}}(n)\right) (d_{u_{i_1}}, x_0^{j_1}) \cdots (d_{u_{i_{\bar{k}}}}, x_0^{j_{\bar{k}}}) \\
&\quad = 0
\end{aligned}$$

using (3.55), the definition of Q, and letting

$$\begin{aligned}
\tilde{S}_{\bar{k}} := \Bigg\{ \begin{pmatrix} j_1 & j_2 & \cdots & j_{\bar{k}} \\ i_1 & i_2 & \ldots & i_{\bar{k}} \end{pmatrix} \in S_{\bar{k}} : j = n - |\eta_0|\,, \\
\eta_0 = x_0^{n_{\bar{k}}} x_{i_{\bar{k}}} \cdots x_0^{n_1} x_{i_1} x_0^{n_0} \Bigg\}.
\end{aligned}$$

But since Q is a polynomial, it is locally convergent. It is already known from Theorem 3.19 that if $Q \circ d_u = 0$ for all locally convergent

d_u in $\mathbb{R}^m[[X_0]]$ then $Q=0$. But this contradicts the assumed property that $Q \neq 0$, which had followed from the assumption that $(c,\eta_0) \neq 0$. Hence, $c=0$, and the theorem is proved. ■

In Chapter 1, the notion of the formal Laplace-Borel transform of a function was introduced. Namely, given a function $u:\mathbb{R}\rightarrow\mathbb{R}$, which is either analytic at a point $t_0 \in \mathbb{R}$ or is a function in the formal sense, one can construct its generating series

$$c_u = \sum_{n=0}^{\infty} (c_u, x_0^n)\, x_0^n$$

directly from its Taylor series expansion. The formal Laplace transform in this setting is the mapping

$$\mathscr{L}_f : u \mapsto c_u, \tag{3.56}$$

and its inverse is the formal Borel transform. In light of Theorems 3.5 and 3.38, an analogous definition is possible for any Fliess operator, convergent or formal. Slightly abusing the notation, elements in $\mathscr{F}$ will be written as F_c.

Definition 3.10 *The* ***formal Laplace transform*** *on $\mathscr{F}$ is defined as the mapping*

$$\begin{aligned} \mathscr{L}_f &: \mathscr{F} \rightarrow \mathbb{R}^{\ell}\langle\langle X\rangle\rangle \\ &: F_c \mapsto c. \end{aligned}$$

The corresponding inverse transform, the ***formal Borel transform*** *on $\mathbb{R}^{\ell}\langle\langle X\rangle\rangle$, is*

$$\begin{aligned} \mathscr{B}_f &: \mathbb{R}^{\ell}\langle\langle X\rangle\rangle \rightarrow \mathscr{F} \\ &: c \mapsto F_c. \end{aligned}$$

Note that when $m=0$, this definition is consistent with (3.56) in the sense that a given function u can be represented as a constant operator F_{c_u}, that is, $u = F_{c_u}[v]$ for any signal v, and $\mathscr{L}_f[u] = c_u = \mathscr{L}_f[F_{c_u}]$. In fact, as the following example shows, this point of view is implicit in the classical treatment of linear time-invariant systems.

Example 3.45 Consider a causal linear integral operator

$$y(t) = \int_0^t h(t-\tau)u(\tau)\,d\tau,$$

where the kernel function h is analytic at $t = 0$. This operator is completely characterized by the integral Laplace transform of it kernel function, namely its *transfer function* $H(s) := \mathscr{L}[h](s) = \sum_{k>0}(c_h, x_0^k)s^{-k}$. In terms of the formal Laplace transform, this is equivalent to taking the transform of a *signal*, specifically, $c_h = \mathscr{L}[h]$, as opposed to the *operator* $u \mapsto y$. On the other hand, the formal Laplace transform of the operator is the linear series $c = \sum_{j\geq 0}(c_h, x_0^k)x_0^k x_1 = c_h x_1$ (see Section 1.3). □

It is next shown that many of the familiar properties of the integral Laplace transform also have their formal counterparts. To facilitate the analysis, the concept of a generalized series appearing in the previous section is further refined.

Definition 3.11 *A **Dirac series**, δ_i, is a generalized series with the defining property that $F_{\delta_i}[u] = u_i(t)$, or equivalently, $\delta_i \circ c_u = c_{u_i}$, $i = 1, 2, \ldots, m$.*

Theorem 3.39 *For any $c, d \in \mathbb{R}^\ell\langle\langle X\rangle\rangle$, the following identities hold:*

1. *Linearity*

$$\mathscr{L}_f\left[\alpha F_c + \beta F_d\right] = \alpha\mathscr{L}_f\left[F_c\right] + \beta\mathscr{L}_f\left[F_d\right]$$
$$\mathscr{B}_f\left[\alpha c + \beta d\right] = \alpha\mathscr{B}_f\left[c\right] + \beta\mathscr{B}_f\left[d\right],$$

   where $\alpha, \beta \in \mathbb{R}$.

2. *Integration*

$$\mathscr{L}_f\left[I^n F_c\right] = x_0^n c$$
$$\mathscr{B}_f\left[x_0^n c\right] = I^n F_c,$$

   where $I(\cdot)$ denotes the formal integration operator.

3. *Differentiation*

$$\mathscr{L}_f\left[DF_c\right] = x_0^{-1}(c) + \sum_{i=1}^m \delta_i \sqcup\!\sqcup \left(x_i^{-1}(c)\right)$$

$$\mathscr{B}_f\left[x_0^{-1}(c) + \sum_{i=1}^m \delta_i \sqcup\!\sqcup \left(x_i^{-1}(c)\right)\right] = DF_c,$$

where $D(\cdot)$ is the formal differentiation operator. If x_0^n is a left factor of c, that is, $c = x_0^n c'$ for some $c' \in \mathbb{R}^\ell\langle\langle X\rangle\rangle$, then

$$\begin{aligned}\mathscr{L}_f\left[D^n F_c\right] &= x_0^{-n}(c)\\ \mathscr{B}_f\left[x_0^{-n}(c)\right] &= D^n F_c.\end{aligned}$$

4. Products

$$\begin{aligned}\mathscr{L}_f\left[F_c F_d\right] &= \mathscr{L}_f\left[F_c\right] ⧢ \mathscr{L}_f\left[F_d\right]\\ \mathscr{B}_f\left[c ⧢ d\right] &= \mathscr{B}_f\left[c\right]\mathscr{B}_f\left[d\right].\end{aligned}$$

Proof: The properties of linearity and integration are trivial to verify. The product property follows directly from Theorem 3.11. Thus, only the differentiation property requires some justification. It was shown in Theorem 3.4 that the derivative of a convergent Fliess operator is

$$\begin{aligned}\frac{d}{dt}F_c[u](t) &= F_{x_0^{-1}(c)}[u](t) + \sum_{i=1}^m u_i(t)F_{x_i^{-1}(c)}[u](t)\\ &= F_{x_0^{-1}(c)+\sum_{i=1}^m \delta_i ⧢ (x_i^{-1}(c))}[u].\end{aligned}$$

For a formal operator, one can easily show that the composition product satisfies the identity

$$\begin{aligned}x_0^{-1}(c\circ c_u) &= x_0^{-1}(c)\circ c_u + \sum_{i=1}^m c_{u_i} ⧢ [x_i^{-1}(c)\circ c_u]\\ &= \left(x_0^{-1}(c) + \sum_{i=1}^m \delta_i ⧢ (x_i^{-1}(c))\right)\circ c_u\end{aligned}$$

(see Problem 2.7.7). Thus, the first pair of identities in part *3* is established. Now if x_0 is a left factor of c, then $F_{x_i^{-1}(c)}[u](t) = 0$ for $i = 1, 2, \cdots, m$. In this case, $\frac{d}{dt}F_c[u](t) = F_{x_0^{-1}(c)}[u](t)$. Proceeding inductively, the second pair of identities follow. ■

Example 3.46 Consider a generalization of the Wiener system in Example 3.7:

$$y(t) = \exp\left(\int_0^t u_1(\tau) + u_2(\tau)\,d\tau\right).$$

Setting $X = \{x_1, x_2\}$, observe that

$$\begin{aligned} y(t) &= \sum_{k=0}^{\infty} \frac{1}{k!} \left(E_{x_1}[u](t) + E_{x_2}[u](t)\right)^k \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} \left(E_{(x_1+x_2)^{\sqcup\!\sqcup k}}[u](t)\right) \\ &= \sum_{k=0}^{\infty} E_{(x_1+x_2)^k}[u](t) \\ &= F_c[u](t), \end{aligned}$$

where $c = (x_1 + x_2)^* := \sum_{k\geq 0}(x_1 + x_2)^k$. Therefore,

$$\mathscr{L}_f[F_c] = (x_1 + x_2)^*.$$

This result can be viewed as an operator version of the integral transform pair

$$e^t \stackrel{\mathscr{L}}{\Longleftrightarrow} (1 - s)^{-1}.$$

Other formal Laplace-Borel transform pairs are given in Table 3.4. □

Example 3.47 Suppose F_c has the generating series $c = \mathrm{char}(X)$. For any fixed word $\eta \in X^*$

$$\begin{aligned} \mathscr{L}_f[F_c E_\eta] &= \mathscr{L}_f[F_c] \sqcup\!\sqcup \mathscr{L}_f[E_\eta] \\ &= c \sqcup\!\sqcup \eta \\ &= \sum_{\nu\in X^*} \binom{\nu}{\eta} \nu, \end{aligned}$$

where $\binom{\nu}{\eta}$ denotes the number of subwords of ν which are equal to η (see Problem 2.4.4). For example, if $X = \{x_0, x_1\}$ and $\eta = x_1x_0$ then

$$\begin{aligned} c \sqcup\!\sqcup \eta &= (x_0 + x_1) \sqcup\!\sqcup x_1x_0 \\ &= x_0x_1x_0 + 2x_1x_0x_0 + 2x_1x_1x_0 + x_1x_0x_1 \\ &= \binom{x_0x_1x_0}{x_1x_0} x_0x_1x_0 + \binom{x_1x_0x_0}{x_1x_0} x_1x_0x_0 + \binom{x_1x_1x_0}{x_1x_0} x_1x_1x_0 + \\ &\quad \binom{x_1x_0x_1}{x_1x_0} x_1x_0x_1. \end{aligned}$$

□

Table 3.4. Elementary formal Laplace-Borel transform pairs.

F_c	$\mathscr{L}_f[F_c]$
$u \mapsto 1$	1
$u \mapsto t^n$	$n!\, x_0^n$
$u \mapsto \left(\sum_{i=0}^{n-1} \binom{i}{n-1} \frac{(at)^i}{i!}\right) e^{at}$	$(1-ax_0)^{-n}$
$u \mapsto \frac{1}{n!}\left(\int_{t_0}^t \sum_{j=1}^k u_{i_j}(\tau)d\tau\right)^n$	$(x_{i_1}+x_{i_2}+\cdots+x_{i_k})^n$
$u \mapsto \sum_{n\geq 0} \frac{a_n}{n!}\left(\int_{t_0}^t \sum_{j=1}^k u_{i_j}(\tau)d\tau\right)^n$	$\sum_{n\geq 0} a_n(x_{i_1}+x_{i_2}+\cdots+x_{i_k})^n$
$u \mapsto e^{\int_{t_0}^t \sum_{j=1}^k u_{i_j}(\tau)d\tau}$	$(x_{i_1}+x_{i_2}+\cdots+x_{i_k})^*$
$u \mapsto \int_{t_0}^t \sum_{j=1}^k u_{i_j}(\tau)d\tau\, e^{\int_{t_0}^t \sum_{j=1}^k u_{i_j}(\tau)d\tau}$	$\frac{x_{i_1}+x_{i_2}+\cdots+x_{i_k}}{[1-(x_{i_1}+x_{i_2}+\cdots+x_{i_k})]^2}$
$u \mapsto \cos\left(\int_{t_0}^t \sum_{j=1}^k u_{i_j}(\tau)d\tau\right)$	$\frac{1}{1+(x_{i_1}+x_{i_2}+\cdots+x_{i_k})^2}$
$u \mapsto \sin\left(\int_{t_0}^t \sum_{j=1}^k u_{i_j}(\tau)d\tau\right)$	$\frac{x_{i_1}+x_{i_2}+\cdots+x_{i_k}}{1+(x_{i_1}+x_{i_2}+\cdots+x_{i_k})^2}$

Example 3.48 Let $X = \{x_1, x_2\}$,

$$y(t) = \cos\left(\int_0^t u_1(\tau) + u_2(\tau)\, d\tau\right),$$

and $d = (d_1, d_2) \in \mathbb{R}^2\langle\langle X\rangle\rangle$ be arbitrary. From Table 3.4 it follows that $y = F_c[u]$, where

$$c = \frac{1}{1+(x_1+x_2)^2} := \sum_{k=0}^{\infty}(-1)^k(x_1+x_2)^{2k}.$$

Setting $F_e = F_c \circ F_d$, the formal Laplace transform of F_e is then

$$\mathscr{L}_f[F_e] = c \circ d = \sum_{k=0}^{\infty}(-1)^k(x_1+x_2)^{2k} \circ d.$$

□

Example 3.49 For any $c \in \mathbb{R}^m\langle\langle X\rangle\rangle$, $x_i \in X$ with $i = 1, 2, \cdots, m$, and $n \geq 0$, recall that

$$x_i^n \circ c = \frac{1}{n!}(x_0 c_i)^{\sqcup\!\sqcup\, n}$$

(see Problem 2.7.7(a)). Applying the formal Borel transform to both sides of this identity gives

$$\begin{aligned}\mathscr{B}_f\left[x_i^n \circ c\right] &= \mathscr{B}_f\left[\frac{1}{n!}(x_0 c_i)^{\sqcup\!\sqcup\, n}\right] \\ &= \frac{1}{n!}\left(\mathscr{B}_f[x_0 c_i]\right)^n \\ &= \frac{1}{n!}\left(\int_0^t F_{c_i}[u](\tau)\, d\tau\right)^n.\end{aligned}$$

□

Example 3.50 Consider the linear differential equation

$$\frac{d^n y(t)}{dt^n} + \sum_{i=0}^{n-1} a_i \frac{d^i y(t)}{dt^i} = \sum_{i=0}^{n-1} b_i \frac{d^i u(t)}{dt^i}$$

with initial conditions $y^{(i)}(0) = 0$, $u^{(i)}(0) = 0$, and where $a_i, b_i \in \mathbb{R}$, $i = 0, 1, \ldots, n-1$ for $n \geq 1$. The goal is to construct a series solution $y = F_c[u]$ for some $c \in \mathbb{R}\langle\langle X\rangle\rangle$. First, integrate both sides of the equation n times and then apply the formal Laplace transform to get

$$\left(\delta + \sum_{i=0}^{n-1} a_i\, x_0^{n-1-i} x_1\right) \circ c = \sum_{i=0}^{n-1} b_i\, x_0^{n-1-i} x_1,$$

or equivalently,

$$\left(1 + \sum_{i=0}^{n-1} a_i\, x_0^{n-i}\right) c = \sum_{i=0}^{n-1} b_i\, x_0^{n-1-i} x_1.$$

Therefore,

$$c = \left(1 + \sum_{i=0}^{n-1} a_i\, x_0^{n-i}\right)^{-1} \sum_{i=0}^{n-1} b_i\, x_0^{n-1-i} x_1.$$

Rephrased in the language of the integral Laplace transform, this is equivalent to

$$\begin{aligned} Y(s) &= \left(1 + \sum_{i=0}^{n-1} a_i \frac{1}{s^{n-i}}\right)^{-1} \left(\sum_{i=0}^{n-1} b_i \frac{1}{s^{n-i}}\right) U(s) \\ &= \left(s^n + \sum_{i=0}^{n-1} a_i s^i\right)^{-1} \left(\sum_{i=0}^{n-1} b_i s^i\right) U(s). \end{aligned}$$

□

Example 3.51 Consider the nonlinear differential equation

$$\frac{d^n y(t)}{dt^n} + \sum_{i=0}^{n-1} a_i \frac{d^i y(t)}{dt^i} + \sum_{j=2}^{k} p_j\, u(t) y^j(t) = \sum_{i=0}^{n-1} b_i \frac{d^i u(t)}{dt^i}$$

with $y^{(i)}(0) = 0$, $u^{(i)}(0) = 0$, and where $a_i, b_i, p_j \in \mathbb{R}$, $i = 0, 1, \ldots, n-1$ and $j = 2, \ldots, k$ for $n \geq 1$, $k \geq 2$. As in the previous example, integrate both side of the equation n times and assume $y = F_c[u]$. Applying the formal Laplace transform in this case gives

$$\left(1 + \sum_{i=0}^{n-1} a_i\, x_0^{n-i}\right) c + \sum_{j=2}^{k} p_j\, x_0^{n-1} x_1 (c^{\sqcup\!\sqcup\, j}) = \sum_{i=0}^{n-1} b_i\, x_0^{n-1-i} x_1.$$

Defining

$$d_j = -\left(1 + \sum_{i=0}^{n-1} a_i\, x_0^{n-i}\right)^{-1} p_j x_0^{n-1} x_1$$

and

$$c_1 = \left(1 + \sum_{i=0}^{n-1} a_i\, x_0^{n-i}\right)^{-1} \left(\sum_{i=0}^{n-1} b_i\, x_0^{n-1-i} x_1\right),$$

it follows that

$$c = \sum_{j-2}^{k} d_j (c^{\sqcup\!\sqcup\, j}) + c_1.$$

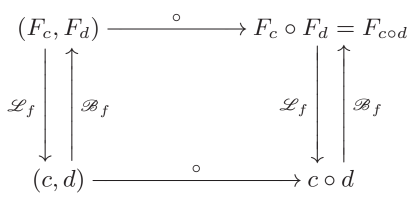


Fig. 3.21. The monoid isomorphism $\mathscr{L}_f$ between $(\mathscr{F}, \circ, I)$ and $(\mathbb{R}^m\langle\langle X \rangle\rangle, \circ, \delta)$.

An inductive solution is given by $c = \lim_{n\to\infty} c_n$, where

$$c_{n+1} = \sum_{j-2}^{k} d_j(c_n^{\sqcup\!\sqcup\, j}) + c_1, \quad n \geq 1,$$

provided the limit exists. This can be verified using the fact that the d_j and c_1 are all proper. □

Given two linear integral operators with kernel functions h_1 and h_2 defined on $[0, +\infty)$, respectively, their composition has the kernel function $h_1 * h_2$, where $*$ denotes the usual convolution product on the set of real-valued functions. The set of all such functions forms a monoid M if the Dirac delta function, δ, is admitted as the unit. The integral Laplace transform satisfies the well known identity

$$\mathscr{L}[h_1 * h_2](s) = \mathscr{L}[h_1](s)\mathscr{L}[h_2](s) = H_1(s)H_2(s)$$

and maps M to the monoid of corresponding transfer functions under the pointwise product in $\mathbb{C}$ with $\mathscr{L}(\delta) = 1$ (see Problem 2.1.1). This section is concluded by observing that the formal Laplace transform acts analogously as a monoid isomorphism in the context of Fliess operators.

Theorem 3.40 *For any $c \in \mathbb{R}^\ell\langle\langle X \rangle\rangle$ and $d \in \mathbb{R}^m\langle\langle X \rangle\rangle$:*

$$\mathscr{L}_f(F_c \circ F_d) = \mathscr{L}_f(F_c) \circ \mathscr{L}_f(F_d)$$
$$\mathscr{B}_f(c \circ d) = \mathscr{B}_f(c) \circ \mathscr{B}_f(d).$$

Proof: The proof follows directly from the definitions. For any $F_c, F_d \in \mathscr{F}$

$$\mathscr{L}_f(F_c \circ F_d) = \mathscr{L}_f(F_{c\circ d}) = c \circ d$$

$$= \mathscr{L}_f(F_c) \circ \mathscr{L}_f(F_d).$$

Similarly, for any formal power series c and d,

$$\begin{aligned}\mathscr{B}_f(c \circ d) &= F_{c\circ d} = F_c \circ F_d \\ &= \mathscr{B}_f(c) \circ \mathscr{B}_f(d).\end{aligned}$$

■

Setting $\ell = m$ and recalling that $F_\delta = I$ acts as the unit on $\mathscr{F}$, the commutative diagram in Figure 3.21 shows the monoid isomorphism relating the monoids $(\mathscr{F}, \circ, I)$ and $(\mathbb{R}^m\langle\langle X\rangle\rangle, \circ, \delta)$.

Problems

Section 3.1

Problem 3.1.1 Let t_0 and t_1 be fixed real numbers with $t_0 < t_1$.

(a) Show that if $u \in L^m_{\mathfrak{p}+1}[t_0, t_1]$ for any $\mathfrak{p} \in [1, \infty)$ then $u \in L^m_{\mathfrak{p}}[t_0, t_1]$. In addition, show that every $u \in L^m_\infty[t_0, t_1]$ is also in $L^m_{\mathfrak{p}}[t_0, t_1]$ for every $\mathfrak{p} \in [1, \infty)$.
(b) Repeat part (a) for the extended spaces $L^m_{\mathfrak{p},e}(t_0)$ and $L^m_{\infty,e}(t_0)$.

Problem 3.1.2 Prove that $L^m_{\mathfrak{p}}[t_0, \infty) \subset L^m_{\mathfrak{p},e}(t_0)$ for any $p \in [1, \infty]$.

Section 3.2

Problem 3.2.1 Show that for any $\mathfrak{p} \in (1, \infty]$ and any finite interval $[t_0, t_0 + T]$

$$\|u\|_1 \leq \|u\|_{\mathfrak{p}} \, T^{\frac{1}{\mathfrak{q}}},$$

when $u \in L^m_{\mathfrak{p}}[t_0, t_0 + T]$, and $\mathfrak{p}$ and $\mathfrak{q}$ are conjugate exponents.
Remark: Consider a subset $\Omega \subset \mathbb{R}$ and two functions $u, v : \Omega \to \mathbb{R}^m$, where $\|u\|_{\mathfrak{p}}$ and $\|v\|_{\mathfrak{q}}$ are well defined in the sense that

$$\|u\|_{\mathfrak{p}} = \left(\int_\Omega |u(t)|^{\mathfrak{p}} \, dt\right)^{\frac{1}{\mathfrak{p}}} < \infty.$$

Hölder's inequality states that

$$\int_{\Omega} \left|u^T(t)v(t)\right| \; dt \leq \|u\|_{\mathfrak{p}} \|v\|_{\mathfrak{q}}$$

when $\mathfrak{p}$ and $\mathfrak{q}$ are conjugate exponents. The special case where $\mathfrak{p} = \mathfrak{q} = 2$ is known as the Schwarz inequality.

Problem 3.2.2 Reconsider the Wiener system in Example 3.3.

(a) Determine the generating series c assuming $z(0) = a$.
(b) For what values of a is c well defined?

Problem 3.2.3 A function $f : \mathbb{R} \to \mathbb{R}$ is said to satisfy a *Lipschitz condition* on $J = [a, b]$ if there exists a constant $L > 0$ such that

$$|f(z_1) - f(z_2)| \leq L\,|z_1 - z_2|\,, \;\; \forall z_1, z_2 \in J.$$

(a) Show that if f satisfies a Lipschitz condition on J then it is absolutely continuous on J.
(b) Show that if f is continuous on J and has a bounded derivative on (a, b), i.e., $|f'(z)| < M$ everywhere on (a, b) for some finite M, then f satisfies a Lipschitz condition on J.
(c) What can one conclude about f when $0 < L < 1$?

Remark: The mean value theorem is useful in this problem.

Problem 3.2.4 Consider a function $f : \mathbb{R} \to \mathbb{R}$.

(a) Show that if f is absolutely continuous on $J = [a, b]$ then it is continuous on J.
(b) Show that if f is *continuously* differentiable on J then it is absolutely continuous on J.

Problem 3.2.5 Let $X = \{x_0, x_1, \ldots, x_m\}$, $c \in \mathbb{R}_{LC}\langle\langle X\rangle\rangle$, and $u_i \in C^1[t_0, t_0 + T]$ for $i = 1, 2, \ldots, m$. Define $y = F_c[u]$.

(a) Derive a formula for d^2y/dt^2.
(b) Suppose $m = 1$, and c is a linear series. Compute dy/dt and d^2y/dt^2.

Problem 3.2.6 Suppose $c = (c, x_ix_j)x_ix_j + (c, x_jx_i)x_jx_i$.

(a) For the input $\bar{u}$ shown in Figure 3.4 compute $F_c[\bar{u}](t_1+t_2)$ assuming $t_0 = 0$.
(b) Compute $G(c, \alpha) := \left.\dfrac{\partial^2}{\partial t_1 \partial t_2} F_c[\bar{u}](t_1 + t_2)\right|_{t_1=t_2=0^+}$.

(c) Compute $\dfrac{\partial^2 G}{\partial\alpha_{i2}\partial\alpha_{j1}}$ and $\dfrac{\partial^2 G}{\partial\alpha_{j2}\partial\alpha_{i1}}$. Is the result what is expected? Explain.

(d) Suppose now that $x_i = x_0$ so that $\alpha_{0k} = 1$, $k = 1,2$. If partials $\partial/\partial\alpha_{0k}$ are not taken in part (c), what is the result?

Problem 3.2.7 Consider a series of functions

$$f(z) = f_1(z) + f_2(z) + \dots,$$

where each $f_i : W_{\mathbb{C}} \to \mathbb{C}$ is an analytic function on a region $W_{\mathbb{C}}$ in the complex plane. Show that if the series converges uniformly on every compact subset of $W_{\mathbb{C}}$ then f is also analytic on $W_{\mathbb{C}}$.
Remark: This is a standard problem in complex analysis. It can be solved by forming the partial sums $s_n = \sum_{i=1}^{n} f_i$ and applying Morera's theorem. For more information, see, for example, Chapter 5 of [2].

Section 3.3

Problem 3.3.1 For each formal power series c over X^* with coefficients given below, determine its minimal Gevrey order and whether it is locally convergent, globally convergent, or neither. When appropriate determine at least one pair of real numbers $R, T > 0$ such that the Fliess operator F_c is well defined on $B_{\mathfrak{p}}(R)[0,T]$.

(a) $(c,\eta) = 2^{|\eta|}\sin(\frac{\pi}{2}\,|\eta|)$
(b) $(c,\eta) = 2^{|\eta|+1}$
(c) $(c,\eta) = |\eta|^2$, $\eta \in X^*$
(d) $(c,\eta) = 5^{|\eta|}(|\eta|+1)!$
(e) $(c,\eta) = |\eta|^{|\eta|}$
(f) $(c,\eta) = (|\eta|!)^2$

Remark: Stirling's approximation formula

$$k! \sim \sqrt{2\pi k}\, k^k \mathrm{e}^{-k}, \quad k \gg 1$$

is useful for part (e).

Problem 3.3.2 Provide an example, if possible, for each scenario below concerning linear series (see (2.48)). If no such example exists, give a justification.

(a) A series c which is linear but not globally convergent.
(b) A series d which is globally convergent but not linear.

Section 3.4

Problem 3.4.1 Express the input-output mapping of the Wiener system in Example 3.7 as a Volterra operator.

Problem 3.4.2 If the order of the blocks in a Wiener system are reversed, the resulting system is called a *Hammerstein system*. An example of a single-input, single-output Hammerstein system is shown in Figure 3.22. Suppose $g(u) = u^2$.

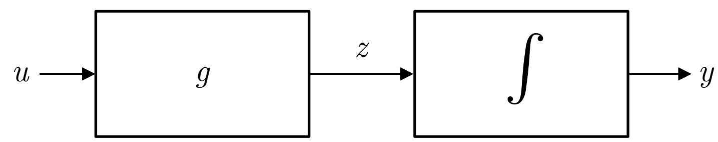


Fig. 3.22. The Hammerstein system in Problem 3.4.2.

(a) Determine a Volterra operator representation of the input-output map $u \mapsto y$.
(b) Determine a Fliess operator representation of $u \mapsto y$.
(c) Suppose $u(t) = \int_0^t v(\tau)\, d\tau$. Determine a Fliess operator representation of $v \mapsto y$. This is an example of what is called *dynamic extension*.

Section 3.5

Problem 3.5.1 Provide a proof for Theorem 3.13 for the special case where $s_c^* = s_d^*$.

Problem 3.5.2 Show that if $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{m\times m}$, then

$$(A \oplus B)^k = \sum_{j=0}^{k} (A^{k-j} \otimes B^j) \binom{k}{j}.$$

Remark: See the references [16, 15, 17, 18] for various elementary identities involving the Kronecker product.

Problem 3.5.3 Show that if $d \in \mathbb{R}_{LC}\langle\langle X\rangle\rangle$ with growth constants $K_d, M_d > 0$, then for any $k \geq 1$:

$$\left|(d^{\,\sqcup\!\sqcup\, k}, \nu)\right| \leq K_d^k M_d^{|\nu|} \frac{(|\nu|+k-1)!}{(k-1)!}, \;\; \forall \nu \in X^*.$$

Show that this implies the more generous upper bound

$$\left|(d^{\,\sqcup\!\sqcup\, k}, \nu)\right| \leq K_d^k (kM_d)^{|\nu|}\, |\nu|!, \;\; \forall \nu \in X^*.$$

What kind of system interconnection could produce the generating series $d^{\,\sqcup\!\sqcup\, k}$?
Remark: See Problem 2.4.8.

Problem 3.5.4 Show that $(1-x_1)^{\,\sqcup\!\sqcup\, -1} = \sum_{k\geq 0} k! x_1^k$, where $c^{\,\sqcup\!\sqcup\, -1}$ denotes the shuffle inverse of $c \in \mathbb{R}\langle\langle X\rangle\rangle$ as defined in Problem 2.4.11. Hence, the shuffle inverse does *not* preserve global convergence.
Remark: It is shown in [88] that the shuffle inverse does preserve local convergence.

Section 3.6

Problem 3.6.1 Verify the integer sequence identity (3.19) used in the analysis of the cascade connection of two linear time-invariant systems.
Remark: An inductive proof is possible, but an alternative approach is to use the integral formula

$$\int_0^1 (1-x)^{i-1} x^{j-1}\, dx = B(i,j),$$

where $B(x,y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$ is the beta function.

Problem 3.6.2 Consider the system shown in Figure 3.23 comprised of a harmonic oscillator with transfer function $H(s) = \omega_o/(s^2+\omega_o^2)$ followed by a saturation type function. Derive a Fliess operator representation of the input-output mapping $u \mapsto y$. State any required assumptions and discuss the convergence properties of the proposed model.

Problem 3.6.3 Let $X = \{x_0, x_1, \ldots, x_m\}$ be a fixed alphabet. Consider a Fliess operator F_c, $c \in \mathbb{R}^m\langle\langle X\rangle\rangle$, and a function $f_d : \mathbb{R}^m \to \mathbb{R}^m$

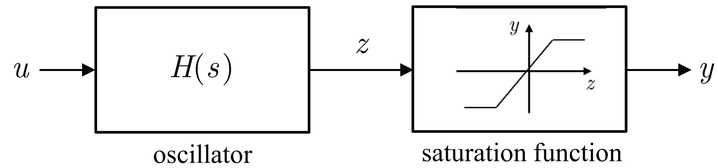


Fig. 3.23. The system with saturation considered in Problem 3.6.2.

with generating series $d \in \mathbb{R}^m\langle\langle X\rangle\rangle$ of the form $d = \sum_{x_j \in X}(d, x_j)x_j$. That is,

$$f_d(z_1, \ldots, z_m) = \sum_{j=0}^{m}(d, x_j)z_j,$$

with $z_0 := 1$.

(a) Show that the Hammerstein-Fliess system $F_c \circ f_d$ has the generating series

$$c \circ d = \sum_{\eta \in X^*}(c, \eta)d^\eta,$$

where $d^\emptyset := 1$ and $d_0 := 1$. That is, $F_c \circ f_d = F_{c \circ d}$.

(b) Is $c \circ d$ always well defined?

(c) Is it true that if c is locally convergent then so is $c \circ d$?

Problem 3.6.4 Consider the Fliess operators F_c and F_d, where $c = \sum_{k \geq 0} x_1^k$ and $d = x_1$.

(a) Verify that $(c \circ d, x_0^k x_1^k) = k!$, $k \geq 0$.

(b) Determine $F_{c \circ d}[u]$ and its convergence properties.

Problem 3.6.5

Let $X = \{x_0, x_1\}$. Assume $c \in \mathbb{R}\langle\langle X\rangle\rangle$ is a maximal series with growth constants $K, M > 0$ and Gevrey order s. Consider the input-output system $y = F_c[u]$.

(a) Determine a differential equation in terms of u, y, and its derivatives that is satisfied when $s = 1$. Be sure to include initial conditions.

(b) Repeat part (a) when $s = 0$.

(c) Repeat part (a) when $0 < s < 1$.

Problem 3.6.6 Verify the inequality

$$\sum_{k=0}^{n} \binom{n}{k}^{-1} \leq 2 + \frac{n-1}{n}, \quad n \geq 1$$

used in Example 3.23.

Problem 3.6.7 Consider two single-input, single-output Fliess operators F_c and F_d connected in series so that $y = (F_c \circ F_d)[u]$. Assume the generating series c and d are only locally convergent with minimal growth constants $K_c, M_c > 0$ and $K_d, M_d > 0$, respectively.

(a) What is the minimal geometric growth constant for the composite system?
(b) Is it possible for the composite system to have a well defined output over a *longer* interval of time than one or both of its subsystems? Either provide such an example or prove such an example does not exist.
(c) What are the *practical* consequences of the answer to part (b)?

Section 3.7

Problem 3.7.1 Let F_c and F_d be two Fliess operators with generating series $c, d \in \mathbb{R}^m_{LC}\langle\langle X \rangle\rangle$. Show that the feedback connection of these operators is always well-posed.
Remark: As with the cascade connection, Theorem 3.2 is useful here.

Problem 3.7.2 Consider the field of (irreducible) rational functions in $s \in \mathbb{C}$ with real coefficients denoted by $\mathbb{R}(s)$. Let $\mathbb{R}_{p\bar{0}}(s)$ be the subfield of proper elements g of $\mathbb{R}(s)$ with the defining property that $g(+\infty)$ exists and is not zero. Let $S = \mathbb{R}_{p0}(s)$ denote the ring of strictly proper elements h of $\mathbb{R}(s)$. Therefore, $h(+\infty) = 0$. Observe that $g \in \mathbb{R}_{p\bar{0}}(s)$ if and only if $g = K + h$ for some $K \neq 0$ and $h \in \mathbb{R}_{p0}(s)$. Show that $\mathbb{R}_{p\bar{0}}(s)$ acts freely from the right as a right transformation group on $\mathbb{R}_{p0}(s)$, where the product hg is defined in the usual fashion when $g, h \in \mathbb{R}(s)$.

Problem 3.7.3 For the series $c = 2x_1^2$ and $d = 3x_0 - x_1$ in $\mathbb{R}\langle X \rangle$, compute the following:

(a) $c \circ d$
(b) $c \,\tilde{\circ}\, d_\delta$
(c) $c_\delta \circ d_\delta$.

Problem 3.7.4 Prove the following propositions:

(a) For any $c \in \mathbb{R}^{\ell}_{LC}\langle\langle X\rangle\rangle$ and $d \in \mathbb{R}^{m}_{LC}\langle\langle X\rangle\rangle$, $F_c \circ (I + F_d) = F_{c\,\tilde{\circ}\, d_\delta}$. Recall this is the defining property of the mixed composition product given in Theorem 3.24.

(b) The mixed composition product provides an ultrametric contraction on $\delta + \mathbb{R}^m\langle\langle X\rangle\rangle$ as described in Theorem 3.25.

(c) For all $c, d, e \in \mathbb{R}^m\langle\langle X\rangle\rangle$, $(c \circ d)\,\tilde{\circ}\, e_\delta = c \circ (d\,\tilde{\circ}\, e_\delta)$.

Problem 3.7.5 Verify the following properties of the mixed composition product by checking their corresponding identities in the Fliess operator algebra:

(a) The following distributivity property holds for all $c, d, e \in \mathbb{R}^m_{LC}\langle\langle X\rangle\rangle$:

$$(c \sqcup\!\sqcup d)\,\tilde{\circ}\, e_\delta = (c\,\tilde{\circ}\, e_\delta) \sqcup\!\sqcup (d\,\tilde{\circ}\, e_\delta)$$

(cf. Problem 2.7.7(d)).

(b) For any $c, d, e \in \mathbb{R}^m_{LC}\langle\langle X\rangle\rangle$ it follows that

$$\begin{aligned}(c \circ d)\,\tilde{\circ}\, e_\delta &= c \circ (d\,\tilde{\circ}\, e_\delta)\\ (c\,\tilde{\circ}\, d_\delta)\,\tilde{\circ}\, e_\delta &= c\,\tilde{\circ}\,(d\,\tilde{\circ}\, e_\delta + e)_\delta.\end{aligned}$$

Problem 3.7.6 Consider the composition group $G = (\delta + \mathbb{R}\langle\langle X\rangle\rangle, \circ, \delta)$ with $X = \{x_0, x_1\}$.

(a) Show that G has a faithful representation $\pi : G \to GL(\mathbb{R}^\infty)$, where $\pi(c_\delta)$ is given by

$$\left[\begin{array}{c|c|cc|cccc|c}
1 & (c,\emptyset) & (c,x_0) & (c,x_1) & (c,x_0^2) & (c,x_0x_1) & (c,x_1x_0) & (c,x_1^2) & \cdots\\ \hline
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots\\ \hline
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \cdots\\
0 & 0 & (c,\emptyset) & 1 & (c,x_0) & (c,x_1) & 0 & 0 & \cdots\\ \hline
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \cdots\\
0 & 0 & 0 & 0 & (c,\emptyset) & 1 & 0 & 0 & \cdots\\
0 & 0 & 0 & 0 & (c,\emptyset) & 0 & 1 & 0 & \cdots\\
0 & 0 & 0 & 0 & (c,\emptyset)^2 & (c,\emptyset) & (c,\emptyset) & 1 & \cdots\\ \hline
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}\right].$$

The coefficients of c_δ are ordered lexicographically along the top row with $x_0 < x_1$, and the partitioning is done according to word length.

(b) Show that $\det(\pi(c_\delta)) = 1$, and that the matrices in this representation are upper triangular only in the case where $c_\delta = \delta + c$ with c proper. This case is most similar to the representation for the Faà di Bruno group G_{FdB} (see Problem 2.6.5).

Problem 3.7.7 Consider the feedback connection of two multivariable linear time-invariant systems as described in Example 3.28.

(a) Show that the feedback equation (3.29) has the solution

$$H_{cl} = (I - HG)^{-1}H = \sum_{k=0}^{\infty}(HG)^k H. \tag{3.57}$$

(b) Show that this solution is equivalent to (3.30).
(c) The feedback product as described in Theorem 3.28 is clearly the nonlinear analogue of (3.30). Is there a nonlinear version of (3.57)? If so, derive it. If not, explain why not.

Problem 3.7.8 Verify formula (3.44) for the largest coefficients in the Devlin polynomials.

Problem 3.7.9 Consider two feedback schemes involving proportional-integral-derivative (PID) compensation as shown in Figure 3.24, where $c \in \mathbb{R}\langle\langle X\rangle\rangle$ and $X = \{x_0, x_1\}$. Assume for the system in Figure 3.24(a) that

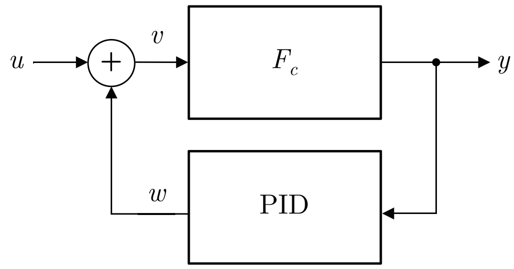
$$w(t) = K_P y(t) + K_I \int_0^t y(\tau)\, d\tau + K_D \frac{dy}{dt},$$

where K_P, K_I and K_D are fixed real numbers. An analogous expression holds for the mapping from w to v in the system shown in the other figure. Determine the generating series d, if possible, so that $w = F_d[v]$ in Figure 3.24(a) if all the PID coefficients are zero except for the following:

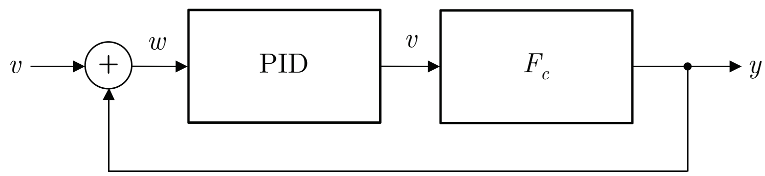
(a) K_P
(b) K_I
(c) K_D.

Repeat the exercise so that $y = F_d[w]$ in Figure 3.24(b).

Section 3.8



(a) PID compensator in the feedback loop.



(b) PID pre-compensator with unity feedback

Fig. 3.24. Two feedback schemes involving PID compensation in Problem 3.7.9.

Problem 3.8.1 Explicitly compute the formal Laplace transform of the following functions:

(a) $y(t) = 1$
(b) $y(t) = t^n/n!$
(c) $y(t) = e^{-\alpha t}$, $\alpha \in \mathbb{R}$
(d) $y(t) = cos(\omega t)$, $\omega \in \mathbb{R}$

Problem 3.8.2 For each $c \in \mathbb{R}\langle\langle X\rangle\rangle$ with $X = \{x_0, x_1\}$ below, compute the unit step response $y_s(t)$ of F_c on an interval $[0, T]$. Specify the largest possible T on which this output is well defined.

(a) $c = (1 - x_0)^{-1}x_1x_0$
(b) $c = (1 - x_1)^{-1}$
(c) $c = 1 + x_1^2$

Problem 3.8.3 For each input-output differential equation below, solve for y using the Laplace-Borel transform.

(a) $y' + ay = bu$, $y(0) = 0$, $a, b \in \mathbb{R}$
(b) $y' + y^2 = uy$, $y(0) = 0$

Problem 3.8.4 In many communication and control systems, a phase-locked loop (PLL) is used to synchronize an incoming external signal with an internally generated signal. For two sinusoids at the same carrier frequency, ω_c, this amounts to driving their phase difference to zero. A simple example of such a system is shown in Figure 3.25. The voltage controlled oscillator produces a sinusoid at the carrier

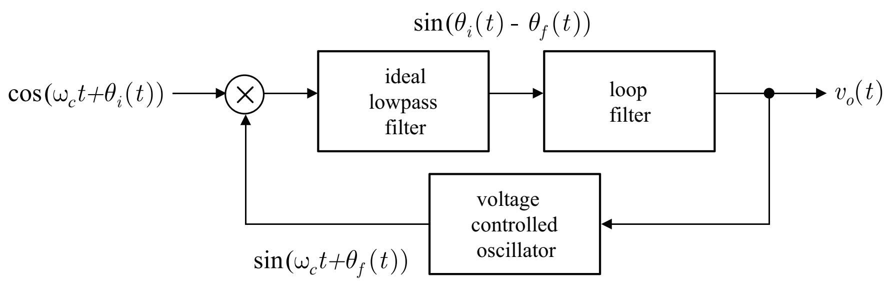


Fig. 3.25. Phase-locked loop in Problem 3.8.4.

frequency with phase

$$\theta_f(t) = \int_0^t v_o(\tau)\, d\tau.$$

For simplicity assume that the loop filter has transfer function $H(s) = 1$.

(a) Derive a differential equation in terms of the instantaneous input frequency $\omega_i = d\theta_i/dt$ and the phase difference $\Delta\theta = \theta_i - \theta_f$. The mapping $\omega_i \mapsto \Delta\theta$ describes how the PLL tracks the change in phase of the input signal.
(b) Assuming that there exists a $c \in \mathbb{R}\langle\langle X\rangle\rangle$ such that $F_c : \omega_i \mapsto \Delta\theta$, derive an algebraic equation that c must satisfy.
(c) How can c be determined from this equation?

Problem 3.8.5 Suppose the formal Laplace transform is extended to the group $\mathscr{F}_\delta$ so that $\mathscr{L}_f : \mathscr{F}_\delta \to \delta + \mathbb{R}^m\langle\langle X\rangle\rangle : F_{c_\delta} \mapsto c_\delta$. Is $\mathscr{L}_f$ a *group* isomorphism? Explain.

Bibliographic Notes

Section 3.1 Fliess operators were first described by M. Fliess in [68, 69, 71, 72, 74], primarily in the context of nonlinear realization theory. The foundation of Fliess's theory rests on the pioneering work by K.-T. Chen on iterated path integrals [31, 32, 33, 34, 35, 36, 37, 38, 40, 41, 42, 43] (see also [192]). The work of Sussmann in [184, 185, 186, 187] addressed important issues related to the convergence of Chen-Fliess series and introduced a formal product representation of a Fliess operator. Many subsequent enhancements and applications were introduced by Jakubczyk in [117, 118, 119, 120, 121, 122] and by Sontag and Wang in [180, 182, 200, 201, 202, 203] The Ph.D. dissertation of Wang [199] provides an excellent introduction to the subject.

Section 3.2 In the classical analysis of Fliess operators, it was normally assumed that the set of admissible inputs was $L_\infty^m[0,T]$ for some $T > 0$ [69, 72, 74, 117, 119, 184]. In later applications of this theory, however, it was more natural to consider other classes of inputs, for example, L_2 inputs [93, 170, 171]. This motivated the development of Theorem 3.1, which first appeared in [95]. The proof given here, however, is an improvement over the original in the sense that the region of convergence is less conservative. The primary innovation in this regard is Lemma 3.1, first proposed by Duffaut Espinosa in [54]. Wang provides in [199, Section 2.3] a detailed treatment of input-output properties of Fliess operators. Theorems 3.3, 3.4, 3.5, 3.6, and 3.7 are all based on this work. The uniqueness of the generating series, Theorem 3.5, was originally addressed in [69, 70, 73, 200]. See also Isidori's book [113, Chapter 3, Lemma 1.2]. Weierstrass's example concerning absolute continuity and Theorem 3.3 are covered in Chapter 11 of [191].

Section 3.3 The main results of this section, Theorems 3.8 and 3.9, have appeared in a number of special cases starting with the the original papers of Fliess, for example, [69]. Here the case where $s = 0$ and $p = \infty$ was presented, primary in the context of rational series. The condition on p was relaxed to $p \geq 1$ by Gray and Wang in [95]. The condition on the Gevrey order was generalized by Winter Arboleda et al. to $0 \leq s < 1$ in [206] (see also [205]).

Section 3.4 Volterra operators is a classical subject originating with V. Volterra in the latter part of the 19th century [197, 198]. As explained, for example, in [74], a variety of different approaches has been taken in the literature. The book by Schetzen gives a detailed treatment of the more traditional approach to the subject [172]. Volterra series as defined by Schetzen do not assume causality and require boundedness assumptions to ensure that the integrals over $(-\infty, \infty)$ are finite. In this setting, it is more natural to do steady-state analysis in the frequency domain using the multi-variable Fourier transform. Additional analytical results along these lines appear in the paper by Boyd et al. [14]. This approach was adapted to control theory in the book by Rugh using the multi-variable Laplace transform [166]. The class of Volterra series that can be written in terms of Chen-Fliess series, on the other hand, are inherently causal, frequently have only a finite radius of convergence, and are well suited for transient response analysis. In this context, the formal Laplace transform, or equivalently, the Laplace-Borel transform, is used for symbolic computation in lieu of the multi-dimensional Laplace transform [74, 135, 136]. Theorem 3.10 is due to Fliess [63, 66, 69]. See also [113, Chapter 3].

Section 3.5 The generating series for the parallel connections in Theorem 3.11 were first described by Fliess in [69]. The local convergence Theorem 3.12 can be found in the Ph.D. dissertation of Wang [199]. The global convergence results in Theorem 3.13 are based on the work of Venkatesh [195] and Winter Arboleda [205]. The remaining material in this section addressing the radius of convergence follows from the analysis by Thitsa and Gray in [190]. The only exception is the material at the end of the section regarding the boundary of the region of convergence for the parallel product connection. This issue was treated by Gray et al. in [92].

Section 3.6 The generating series for the Wiener-Fliess connection in Theorem 3.17 appeared in [81, 94]. (See the bibliographic notes for Section 2.7 for additional context regarding this result.) The cascade connection of Chen-Fliess series as described in Theorem 3.18 was first given by Ferfera in [58, 59], albeit in a different notation. Theorem 3.19 was first developed in an analytic setting in [141]. The ideas surrounding this result are certainly implicit in the early work of Fliess et al., see for example, [69, 74], but the approach of utilizing series composition was never explicitly mentioned or developed. The first local convergence regarding the Wiener-Fliess connection was presented in [81]. The radius of convergence for this connection given in Theorem 3.20 as well as the global result given in Theorem 3.21 (with $s = 0$) was given by Gray and Thitsa in [94]. The more general case is treated by Venkatesh in [195]. Local convergence of the cascade connection of two Fliess operators was first proved by Gray and Li in [91]. The radius of convergence for this connection given in Theorem 3.22 was developed in [190], as is Corollary 3.3. Finally, the global convergence of this cascade connection as described in Theorem 3.23 was proved for the $s = 0$ case in [190]. The general case is treated by Winter Arboleda in [205]. Also, see closely related results by Venkatesh in [195].

Section 3.7

The feedback connection of two Fliess operators was first treated in the Ph.D. dissertation of Ferfera [58]. The idea of using a contraction mapping theorem to prove the existence of a feedback product was discussed there, but the details of such a program first appeared in [91] with some additional improvements in [96]. The mixed composition product was further characterized in the Ph.D. dissertation of Li [140] under the name of the *modified* composition product. While central to the analysis, the role of this product as a right action in a transformation group was not full recognized until it appeared in [84]. See also [89] for a generalization. The characterization of the mixed composition product in Theorem 3.24 is based on an analogous treatment of the composition product appearing in [83] and subsequently by Foissy in [76]. Theorems 3.25 and 3.26 are based on the treatment in [91]. Lemmas 3.4 and 3.5 are a compilation of properties taken from [76, 86, 91, 140]. The first attempt to provide a computation framework for computing the feedback product in the single-input, single-output case was made by Gray et al. with the presentation of the output feedback Hopf algebra in [83]. Subsequent work by Foissy in [76] provided a grading that removed certain technical restrictions in the original approach. The group appearing in Theorem 3.27 and the feedback product in Theorem 3.28 follows the multivariable treatment presented in [86]. Lemma 3.6 and Theorem 3.29 also appeared in [86] as the multivariable extension of the SISO versions in [76]. A purely combinatorial treatment of the output feedback Hopf algebra is presented in [53] from a pre-Lie point of view. Theorem 3.30 is adapted from [87]. Theorems 3.31 and 3.32 are based on the presentation in [85]. The polynomials of Devlin in Example 3.33 first appeared in [49] in the context of Hilbert's sixteenth problem. Their connection to unity feedback systems was described in [56]. Related analysis appears in [57]. Finally, the convergence analysis of the feedback connection in Theorems 3.33-3.37 and the related examples are largely based on the work of Thitsa and Gray [189, 190]. The only exception is Example 3.39, which first appeared in [90].

Section 3.8 The formal Laplace-Borel transform has its roots in the work of E. Borel, who introduced the idea in the context of studying divergent series [11, pp. 242-245]. The treatment of the topic here is based largely on [50, pp. 232–235], in addition to [177, Section 10.2]. The formal Laplace-Borel transform was first used by Fliess, et al. for nonlinear systems analysis in [64, 69, 72, 74, 135]. A related approach was later developed by Hoang Ngoc Minh in [108]. The formal Laplace transform of a system employing the notion of a composition product was presented in [140, 141], but at the time there was no explicit uniqueness theorem for the generating series of a formal Fliess operator, so the analysis was restricted there to the locally convergent case. This deficiency was later addressed in [96], though to date a truly combinatorial treatment of this problem has not appeared to the author's knowledge. Theorem 3.38 and the supporting Lemmas 3.8, 3.9 and Lemma 3.10 are all taken from this reference. The formula in Lemma 3.8, however, has appeared in a number of earlier places, for example, in the work of Crouch and Lamnabhi-Lagarrigue [137], and Sontag and Wang [200]. Theorems 3.39 and 3.40 are based on the presentation in [140, 141]. Finally, Example 3.51 was adapted from [74, 140, 141].

4. Rational Series and Linear Representations

In Chapter 1 a series $c \in \mathbb{R}\langle\langle X\rangle\rangle$ was said to be rational if it assumes either the form ba^{-1} or $ba^{-1}x_i$, where $i \neq 0$ and, a and b are polynomials in x_0. The series a^{-1} is the inverse of a in the sense that $aa^{-1} = a^{-1}a = 1$, but thus far no general representation of a^{-1} has been developed. In this chapter, the notion of rationality is generalized to its fullest extent. It relies on the availability of four *rational operations* defined on $\mathbb{R}\langle\langle X\rangle\rangle$: addition, scalar multiplication, the Cauchy product, and inversion. Then the issue of determining when a series has a *linear* representation is addressed. Such series are said to be *recognizable*. A fundamental result in this area is Schützenberger's Representation Theorem, which states that a series is rational if and only if it is recognizable. Next, a Hankel matrix characterization of rationality is given. This turns out to be useful for characterizing the *minimality* of linear representations. It also has a canonical factorization that will be used in Chapter 6 to describe state space realizations of Fliess operators with rational generating series. Finally, the shuffle product and the composition product are considered on the set of rational series. It is shown that the shuffle product preserves rationality, while the composition product does not unless certain conditions are met. In addition, it is shown that the feedback product also does not preserve rationality.

4.1 Rational Series

Consider a fixed alphabet $X = \{x_0, x_1, \ldots, x_m\}$. A series $c \in \mathbb{R}\langle\langle X\rangle\rangle$ is called *Cauchy invertible* or simply *invertible* (in this chapter) if there exists a series $c^{-1} \in \mathbb{R}\langle\langle X\rangle\rangle$ such that $cc^{-1} = c^{-1}c = 1$. In the event that c is not proper, it is always possible to write

$$c = (c, \emptyset)(1 - c'),$$

where $(c, \emptyset)$ is nonzero, and $c' \in \mathbb{R}\langle\langle X\rangle\rangle$ is proper. It then follows that

$$c^{-1} = \frac{1}{(c, \emptyset)}(1 - c')^{-1} = \frac{1}{(c, \emptyset)}(c')^*,$$

where

$$(c')^* := \sum_{i=0}^{\infty}(c')^i$$

(see Example 2.13). In fact, c is invertible if and *only if* c is not proper (see Problem 4.1.1). Now let S be a subalgebra of the $\mathbb{R}$-algebra $\mathbb{R}\langle\langle X\rangle\rangle$ with the Cauchy product. S is said to be *rationally closed* when every invertible $c \in S$ has $c^{-1} \in S$ (or equivalently, every proper $c' \in S$ has $(c')^* \in S$). The *rational closure* of any subset $E \subset \mathbb{R}\langle\langle X\rangle\rangle$ is the smallest rationally closed subalgebra of $\mathbb{R}\langle\langle X\rangle\rangle$ containing E.

Definition 4.1 *A series $c \in \mathbb{R}\langle\langle X\rangle\rangle$ is* ***rational*** *if it belongs to the rational closure of $\mathbb{R}\langle X\rangle$.*

Thus, a given rational series can be obtained from a finite set of polynomials by performing a finite number of additions, scalar products, Cauchy products, and inversions (or star operations), the so called *rational operations*.

Example 4.1 Suppose $X = \{x_0, x_1\}$ and $E = \{x_0, 1 + x_0x_1\}$. Then the rational closure of E contains elements like: x_0, $1 + x_0x_1$, $(1 + x_0x_1)^{-1}$, 1, x_0x_1, $x_0^2x_1$, $2(x_0+x_0^2x_1)$, $3+x_0+x_0x_1x_0$, $x_0(1+x_0x_1)^{-1}$,..., but not, for example, $x_0 + x_1$ or x_1^2. □

Example 4.2 Suppose $X = \{x\}$, $\{\alpha_i\}_{i\in\mathbb{N}_0}$ is a sequence of real numbers, and

$$c = \sum_{i=0}^{\infty} \alpha_i x^i.$$

If $\alpha_i = \alpha^i$ for every $i \in \mathbb{N}_0$ and some $\alpha \in \mathbb{R}$ then the series c is clearly rational because $c = (1 - \alpha x)^{-1}$. On the other hand, when $\alpha_i = \alpha^i/i!$, the series

$$c = \sum_{i=0}^{\infty} \frac{(\alpha x)^i}{i!} = e^{\alpha x}$$

is not rational since it is not the result of applying a finite number of rational operations to any finite set of polynomials in $\mathbb{R}\langle X\rangle$. □

Let $\mathbb{R}\langle\langle X\rangle\rangle^{n\times n}$ denote the set of $n \times n$ matrices with components from $\mathbb{R}\langle\langle X\rangle\rangle$. Convergence in $\mathbb{R}\langle\langle X\rangle\rangle^{n\times n}$ is defined componentwise using the ultrametric dist on $\mathbb{R}\langle\langle X\rangle\rangle$, that is, if $\mathbf{A}, \mathbf{B} \in \mathbb{R}\langle\langle X\rangle\rangle^{n\times n}$ then $\mathrm{dist}(\mathbf{A}, \mathbf{B}) := \max_{i,j} \mathrm{dist}(a_{ij}, b_{ij})$. A matrix $\mathbf{C} \in \mathbb{R}\langle\langle X\rangle\rangle^{n\times n}$ is called proper when every component c_{ij} is proper. Similar to the scalar case, one can verify that $\mathbf{C}^*$ is well defined if $\mathbf{C}$ is proper (see Problem 4.1.4). The following lemma describes more precisely the nature of the components of $\mathbf{C}^*$. This characterization will be employed shortly to describe the relationship between rational and recognizable series.

Lemma 4.1 *If* $\mathbf{C} \in \mathbb{R}\langle\langle X\rangle\rangle^{n\times n}$ *is proper then*

$$\mathbf{C}^* = \sum_{i=0}^{\infty} \mathbf{C}^i$$

has components in the rational closure of the components of $\mathbf{C}$.

Proof: The proof is by induction on n. The result is immediate when $n = 1$. So select some $n \geq 1$ and assume the lemma holds up to this fixed value. Analogous to the scalar case, it is easily verified that $\mathbf{C}^* \in \mathbb{R}^{(n+1)\times(n+1)}$ is the unique solution to the matrix equations

$$(I_{n+1} - \mathbf{C})\mathbf{C}^* = I_{n+1}, \quad \mathbf{C}^*(I_{n+1} - \mathbf{C}) = I_{n+1},$$

where I_{n+1} is an $(n+1)\times(n+1)$ identity matrix (see Problem 4.1.5). Consequently,

$$\mathbf{C}^* = I_{n+1} + \mathbf{C}\mathbf{C}^* = I_{n+1} + \mathbf{C}^*\mathbf{C}. \tag{4.1}$$

Now partition the rows and columns of the matrix $\mathbf{C}$ in the following manner:

$$\mathbf{C} = \begin{array}{c} \begin{array}{cc} \scriptstyle n & \scriptstyle 1 \end{array} \\ \left[\begin{array}{c|c} \mathbf{C}_1 & \mathbf{C}_4 \\ \hline \mathbf{C}_3 & \mathbf{C}_2 \end{array}\right] \begin{array}{c} \scriptstyle n \\ \scriptstyle 1 \end{array} \end{array}$$

It can be verified by direct substitution into (4.1) that

$$\mathbf{C}^* = \begin{bmatrix} \Delta_1^* & \mathbf{C}_1^*\mathbf{C}_4\Delta_2^* \\ \mathbf{C}_2^*\mathbf{C}_3\Delta_1^* & \Delta_2^* \end{bmatrix}, \tag{4.2}$$

where $\Delta_1 = \mathbf{C}_1 + \mathbf{C}_4\mathbf{C}_2^*\mathbf{C}_3$ and $\Delta_2 = \mathbf{C}_2 + \mathbf{C}_3\mathbf{C}_1^*\mathbf{C}_4$ (see Problem 4.1.6). Since the lemma is assumed to hold for matrices of dimension n or less, all the components of $\mathbf{C}_1^*$ and $\mathbf{C}_2^*$ are in the rational closure of the components of $\mathbf{C}$. The same is obviously true for the components of Δ_1^* and Δ_2^*, and thus for all the components of $\mathbf{C}^*$. Hence, by induction, the lemma is true for all $n \geq 1$. ■

4.2 Recognizable Series

As discussed in Section 2.1, the collection of $\mathbb{R}$-linear mappings on the vector space $\mathbb{R}^n$, represented by the set of matrices $\mathbb{R}^{n\times n}$, forms a monoid under matrix multiplication. The following definition utilizes this fact to describe the central notion behind recognizability of a formal power series.

Definition 4.2 *A* ***linear representation*** *of a series $c \in \mathbb{R}\langle\langle X\rangle\rangle$ is any triple (μ, γ, λ), where*

$$\mu : X^* \to \mathbb{R}^{n\times n}$$

is a monoid homomorphism, and $\gamma, \lambda^T \in \mathbb{R}^{n\times 1}$ are such that

$$(c, \eta) = \lambda\mu(\eta)\gamma, \ \forall \eta \in X^*.$$

The integer $n \geq 1$ is the dimension of the representation.

Definition 4.3 *A series $c \in \mathbb{R}\langle\langle X\rangle\rangle$ is called* ***recognizable*** *if it has a linear representation.*

Given any linear representation (μ, γ, λ) of c, the homomorphism μ is uniquely specified by its image on $X = \{x_0, x_1, \ldots, x_m\}$, specifically by the set of matrices $N = \{N_0, N_1, \ldots, N_m\}$, where $N_i = \mu(x_i),\ i = 0, 1, \ldots, m$ (see Problem 2.1.3). In which case, c can be written in the form

$$\begin{aligned} c &= \sum_{\eta\in X^*} (\lambda\mu(\eta)\gamma)\, \eta \\ &= \sum_{k=0}^{\infty} \sum_{i_1,\ldots,i_k=0}^{m} (\lambda N_{i_k} \cdots N_{i_1}\gamma)\, x_{i_k} \cdots x_{i_1}. \end{aligned} \tag{4.3}$$

Given one linear representation for c, it is trivial to produce another representation of the same dimension. The set of all nonsingular matrices in $\mathbb{R}^{n\times n}$ defines the general linear group $GL_n(\mathbb{R})$. This group acts as a transformation group on the set of n dimensional linear representations via the mapping

$$\mathcal{A}_T : (\mu_N, \gamma, \lambda) \to (\mu_{\tilde{N}}, \tilde{\gamma}, \tilde{\lambda}),$$

where μ_N is the monoid homomorphism specified by the set of matrices N, and

$$\tilde{N}_i = TN_iT^{-1},\ \tilde{\gamma} = T\gamma,\ \ \tilde{\lambda} = \lambda T^{-1}.$$

Clearly $(\mu_{\tilde{N}}, \tilde{\gamma}, \tilde{\lambda})$ is another representation of c. Since there are uncountable infinitely many such T in $GL_n(\mathbb{R})$, the set of linear representations of c has the same cardinality. In fact, the action $\mathcal{A}$ does not account for all possible representations of c since representations of different dimensions can also exist (see Problem 4.4.6).

Example 4.3 Suppose $X = \{x_0\}$ and $c \in \mathbb{R}[[X_0]]$ is recognizable with representation (μ_A, z_0, C) and $A = \mu(x)$. Using the homomorphism

$$\begin{aligned} \rho : X^* &\to \mathbb{N} \\ : \eta &\mapsto |\eta| \end{aligned}$$

described in Example 2.3, the mapping

$$\mu_A(\eta) = A^{\rho(\eta)}$$

is a homomorphism of X^* into $\mathbb{R}^{n\times n}$. Therefore, the series c can be written in the form

$$\begin{aligned} c &= \sum_{\eta\in X^*} (CA^{\rho(\eta)}z_0)\,\eta = \sum_{i=0}^{\infty} (CA^i z_0)\,x_0^i \\ &= C(Ax_0)^* z_0. \end{aligned}$$

□

A central question is how to determine when a given series is recognizable. Ultimately, this will be answered via three different methods, each of which provides some insight into the nature of recognizable series. But first a necessary condition for recognizability is presented. In applications, it can be useful for quickly determining if a specific series has any hope of being recognizable.

Theorem 4.1 *If $c \in \mathbb{R}\langle\langle X\rangle\rangle$ is recognizable, then c is globally convergent with Gevrey order $s^* = 0$.*

Proof: Since c is recognizable, its coefficients can be written in the form

$$(c, x_{i_k} \cdots x_{i_1}) = \lambda N_{i_k} \cdots N_{i_1} \gamma.$$

Assuming $c \neq 0$, otherwise the problem is trivial, define the positive constants $K = \|\lambda\|\,\|\gamma\|$ and $M = \max_i \|N_i\|$, where $\|\cdot\|$ when applied to a matrix denotes the induced matrix norm. For any $k \geq 0$, it follows from the Cauchy-Schwarz inequality that

$$\begin{aligned}
|(c, x_{i_k} \cdots x_{i_1})| &= |\lambda N_{i_k} \cdots N_{i_1} \gamma| \\
&\leq \|\lambda\|\,\|N_{i_k} \cdots N_{i_1} \gamma\| \\
&\leq \|\lambda\|\,\|N_{i_k} \cdots N_{i_1}\|\,\|\gamma\| \\
&\leq \|\lambda\|\,\|N_{i_k}\| \cdots \|N_{i_1}\|\,\|\gamma\| \\
&\leq K\,M^k.
\end{aligned}$$

■

Thus, if a series c is recognizable, its coefficients can have at most a geometric growth rate. The following property provides the first necessary and sufficient test for recognizability. The subsequent two sections provide alternative characterizations of recognizability in terms of rationality and the Hankel mapping of c.

Definition 4.4 *A subset $V \subset \mathbb{R}\langle\langle X\rangle\rangle$ is called* ***stable*** *when $\xi^{-1}(c) \in V$ for all $c \in V$ and $\xi \in X^*$.*

Theorem 4.2 *A series $c \in \mathbb{R}\langle\langle X\rangle\rangle$ is recognizable if and only if there exists a stable finite dimensional $\mathbb{R}$-vector subspace of $\mathbb{R}\langle\langle X\rangle\rangle$ containing c.*

Proof: Suppose c is recognizable. Then there exists a linear representation (μ, γ, λ) of finite dimension n such that

$$(c, \eta) = \lambda \mu(\eta) \gamma, \ \forall \eta \in X^*.$$

Define the set of series $\{\bar{c}_i\}_{i=1}^n$ in $\mathbb{R}\langle\langle X\rangle\rangle$ by

$$(\bar{c}_i, \eta) = [\mu(\eta)\gamma]_i, \ i = 1, \ldots, n,$$

where $[v]_i$ denotes the i-th component of vector $v \in \mathbb{R}^n$. Let V be the $\mathbb{R}$-vector subspace of $\mathbb{R}\langle\langle X \rangle\rangle$ defined by the span of $\{\bar{c}_i\}_{i=1}^n$. Observe that for any $\eta \in X^*$,

$$(c,\eta) = \lambda\mu(\eta)\gamma = \sum_{i=1}^n \lambda_i[\mu(\eta)\gamma]_i = \sum_{i=1}^n \lambda_i(\bar{c}_i,\eta),$$

and thus,

$$\begin{aligned} c &= \sum_{\eta\in X^*}(c,\eta)\eta = \sum_{\eta\in X^*}\left(\sum_{i=1}^n \lambda_i(\bar{c}_i,\eta)\right)\eta \\ &= \sum_{i=1}^n \lambda_i\left(\sum_{\eta\in X^*}(\bar{c}_i,\eta)\eta\right) = \sum_{i=1}^n \lambda_i\bar{c}_i. \end{aligned}$$

Hence, $c \in V$. Now select any $\hat{c} \in V$. Clearly, there exists real numbers $\{\hat{\lambda}_i\}_{i=1}^n$ such that

$$\begin{aligned} (\hat{c},\eta) &= \sum_{i=1}^n \hat{\lambda}_i(\bar{c}_i,\eta) \\ &= \sum_{i=1}^n \hat{\lambda}_i[\mu(\eta)\gamma]_i \\ &= \hat{\lambda}\mu(\eta)\gamma. \end{aligned}$$

Consequently, for any $\xi \in X^*$

$$\begin{aligned} \xi^{-1}(\hat{c}) &= \sum_{\eta\in X^*}(\hat{c},\xi\eta)\eta = \sum_{\eta\in X^*}(\hat{\lambda}\mu(\xi\eta)\gamma)\eta \\ &= \sum_{\eta\in X^*}(\hat{\lambda}\mu(\xi)\mu(\eta)\gamma)\eta \\ &= \sum_{\eta\in X^*}\sum_{i=1}^n[\hat{\lambda}\mu(\xi)]_i[\mu(\eta)\gamma]_i\eta \\ &= \sum_{i=1}^n[\hat{\lambda}\mu(\xi)]_i\bar{c}_i. \end{aligned}$$

So $\xi^{-1}(\hat{c}) \in V$, and this proves the theorem in one direction.

Now conversely, let V be a stable n-dimensional $\mathbb{R}$-vector subspace of $\mathbb{R}\langle\langle X \rangle\rangle$ containing c. Let $\{\bar{c}_i\}_{i=1}^n$ be any basis for V. Then it is possible to write

$$c = \sum_{i=1}^{n} \lambda_i \bar{c}_i \tag{4.4}$$

for some $\lambda_i \in \mathbb{R}$, $i = 1, \ldots, n$. Since V is stable, it follows that $x^{-1}(\bar{c}_i) \in V$ for every letter $x \in X$ and every $i = 1, 2, \ldots, n$. Thus, it is also possible to write

$$x^{-1}(\bar{c}_i) = \sum_{j=1}^{n} [\mu(x)]_{ij} \bar{c}_j,$$

where $[\mu(x)]_{ij} \in \mathbb{R}$, $i, j = 1, \ldots, n$. This clearly defines a mapping $\mu : X \mapsto \mathbb{R}^{n\times n}$. Furthermore, this mapping can be uniquely extended in the usual manner to a monoid homomorphism $\mu : X^* \to \mathbb{R}^{n\times n}$, i.e.,

$$\mu(x_{i_k} x_{i_{k-1}} \cdots x_{i_1}) = \mu(x_{i_k})\mu(x_{i_{k-1}}) \cdots \mu(x_{i_1}) \tag{4.5}$$

(see Problem 2.1.3). It is now shown inductively on the length of an arbitrary $\eta \in X^*$ that

$$\eta^{-1}(\bar{c}_i) = \sum_{j=1}^{n} [\mu(\eta)]_{ij} \bar{c}_j. \tag{4.6}$$

The empty word case is trivial, and the case where $\eta \in X$ was established above. Suppose that up to some fixed $k \geq 1$ the identity (4.6) holds for every $\eta \in X^k$. Then for any $x \in X$ it follows that

$$\begin{aligned}
(\eta x)^{-1}(\bar{c}_i) &= x^{-1}(\eta^{-1}(\bar{c}_i)) \\
&= x^{-1}\left(\sum_{k=1}^{n} [\mu(\eta)]_{ik} \bar{c}_k\right) \\
&= \sum_{k=1}^{n} [\mu(\eta)]_{ik}\, x^{-1}(\bar{c}_k) \\
&= \sum_{k=1}^{n} [\mu(\eta)]_{ik} \sum_{j=1}^{n} [\mu(x)]_{kj} \bar{c}_j \\
&= \sum_{j=1}^{n} \left(\sum_{k=1}^{n} [\mu(\eta)]_{ik} [\mu(x)]_{kj}\right) \bar{c}_j \\
&= \sum_{j=1}^{n} [\mu(\eta x)]_{ij} \bar{c}_j.
\end{aligned}$$

Hence, equation (4.6) holds for all $\eta \in X^{k+1}$ and thus for every word in X^*. Now define $\gamma \in \mathbb{R}^{n\times 1}$ componentwise by $\gamma_j = (\bar{c}_j, \emptyset)$. Observe that for any $\eta \in X^*$

$$\begin{aligned}(\bar{c}_i, \eta) &= (\eta^{-1}(\bar{c}_i), \emptyset) \\ &= \sum_{j=1}^{n} [\mu(\eta)]_{ij} (\bar{c}_j, \emptyset) \\ &= [\mu(\eta)\gamma]_i,\end{aligned}$$

and thus, from equation (4.4),

$$\begin{aligned}(c, \eta) &= \sum_{i=1}^{n} \lambda_i (\bar{c}_i, \eta) \\ &= \sum_{i=1}^{n} \lambda_i [\mu(\eta)\gamma]_i \\ &= \lambda\mu(\eta)\gamma.\end{aligned}$$

So c is recognizable. ■

Example 4.4 Suppose $p \in \mathbb{R}\langle X\rangle$ where $X = \{x_0, x_1, \ldots, x_m\}$. Let V_p be the $\mathbb{R}$-vector subspace of $\mathbb{R}\langle X\rangle \subset \mathbb{R}\langle\langle X\rangle\rangle$, where $\hat{p} \in V_p$ if $\deg(\hat{p}) \leq \deg(p)$. Clearly $p \in V_p$, and it is easily verified that $\dim(V_p) = \sum_{i=0}^{\deg(p)} (m+1)^i$. In addition, V_p is stable since $\xi^{-1}(\hat{p}) \leq \deg(p)$ for any $\xi \in X^*$ and $\hat{p} \in V_p$. Thus, every polynomial p, which was already known to be rational, is now also seen to be recognizable. □

Example 4.5 Suppose $p \in \mathbb{R}\langle X\rangle$ with $(p, \emptyset) = 1$. Then $p^{-1} = (p')^*$, where $p' = 1 - p$. Since p' is recognizable there exists a finite-dimensional $\mathbb{R}$-vector subspace $V_{p'}$ containing p'. Define a second subspace

$$V_{p^{-1}} = \{\alpha + \hat{p}p^{-1} : \alpha \in \mathbb{R},\ \hat{p} \in V_{p'}\}.$$

Clearly, $p^{-1} \in V_{p^{-1}}$ since $p^{-1} = 1 + p'p^{-1}$. The subspace $V_{p^{-1}}$ is finite dimensional since $V_{p'}$ is. The stability of $V_{p^{-1}}$ follows directly from Lemma 2.1 using the fact that p' is proper. Specifically, for any $x \in X$:

$$x^{-1}(\alpha + \hat{p}p^{-1}) = x^{-1}(\hat{p}p^{-1})$$

$$
\begin{aligned}
&= x^{-1}(\hat{p})p^{-1} + (\hat{p},\emptyset)\, x^{-1}(p^{-1}) \\
&= x^{-1}(\hat{p})p^{-1} + (\hat{p},\emptyset)\, x^{-1}(p'p^{-1}) \\
&= x^{-1}(\hat{p})p^{-1} + (\hat{p},\emptyset)x^{-1}(p')p^{-1} \\
&= \left[x^{-1}(\hat{p}) + (\hat{p},\emptyset)x^{-1}(p')\right] p^{-1}.
\end{aligned}
$$

But $x^{-1}(\hat{p}), x^{-1}(p') \in V_{p'}$ since $\hat{p}, p' \in V_{p'}$, and $V_{p'}$ is stable. Thus, $x^{-1}(\alpha + \hat{p}p^{-1}) \in V_{p^{-1}}$, and therefore $V_{p^{-1}}$ is also stable (see Problem 4.2.1). Consequently, p^{-1} is recognizable, in addition to being rational. In the next section, it is shown that being both rational and recognizable is no coincidence. □

4.3 Schützenberger's Theorem

The following theorem is fundamental in the theory of rational series.

Theorem 4.3 (Schützenberger) *A formal power series is rational if and only if it is recognizable.*

Proof: Suppose c is recognizable. Let (μ, γ, λ) be any linear representation of c of dimension n. Consider the proper matrix in $\mathbb{R}\langle X\rangle^{n\times n}$:

$$\mathbf{C} = \sum_{x\in X} \mu(x)x.$$

It follows from a defining property of homomorphisms (see equation (2.1)) and a modest generalization of the identity $(\mathrm{char}(X))^i = \mathrm{char}(X^i)$ (see Problem 2.4.6) that

$$
\begin{aligned}
\mathbf{C}^* = \sum_{i=0}^{\infty} \mathbf{C}^i &= \sum_{i=0}^{\infty} \left(\sum_{x\in X} \mu(x)x \right)^i \\
&= \sum_{i=0}^{\infty} \sum_{\eta\in X^i} \mu(\eta)\eta = \sum_{\eta\in X^*} \mu(\eta)\eta.
\end{aligned}
$$

Thus,

$$c = \sum_{\eta\in X^*} (\lambda\mu(\eta)\gamma)\eta = \lambda\mathbf{C}^*\gamma.$$

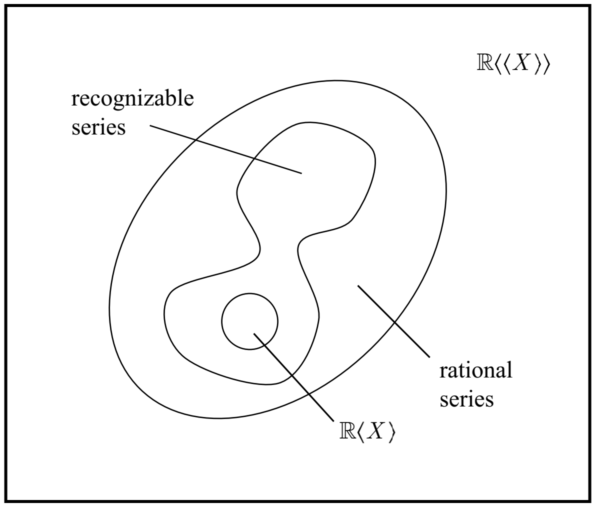


Fig. 4.1. The initial set inclusions in the proof that all rational series are recognizable.

From Lemma 4.1, every component of $\mathbf{C}^*$ is in the rational closure of the polynomials comprising $\mathbf{C}$. Hence, c is rational.

It is next shown that all rational series are recognizable. In light of the result just proven above and Example 4.4, the known set inclusions are as shown in Figure 4.1. Observe it is only necessary to show that the set of recognizable series is rationally closed, and thus, it can not be smaller than the *smallest* rational closure of $\mathbb{R}\langle X\rangle$, namely, the set of rational series. Hence, the set of recognizable series is not a proper subset of the set of rational series, as shown in the figure, but is in fact equivalent to this set.

The set of recognizable series is rationally closed:

1. Suppose series c and d are recognizable. Let V_c and V_d be any corresponding stable finite dimensional $\mathbb{R}$-vector subspaces of $\mathbb{R}\langle\langle X\rangle\rangle$ as per Theorem 4.2. Then $c+d$ is in the inner sum V_c+V_d, which is also stable and a finite dimensional $\mathbb{R}$-vector subspaces of $\mathbb{R}\langle\langle X\rangle\rangle$ (see Problem 4.2.2). Hence, $c+d$ is recognizable. Closure under the scalar product is shown similarly.
2. Suppose series c and d are recognizable with V_c and V_d as defined above. Define the subspace

$$V_{cd} = \{\hat{c}d + \hat{d} : \hat{c} \in V_c,\ \hat{d} \in V_d\}.$$

V_{cd} contains cd and is finite dimensional since both V_c and V_d are. Furthermore, as a direct consequence of Lemma 2.1, V_{cd} is stable (see Problem 4.3.1). Thus, cd is recognizable.

3. Suppose c is recognizable and proper. Define the subspace

$$V_{c^*} = \{\alpha + \hat{c}c^* : \alpha \in \mathbb{R},\ \hat{c} \in V_c\}.$$

Again, $c^* \in V_{c^*}$, and V_{c^*} is a finite dimensional $\mathbb{R}$-vector subspace of $\mathbb{R}\langle\langle X\rangle\rangle$ since V_c is. The stability of V_{c^*} follows from the same type of argument as given in Example 4.5. Specifically, from the identity $c^* = 1 + cc^*$ and the assumption that $(c, \emptyset) = 0$, it follows for any $x \in X$ that:

$$\begin{aligned} x^{-1}(\alpha + \hat{c}c^*) &= x^{-1}(\hat{c}c^*) \\ &= x^{-1}(\hat{c})c^* + (\hat{c}, \emptyset)\, x^{-1}(c^*) \\ &= x^{-1}(\hat{c})c^* + (\hat{c}, \emptyset)\, x^{-1}(cc^*) \\ &= x^{-1}(\hat{c})c^* + (\hat{c}, \emptyset) x^{-1}(c)c^* \\ &= \left[x^{-1}(\hat{c}) + (\hat{c}, \emptyset) x^{-1}(c)\right] c^*. \end{aligned}$$

But $x^{-1}(\hat{c}), x^{-1}(c) \in V_c$ since $c, \hat{c} \in V_c$, and V_c is stable. Thus, $x^{-1}(\alpha + \hat{c}c^*) \in V_{c^*}$, and therefore, V_{c^*} is stable.

■

4.4 Hankel Rank of a Rational Series

In this section, a characterization of a rational/recognizable series is given via its Hankel mapping. The following definition generalizes the more familiar notion introduced in Chapter 1.

Definition 4.5 *For any $c \in \mathbb{R}\langle\langle X\rangle\rangle$, the $\mathbb{R}$-linear mapping $\mathcal{H}_c$: $\mathbb{R}\langle X\rangle \to \mathbb{R}\langle\langle X\rangle\rangle$ on the $\mathbb{R}$-vector space $\mathbb{R}\langle X\rangle$ uniquely specified by*

$$(\mathcal{H}_c(\eta), \xi) = (c, \xi\eta),\ \ \forall \xi, \eta \in X^*$$

is called the ***Hankel mapping*** *of c.*

Each $\mathbb{R}$-vector space $\mathbb{R}\langle X\rangle$ and $\mathbb{R}\langle\langle X\rangle\rangle$ is clearly spanned by the set of monomials $\{1\,\eta\}_{\eta\in X^*}$. So the mapping $\mathcal{H}_c : \mathbb{R}\langle X\rangle \to \mathbb{R}\langle\langle X\rangle\rangle$ has a matrix representation, whose (ξ,η) component is given by $[\mathcal{H}_c]_{\xi,\eta} = (\mathcal{H}_c(\eta),\xi) = (c,\xi\eta)$ (see Problem 4.4.1). Its range space, $\mathcal{H}_c(\mathbb{R}\langle X\rangle)$, is a vector subspace of $\mathbb{R}\langle\langle X\rangle\rangle$, but not necessarily finite dimensional. Consider the following definition.

Definition 4.6 *The* ***Hankel rank*** *of* $c \in \mathbb{R}\langle\langle X\rangle\rangle$ *is*

$$\rho_H(c) = \dim(\mathcal{H}_c(\mathbb{R}\langle X\rangle)).$$

Example 4.6 Suppose $X = \{x_0\}$. Given any $c \in \mathbb{R}\langle\langle X\rangle\rangle$ and $x_0^i, x_0^j \in X^*$, observe

$$(\mathcal{H}_c(x_0^i), x_0^j) = (c, x_0^{i+j}).$$

Thus, the linear mapping $\mathcal{H}_c$ has the matrix representation

$$\mathcal{H}_c = \begin{bmatrix} (c,\emptyset) & (c,x_0) & (c,x_0^2) & \cdots \\ (c,x_0) & (c,x_0^2) & (c,x_0^3) & \cdots \\ (c,x_0^2) & (c,x_0^3) & (c,x_0^4) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

which has the well known Hankel structure. As described in Section 1.1, when the rank of this matrix is finite, there exists polynomials $p, q \in \mathbb{R}\langle X\rangle$ so that $c = pq^{-1}$. Hence, the series c is rational. □

The Hankel characterization of rationality for an arbitrary finite alphabet is given below.

Theorem 4.4 *A series* $c \in \mathbb{R}\langle\langle X\rangle\rangle$ *is rational if and only if its Hankel rank,* $\rho_H(c)$*, is finite.*

Proof: Suppose c is rational. Then from Theorems 4.2 and 4.3 there exists a stable $\mathbb{R}$-vector subspace $V_c \subset \mathbb{R}\langle\langle X\rangle\rangle$ containing c with finite dimension n. Let $\{\bar{c}_i\}_{i=1}^n$ be any basis for V_c. Since V_c is stable, for any $\xi \in X^*$ there exists real numbers $\{\alpha_{\xi,i}\}_{i=1}^n$ such that

$$\xi^{-1}(c) = \sum_{i=1}^{n} \alpha_{\xi,i}\,\bar{c}_i.$$

For each $\bar{c}_i$, associate the companion series

$$e_i = \sum_{\xi \in X^*} \alpha_{\xi,i}\, \xi.$$

It is shown next that $\mathcal{H}_c(\mathbb{R}\langle X\rangle)$ lies in the subspace of $\mathbb{R}\langle\langle X\rangle\rangle$ spanned by $\{e_i\}_{i=1}^n$. Therefore, $\rho_H(c) \leq n < \infty$. Select any polynomial $p = \sum_{j=1}^r (p, \eta_j)\eta_j$ and apply $\mathcal{H}_c$:

$$\begin{aligned}
\mathcal{H}_c(p) &= \sum_{j=1}^r \mathcal{H}_c(\eta_j)(p, \eta_j) \\
&= \sum_{j=1}^r \sum_{\xi \in X^*} (\mathcal{H}_c(\eta_j), \xi)\xi\, (p, \eta_j) \\
&= \sum_{j=1}^r \sum_{\xi \in X^*} (c, \xi\eta_j)\xi\, (p, \eta_j) \\
&= \sum_{j=1}^r \sum_{\xi \in X^*} (\xi^{-1}(c), \eta_j)\xi\, (p, \eta_j) \\
&= \sum_{j=1}^r \sum_{\xi \in X^*} \left[\sum_{i=1}^n \alpha_{\xi,i}(\bar{c}_i, \eta_j)\right] \xi\, (p, \eta_j) \\
&= \sum_{i=1}^n \left[\sum_{j=1}^r (\bar{c}_i, \eta_j)(p, \eta_j)\right] e_i.
\end{aligned}$$

Thus, the first half of the theorem is proved.

Now consider the converse claim. Suppose $\rho_H(c)$ is finite, say $\rho_H(c) = n$. The set of words $\{\eta_i\}_{i=1}^n$ in X^*, which label the n linearly independent columns of $\mathcal{H}_c$, describe a basis $\{\mathcal{H}_c(\eta_i)\}_{i=1}^n$ for $\mathcal{H}_c(\mathbb{R}\langle X\rangle)$. For each $x \in X$, one can uniquely define a matrix $\mu(x) \in \mathbb{R}^{n\times n}$ via the system of equations

$$\mathcal{H}_c(x\eta_i) = \sum_{j=1}^n [\mu(x)]_{ji}\, \mathcal{H}_c(\eta_j), \quad i = 1, \ldots, n, \tag{4.7}$$

since $\mathcal{H}_c(x\eta_i)$ is the $x\eta_i$-th column of the matrix $\mathcal{H}_c$. The mapping $\mu : X \to \mathbb{R}^{n\times n}$ can be uniquely extended to a monoid homomorphism on X^* via equation (4.5). (This part of the proof closely parallels the proof of Theorem 4.2.) It is now shown inductively on word length that for an arbitrary $\eta \in X^*$

$$\mathcal{H}_c(\eta\eta_i) = \sum_{j=1}^{n} [\mu(\eta)]_{ji} \mathcal{H}_c(\eta_j). \tag{4.8}$$

The claim is trivial if η is empty. Apply (4.7) for words of length one. Suppose that up to some fixed $k \geq 0$ the identity (4.8) holds for every $\eta \in X^k$. Then for any $x \in X$ and $\xi \in X^*$ observe that

$$\begin{aligned}
(\mathcal{H}_c(x\eta\eta_i), \xi) &= (c, \xi x\eta\eta_i) \\
&= (\mathcal{H}_c(\eta\eta_i), \xi x) \\
&= \sum_{k=1}^{n} [\mu(\eta)]_{ki} (\mathcal{H}_c(\eta_k), \xi x) \\
&= \sum_{k=1}^{n} [\mu(\eta)]_{ki} (\mathcal{H}_c(x\eta_k), \xi) \\
&= \sum_{k=1}^{n} [\mu(\eta)]_{ki} \sum_{j=1}^{n} [\mu(x)]_{jk} (\mathcal{H}_c(\eta_j), \xi) \\
&= \sum_{j=1}^{n} [\mu(x\eta)]_{ji} (\mathcal{H}_c(\eta_j), \xi).
\end{aligned}$$

Thus, equation (4.8) holds for every $\eta \in X^*$. Now since $c \in \mathcal{H}_c(\mathbb{R}\langle X \rangle)$ ($c = \mathcal{H}_c(1)$), there must exist scalars $\{\gamma_i\}_{i=1}^n$ such that $c = \sum_{i=1}^n \gamma_i \mathcal{H}_c(\eta_i)$. Therefore, given any $\eta \in X^*$ observe

$$\begin{aligned}
(c, \eta) &= \sum_{i=1}^{n} \gamma_i (\mathcal{H}_c(\eta_i), \eta) = \sum_{i=1}^{n} \gamma_i (c, \eta\eta_i) \\
&= \sum_{i=1}^{n} \gamma_i (\mathcal{H}_c(\eta\eta_i), \emptyset) = \sum_{i=1}^{n} \gamma_i \sum_{j=1}^{n} [\mu(\eta)]_{ji} (\mathcal{H}_c(\eta_j), \emptyset) \\
&= \lambda\mu(\eta)\gamma,
\end{aligned}$$

where

$$\begin{aligned}
\lambda &= [(\mathcal{H}_c(\eta_1), \emptyset) \ (\mathcal{H}_c(\eta_2), \emptyset) \ \cdots \ (\mathcal{H}_c(\eta_n), \emptyset)] \\
&= [(c, \eta_1) \ (c, \eta_2) \ \cdots \ (c, \eta_n)] \\
\gamma &= [\gamma_1 \ \gamma_2 \ \cdots \gamma_n]^T.
\end{aligned}$$

Thus, c is recognizable, or equivalently, rational. ■

Example 4.7 Let $X = \{x_1\}$ and consider the series $c = \sum_{k\geq 0} x_1^k$. For any $p \in \mathbb{R}\langle X\rangle$ and $\eta \in X^*$ observe that

$$\begin{aligned}(\mathcal{H}_c(p), \eta) &= \sum_{\xi\in X^*} (p, \xi)(\mathcal{H}_c(\xi), \eta)\\ &= \sum_{\xi\in X^*} (p, \xi)(c, \eta\xi).\end{aligned}$$

Therefore,

$$\begin{aligned}(\mathcal{H}_c(p), x_1^i) &= \sum_{j=0}^{\infty} (p, x_1^j)(c, x_1^i x_1^j)\\ &= \sum_{j=0}^{\infty} (p, x_1^j),\end{aligned}$$

or equivalently,

$$\begin{aligned}\mathcal{H}_c(p) &= \sum_{i=0}^{\infty} \left[\sum_{j=0}^{\infty} (p, x_1^j)\right] x_1^i\\ &= c\sum_{j=0}^{\infty} (p, x_1^j).\end{aligned}$$

Thus, it follows that $\rho_H(c) = 1$. It is easy to see that $(N_1, \lambda, \gamma) = (1, 1, 1)$ is a linear representation of c □

Example 4.8 Suppose $X = \{x_0, x_1\}$. Consider a linear series $c \in \mathbb{R}\langle\langle X\rangle\rangle$ with coefficients

$$(c, \eta) = \begin{cases} CA^k B_1 & : \ \eta = x_0^k x_1, \ \ k \geq 0\\ 0 & : \ \text{otherwise,}\end{cases}$$

where (A, B_1, C) is the controllability canonical form of c:

$$A = \begin{bmatrix} 0 & -(\tilde{a}, \emptyset)\\ 1 & -(\tilde{a}, x_0)\end{bmatrix}, \ \ B_1 = \begin{bmatrix} 1\\ 0\end{bmatrix}, \ \ C = \begin{bmatrix} (c, x_1) & (c, x_0 x_1)\end{bmatrix}.$$

This series is rational since it can be directly verified that $c = ba^{-1}x_1$ with

$$a = 1 + (a, x_0)x_0 + (a, x_0^2)x_0^2$$
$$b = (b, \emptyset) + (b, x_0)x_0,$$

where $(a, x_0^i) = (\tilde{a}, x_0^{2-i})$ for $i = 1, 2$, $(b, \emptyset) = (c, x_1)$ and $(b, x_0) = (c, x_0x_1) + (c, x_1)(a, x_0)$ (see Section 1.4 and Problem 4.4.3). The inductive procedure used in the previous proof provides one way to synthesize a linear representation of c, but surprisingly it is distinct from the triple (μ_A, B_1, C). That is, what one commonly calls a *realization* for this linear series does *not* directly yield a linear representation of c. The Hankel matrix for c is

$$\begin{array}{c} \\ \\ \mathcal{H}_c = \\ \\ \\ \end{array}
\begin{array}{c}
\begin{array}{cccccccc} \emptyset & x_0 & x_1 & x_0^2 & x_0x_1 & x_1x_0 & x_1^2 & \end{array} \\
\left[\begin{array}{cccccccc}
0 & 0 & CB_1 & 0 & CAB_1 & 0 & 0 & \cdots \\
0 & 0 & CAB_1 & 0 & CA^2B_1 & 0 & 0 & \cdots \\
CB_1 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
0 & 0 & CA^2B_1 & 0 & CA^3B_1 & 0 & 0 & \cdots \\
CAB_1 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}\right]
\begin{array}{l} \emptyset \\ x_0 \\ x_1 \\ x_0^2 \\ x_0x_1 \\ x_1x_0 \\ x_1^2 \\ \\ \end{array} \\
\begin{array}{cccccccc} \uparrow & & \uparrow & & \uparrow & & & \\ \eta_1 & & \eta_2 & & \eta_3 & & & \end{array}
\end{array}$$

If the polynomials $\tilde{b}$ and $\tilde{a}$ share no common roots, then the columns corresponding the words $\eta_1 = \emptyset$, $\eta_2 = x_1$, and $\eta_3 = x_0x_1$ are linearly independent, while all columns to the right of these columns lie in their span. Thus, $\rho_H(c) = 3$. According to equation (4.7) and employing the Cayley-Hamilton Theorem,

$$\begin{aligned}
\mathcal{H}_c(x_0\eta_1) &= 0 \\
\mathcal{H}_c(x_0\eta_2) &= \mathcal{H}_c(\eta_3) \\
\mathcal{H}_c(x_0\eta_3) &= -(\tilde{a}, \emptyset)\mathcal{H}_c(\eta_2) - (\tilde{a}, x_0)\mathcal{H}_c(\eta_3)
\end{aligned}$$

$$\begin{aligned}
\mathcal{H}_c(x_1\eta_1) &= \mathcal{H}_c(\eta_2) \\
\mathcal{H}_c(x_1\eta_2) &= 0 \\
\mathcal{H}_c(x_1\eta_3) &= 0.
\end{aligned}$$

The corresponding homomorphism, μ_N, is therefore uniquely specified by $N = \{N_0, N_1\}$, where

$$N_0 = \left[\begin{array}{c|cc} 0 & 0 & 0 \\ \hline 0 & 0 & -(\tilde{a},\emptyset) \\ 0 & 1 & -(\tilde{a},x_0) \end{array}\right], \quad N_1 = \left[\begin{array}{c|cc} 0 & 0 & 0 \\ \hline 1 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right].$$

The remaining portion of the linear representation follows from this same construction to be

$$\lambda = \left[\begin{array}{c|cc} 0 & (c,x_1) & (c,x_0x_1) \end{array}\right], \quad \gamma = \left[\begin{array}{c} 1 \\ \hline 0 \\ 0 \end{array}\right].$$

Clearly, the triple (A, B_1, C) is imbedded in this representation. (See Problem 4.4.4 for the most general case.) Note that $\mathcal{H}_c$ given above is a Hankel matrix for a series in two letters, while in contrast, the Hankel matrix $\mathcal{H}_{c_i}$ employed in Chapter 1 (equation (1.27)), is a Hankel matrix for a series in only one letter, namely the series $b_i a_i^{-1}$ in x_0. □

The link between rationality of a series and its Hankel mapping can also be seen through a canonical factorization of the latter.[1] Specifically, the finiteness of the Hankel rank implies the existence of a factorization of $\mathcal{H}_c = Q_c P_c$, where $P_c : \mathbb{R}\langle X\rangle \to \mathbb{R}\langle X\rangle$ is an $\mathbb{R}$-vector space homomorphism whose range has dimensional $\rho_H(c)$, and $Q_c : \mathbb{R}\langle X\rangle \to \mathbb{R}\langle\langle X\rangle\rangle$ is an $\mathbb{R}$-vector space isomorphism onto the range of $\mathcal{H}_c$. To produce such a factorization, define the following equivalence relation on $\mathbb{R}\langle X\rangle$:

$$p \sim p' \Longleftrightarrow \mathcal{H}_c(p) = \mathcal{H}_c(p') \Longleftrightarrow p - p' \in \mathrm{null}(\mathcal{H}_c).$$

In which case, the canonical factorization induced by this relation yields the desired mappings:

$$P_c : \mathbb{R}\langle X\rangle \to \mathbb{R}\langle X\rangle/\sim$$
$$Q_c : \mathbb{R}\langle X\rangle/\sim \to \mathbb{R}\langle\langle X\rangle\rangle$$

(see Figure 4.2). The existence of the vector space isomorphism Q_c means that $\mathbb{R}\langle X\rangle/\sim$ is a $\rho_H(c)$ dimensional subspace of $\mathbb{R}\langle X\rangle$. It is also an $\mathbb{R}\langle X\rangle$-module under the product

$$x_i \cdot P_c(p) := P_c(px_i), \quad x_i \in X, \quad p \in \mathbb{R}\langle X\rangle \tag{4.9}$$

[1] See Section A.2 in Appendix A.

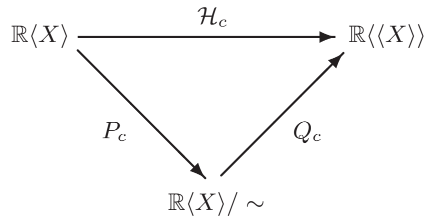


Fig. 4.2. The canonical factorization of the Hankel mapping $\mathcal{H}_c : \mathbb{R}\langle X\rangle \mapsto \mathbb{R}\langle\langle X\rangle\rangle$.

(see Problem 4.4.2).

Next it is shown that this canonical factorization of $\mathcal{H}_c$ leads directly to an $\rho_H(c)$ dimensional representation of c. Set $Z = \mathbb{R}\langle X\rangle/\sim$ and define for each $x_i \in X$ the $\mathbb{R}$-linear map

$$X_i : Z \to Z : z \mapsto x_i \cdot z.$$

Fix a basis for Z, and let N_i denote any matrix representation of X_i in this basis. Define row vector γ to be $P_c(\mathbf{1}) \in Z$ written in terms of this basis. Likewise, the column vector λ is a representation of the $\mathbb{R}$-linear map $Z \to \mathbb{R} : p \mapsto (Q_c(z), \emptyset)$ in terms of this basis. Since $c = \mathcal{H}_c(\mathbf{1})$, it follows that for any $\eta = x_{i_1}x_{i_2}\cdots, x_{i_k} \in X^*$:

$$\begin{aligned}(c, x_{i_1}x_{i_2}\cdots, x_{i_k}) &= (\mathcal{H}_c(\mathbf{1}), x_{i_1}x_{i_2}\cdots x_{i_k}) \\ &= (\mathcal{H}_c(x_{i_1}x_{i_2}\cdots x_{i_k}), \emptyset) \\ &= (Q_cP_c(x_{i_1}x_{i_2}\cdots x_{i_k}), \emptyset) \\ &= (Q_cX_{i_k}P_c(x_{i_1}x_{i_2}\cdots x_{i_{k-1}}), \emptyset) \\ &= (Q_cX_{i_{k-1}}X_{i_k}P_c(x_{i_1}x_{i_2}\cdots x_{i_{k-2}}), \emptyset) \\ &\;\vdots \\ &= (Q_c(X_{i_1}X_{x_{i_2}}\cdots X_{i_k}P_c(\mathbf{1})), \emptyset) \\ &= \lambda N_{i_1}N_{i_2}\cdots N_{i_k}\gamma,\end{aligned}$$

and $(c, \emptyset) = (\mathcal{H}_c(\mathbf{1}), \emptyset) = (Q_c(P_c(\mathbf{1})), \emptyset) = \lambda\gamma$.

A linear representation of c is said to be *minimal* if there exists no other linear representation of c with a lower dimension. One advantage of the Hankel matrix characterization of rationality is that it provides the minimal dimension of its linear representations.

Theorem 4.5 *The minimal dimension of a linear representation of a rational series $c \in \mathbb{R}\langle\langle X\rangle\rangle$ is equivalent to its Hankel rank, $\rho_H(c)$.*

Proof: Suppose c has a linear representation of dimension $n < \rho_H(c)$. Then its underlying maps X_i are on a vector space Z of dimension n. In which case, the linear maps Q_c and P_c in the canonical factorization of $\mathcal{H}_c$ have matrix representations with rank not exceeding n. Therefore, $\mathrm{rank}(\mathcal{H}_c) = \mathrm{rank}(Q_cP_c) \leq n < \rho_H(c)$, which is a contradiction. ■

If (μ, γ, λ) is a minimal linear representation of c then clearly so is $\mathcal{A}_T(\mu, \gamma, \lambda)$ for any $T \in GL_n(\mathbb{R})$. What is not so evident is whether all the minimal linear representations of c lie on an *orbit* of the action $\mathcal{A}$, namely $\mathcal{O}_{(\mu,\gamma,\lambda)}(\mathcal{A}) := \{\mathcal{A}_T(\mu, \gamma, \lambda) : T \in GL_n(\mathbb{R})\}$. The following theorem establishes this claim.

Theorem 4.6 *Let (μ_N, γ, λ) and $(\mu_{\tilde{N}}, \tilde{\gamma}, \tilde{\lambda})$, be two minimal linear representations of a rational series $c \in \mathbb{R}\langle\langle X\rangle\rangle$. Then there exists a matrix $T \in GL_n(\mathbb{R})$ such that $(\mu_{\tilde{N}}, \tilde{\gamma}, \tilde{\lambda}) = \mathcal{A}_T(\mu_N, \gamma, \lambda)$.*

Proof: Since (μ, γ, λ) and $(\tilde{\mu}, \tilde{\gamma}, \tilde{\lambda})$ are both minimal linear representations of the same series c, they have the same Hankel matrix $\mathcal{H}_c$. The canonical factorization $\mathcal{H}_c = Q_cP_c$ depends only on the minimal dimension $\rho_H(c)$ of the vector space $Z = \mathbb{R}\langle X\rangle/\sim$. As each linear map $X_i : Z \to Z : z \mapsto x_i \cdot z$ is independently defined of any particular minimal representation, any two matrix representations of it must be related by similarity. As all such matrices in a given representation are by assumption written in terms of the same basis for Z, $\mu_{\tilde{N}} = \mu_{TNT^{-1}}$ for some $T \in GL_n(\mathbb{R})$. Furthermore, since for every $\eta \in X^*$

$$\begin{aligned}(c, \eta) &= \tilde{\lambda}\mu_{\tilde{N}}(\eta)\tilde{\gamma} \\ &= \tilde{\lambda}T\mu_N(\eta)T^{-1}\tilde{\gamma} \\ &= \lambda\mu_N(\eta)\gamma.\end{aligned}$$

It follows that $\tilde{\lambda} = \lambda T^{-1}$ and $\tilde{\gamma} = T\gamma$, hence proving the theorem. ■

Example 4.9 Example 4.7 can be generalized to an arbitrary alphabet $X = \{x_0, x_1, \ldots, x_m\}$ by considering the series $c = \sum_{\eta\in X^*} \eta$. As before, for any $p \in \mathbb{R}\langle X\rangle$ and $\eta \in X^*$ observe

$$(\mathcal{H}_c(p), \eta) = \sum_{\xi\in X^*} (p, \xi)(c, \eta\xi)$$

$$= \sum_{\xi \in X^*} (p, \xi),$$

and thus,

$$\begin{aligned} \mathcal{H}_c(p) &= \sum_{\eta \in X^*} \sum_{\xi \in X^*} (p, \xi)\, \eta \\ &= c \sum_{\xi \in X^*} (p, \xi). \end{aligned}$$

Therefore, $\rho_H(c) = 1$. In this case, c has a minimal linear representation $N_i = 1$, $i = 0, 1, \ldots, m$ and $\lambda = \gamma = 1$. □

4.5 The Shuffle and Composition Products as Rational Operations

It is possible that other binary operators on $\mathbb{R}\langle\langle X \rangle\rangle$ may be rational in the sense that they produce a rational series when their arguments are rational series. The first theorem states that the shuffle product is one such example.

Theorem 4.7 *If c and d are rational series in $\mathbb{R}\langle\langle X \rangle\rangle$, then $c \sqcup\!\sqcup d$ is also a rational series in $\mathbb{R}\langle\langle X \rangle\rangle$.*

Proof: In light of Theorems 4.2 and 4.3, let V_c and V_d be any stable finite dimensional $\mathbb{R}$-vector subspaces of $\mathbb{R}\langle\langle X \rangle\rangle$ containing c and d, respectively. Let $\{\bar{c}_i\}_{i=1}^{n_c}$ and $\{\bar{d}_i\}_{i=1}^{n_d}$ be any corresponding pair of bases for V_c and V_d. Define the subspace of $\mathbb{R}\langle\langle X \rangle\rangle$:

$$V_{c \sqcup\!\sqcup d} = \text{span}\,\{\bar{c}_i \sqcup\!\sqcup \bar{d}_j : i = 1, \ldots, n_c;\ j = 1, \ldots, n_d\}.$$

Clearly $V_{c \sqcup\!\sqcup d}$ is a finite dimensional subspace of $\mathbb{R}\langle\langle X \rangle\rangle$. Setting

$$c = \sum_{i=1}^{n_c} \alpha_i \bar{c}_i \ \text{ and } \ d = \sum_{j=1}^{n_d} \beta_j \bar{d}_j$$

it follows that

$$c \sqcup\!\sqcup d = \sum_{i,j=1}^{n_c, n_d} \alpha_i \beta_j\, \bar{c}_i \sqcup\!\sqcup \bar{d}_j \in V_{c \sqcup\!\sqcup d}.$$

To see that $V_{c \sqcup\!\sqcup\, d}$ is stable, observe that for any $x \in X$

$$x^{-1}(\bar{c}_i \sqcup\!\sqcup\, \bar{d}_j) = x^{-1}(\bar{c}_i) \sqcup\!\sqcup\, \bar{d}_j + \bar{c}_i \sqcup\!\sqcup\, x^{-1}(\bar{d}_j) \tag{4.10}$$

for all i and j (see Theorem 2.5). Since V_c and V_d are stable, it follows that $x^{-1}(\bar{c}_i \sqcup\!\sqcup\, \bar{d}_j) \in V_{c \sqcup\!\sqcup\, d}$, and thus, $V_{c \sqcup\!\sqcup\, d}$ is stable (see Problem 4.2.3(c)). Hence, $c \sqcup\!\sqcup\, d$ is recognizable, and therefore rational. ■

The following theorem gives a specific linear representation of the shuffle product of two series.

Theorem 4.8 *If $c \in \mathbb{R}\langle\langle X \rangle\rangle$ has a linear representation $(\mu_c, \gamma_c, \lambda_c)$, and $d \in \mathbb{R}\langle\langle X \rangle\rangle$ has representation $(\mu_d, \gamma_d, \lambda_d)$, then $c \sqcup\!\sqcup\, d$ has a linear representation $(\mu, \gamma_c \otimes \gamma_d, \lambda_c \otimes \lambda_d)$, where for each $x_i \in X$, $\mu(x_i) = \mu_c(x_i) \otimes I_{n_d} + I_{n_c} \otimes \mu_d(x_i)$.*

Proof: For any fixed $\nu \in X^*$ and using ideas from Example 2.19, observe that

$$\begin{aligned}(c \sqcup\!\sqcup\, d, \nu) &= \sum_{\eta,\xi \in X^*} (c, \eta)(d, \xi)(\eta \sqcup\!\sqcup\, \xi, \nu) \\ &= \sum_{\eta,\xi \in X^*} \lambda_c \mu_c(\eta) \gamma_c \, \lambda_d \mu_d(\xi) \gamma_d \, (\eta \otimes \xi, \mathrm{sh}^*(\nu)) \\ &= (\lambda_c \otimes \lambda_d) \left[\sum_{\eta,\xi \in X^*} (\mu_c(\eta) \otimes \mu_d(\xi))(\mathrm{sh}^*(\nu), \eta \otimes \xi) \right] (\gamma_c \otimes \gamma_d).\end{aligned}$$

The claim is now proved if it can be shown that

$$\mu(\nu) = \sum_{\eta,\xi \in X^*} (\mu_c(\eta) \otimes \mu_d(\xi))(\mathrm{sh}^*(\nu), \eta \otimes \xi).$$

It is clear for $\nu = \emptyset$ that

$$\begin{aligned}\sum_{\eta,\xi \in X^*} (\mu_c(\eta) \otimes \mu_d(\xi))(\mathrm{sh}^*(\emptyset), \eta \otimes \xi) &= \mu_c(\emptyset) \otimes \mu_d(\emptyset) \\ &= I_{n_c} \otimes I_{n_d} \\ &= \mu(\emptyset).\end{aligned}$$

If $\nu = x_i \in X$, then

$$\sum_{\eta,\xi \in X^*} (\mu_c(\eta) \otimes \mu_d(\xi))(\mathrm{sh}^*(x_i), \eta \otimes \xi)$$

$$
\begin{aligned}
&= \sum_{\eta,\xi\in X^*} (\mu_c(\eta)\otimes\mu_d(\xi))(x_i\otimes \mathbf{1} + \mathbf{1}\otimes x_i, \eta\otimes\xi) \\
&= \mu_c(x_i)\otimes I_{n_d} + I_{n_c}\otimes \mu_d(x_i) \\
&= \mu(x_i).
\end{aligned}
$$

Now assume the identity holds for words up to length $k\geq 0$, and let $\nu\in X^k$. Given any $x_i\in X$ observe that

$$
\begin{aligned}
&\sum_{\eta,\xi\in X^*} (\mu_c(\eta)\otimes\mu_d(\xi))(\mathrm{sh}^*(x_i\nu), \eta\otimes\xi) \\
&\qquad = \sum_{\eta,\xi\in X^*} (\mu_c(\eta)\otimes\mu_d(\xi))(\mathrm{sh}^*(x_i)\mathrm{sh}^*(\nu), \eta\otimes\xi) \\
&\qquad = \sum_{\eta,\xi\in X^*} (\mu_c(\eta)\otimes\mu_d(\xi))((x_i\otimes \mathbf{1} + \mathbf{1}\otimes x_i)\mathrm{sh}^*(\nu), \eta\otimes\xi) \\
&\qquad = \sum_{\eta,\xi\in X^*} (\mu_c(\eta)\otimes\mu_d(\xi))(\mathrm{sh}^*(\nu), x_i^{-1}(\eta)\otimes\xi) + \\
&\qquad\qquad \sum_{\eta,\xi\in X^*} (\mu_c(\eta)\otimes\mu_d(\xi))(\mathrm{sh}^*(\nu), \eta\otimes x_i^{-1}(\xi)) \\
&\qquad = \sum_{\eta,\xi\in X^*} (\mu_c(x_i\eta)\otimes\mu_d(\xi))(\mathrm{sh}^*(\nu), \eta\otimes\xi) + \\
&\qquad\qquad \sum_{\eta,\xi\in X^*} (\mu_c(\eta)\otimes\mu_d(x_i\xi))(\mathrm{sh}^*(\nu), \eta\otimes\xi) \\
&\qquad = \sum_{\eta,\xi\in X^*} (\mu_c(x_i)\mu_c(\eta)\otimes\mu_d(\xi))(\mathrm{sh}^*(\nu), \eta\otimes\xi) + \\
&\qquad\qquad \sum_{\eta,\xi\in X^*} (\mu_c(\eta)\otimes\mu_d(x_i)\mu_d(\xi))(\mathrm{sh}^*(\nu), \eta\otimes\xi) \\
&\qquad = (\mu_c(x_i)\otimes I_{n_d} + I_{n_c}\otimes\mu_d(x_i))\mu(\nu) \\
&\qquad = \mu(x_i)\mu(\nu) \\
&\qquad = \mu(x_i\nu),
\end{aligned}
$$

where the induction hypothesis was employed in the third to the last step. Thus, the identity holds for all $\nu\in X^*$, and the theorem is proved.

■

Example 4.10 Let $X_0=\{x_0\}$ and consider two series $c,d\in\mathbb{R}[[X_0]]$ with linear representations

$$(c, x_0^k) = C_c A_c^k z_c, \quad (d, x_0^k) = C_d A_d^k z_d, \quad k \geq 0.$$

It was shown in Example 3.13 that $c \sqcup\!\sqcup d$ has the linear representation $(A_c \oplus A_d, C_c \otimes C_d, z_c \otimes z_d)$. In the context of the previous theorem,

$$\mu(x_0) = (A_c \otimes I_{n_d}) + (I_{n_c} \otimes A_d) = A_c \oplus A_d,$$

so that $\mu(x_0^k) = (A_c \oplus A_d)^k$, $k \geq 0$. □

Example 4.11 Suppose $X = \{x_0, x_1\}$. Consider two rational linear series $c, d \in \mathbb{R}\langle\langle X \rangle\rangle$ with linear representations coming from linear state space realizations (A_c, B_c, C_c) and (A_d, B_d, C_d), respectively, as in Example 4.8. In which case, $c \sqcup\!\sqcup d$ has the linear representation

$$\begin{aligned}
\mu(x_0) &= \begin{bmatrix} 0 & 0 \\ 0 & A_c \end{bmatrix} \otimes I_{n_d} + I_{n_c} \otimes \begin{bmatrix} 0 & 0 \\ 0 & A_d \end{bmatrix} \\
\mu(x_1) &= \begin{bmatrix} 0 & 0 \\ B_c & 0 \end{bmatrix} \otimes I_{n_d} + I_{n_c} \otimes \begin{bmatrix} 0 & 0 \\ B_d & 0 \end{bmatrix} \\
\gamma &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \lambda = \begin{bmatrix} 0 & C_c \end{bmatrix} \otimes \begin{bmatrix} 0 & C_d \end{bmatrix}.
\end{aligned}$$

□

Next consider the composition product. The following example reveals that the composition product is *not* in general a rational operation. This has important consequences in system theory as it implies that a cascade of two rational systems (i.e., systems with rational generating series) do not always produce another rational system.

Example 4.12 Suppose $X = \{x_0, x_1\}$ and consider the rational series $c = (1 - x_1)^{-1} = x_1^*$. The claim is that c composed with itself is not rational. The main goal is to show that

$$(c \circ c, x_0^{k_0} x_1^{k_1}) = (k_0)^{k_1}, \quad k_0 \geq 0, \quad k_1 \geq 0,$$

or equivalently,

$$(x_1^{-k_1} x_0^{-k_0}(c \circ c), \emptyset) = (k_0)^{k_1}. \tag{4.11}$$

The claim is trivial when $k_0 = k_1 = 0$ provided that $0^0 := 1$. If $k_0 = 1$ and $k_1 = 0$, observe that

$$x_0^{-1}(c \circ c) = \underbrace{x_0^{-1}(c)}_{0} \circ c + c \sqcup\!\sqcup (\underbrace{x_1^{-1}(c)}_{c} \circ c)$$
$$= c \sqcup\!\sqcup (c \circ c).$$

The intermediate claim then is that

$$x_0^{-k_0}(c \circ c) = c^{\sqcup\!\sqcup\, k_0} \sqcup\!\sqcup (c \circ c), \quad k_0 \geq 1.$$

If the identity above holds up to some fixed $k_0 \geq 1$ then

$$\begin{aligned} x_0^{-k_0-1}(c \circ c) &= x_0^{-1}(c^{\sqcup\!\sqcup\, k_0} \sqcup\!\sqcup (c \circ c)) \\ &= x_0^{-1}(c^{\sqcup\!\sqcup\, k_0}) \sqcup\!\sqcup (c \circ c) + c^{\sqcup\!\sqcup\, k_0} \sqcup\!\sqcup x_0^{-1}(c \circ c) \\ &= \left[k_0 c^{\sqcup\!\sqcup\,(k_0-1)} \sqcup\!\sqcup \underbrace{x_0^{-1}(c)}_{0} \right] \sqcup\!\sqcup (c \circ c) + \\ &\quad\; c^{\sqcup\!\sqcup\, k_0} \sqcup\!\sqcup (c \sqcup\!\sqcup (c \circ c)) \\ &= c^{\sqcup\!\sqcup\,(k_0+1)} \sqcup\!\sqcup (c \circ c). \end{aligned}$$

(The identities in Problem 2.4.5(h) and 2.7.7(f) were employed above.) Hence, the intermediate identity in question holds for $k_0 \geq 0$. Next observe that

$$\begin{aligned} x_1^{-1} x_0^{-k_0}(c \circ c) &= x_1^{-1}(c^{\sqcup\!\sqcup\, k_0} \sqcup\!\sqcup (c \circ c)) \\ &= x_1^{-1}(c^{\sqcup\!\sqcup\, k_0}) \sqcup\!\sqcup (c \circ c) + c^{\sqcup\!\sqcup\, k_0} \sqcup\!\sqcup \underbrace{x_1^{-1}(c \circ c)}_{0} \\ &= k_0 c^{\sqcup\!\sqcup\,(k_0-1)} \sqcup\!\sqcup \underbrace{x_1^{-1}(c)}_{c} \sqcup\!\sqcup (c \circ c) \\ &= k_0 c^{\sqcup\!\sqcup\, k_0} \sqcup\!\sqcup (c \circ c). \end{aligned}$$

The second intermediate claim is that

$$x_1^{-k_1} x_0^{-k_0}(c \circ c) = (k_0)^{k_1} c^{\sqcup\!\sqcup\, k_0} \sqcup\!\sqcup (c \circ c).$$

If this is the case up to some fixed $k_1 \geq 1$ then

$$\begin{aligned} x_1^{-k_1-1} x_0^{-k_0}(c \circ c) &= x_1^{-1}((k_0)^{k_1} c^{\sqcup\!\sqcup\, k_0} \sqcup\!\sqcup (c \circ c)) \\ &= (k_0)^{k_1} \Big[x_1^{-1}(c^{\sqcup\!\sqcup\, k_0}) \sqcup\!\sqcup (c \circ c) + \\ &\qquad c^{\sqcup\!\sqcup\, k_0} \sqcup\!\sqcup \underbrace{x_1^{-1}(c \circ c)}_{0} \Big] \end{aligned}$$

$$= (k_0)^{k_1}\left[k_0 c^{\sqcup\!\sqcup\, k_0} \sqcup\!\sqcup\, (c \circ c)\right]$$
$$= (k_0)^{k_1+1} c^{\sqcup\!\sqcup\, k_0} \sqcup\!\sqcup\, (c \circ c).$$

Hence, this claim holds for all $k_1, k_0 \geq 0$. To validate (4.11), simply compare the constant coefficients in the above identity:

$$(x_1^{-k_1} x_0^{-k_0}(c \circ c), \emptyset) = ((k_0)^{k_1} c^{\sqcup\!\sqcup\, k_0} \sqcup\!\sqcup\, (c \circ c), \emptyset)$$
$$(c \circ c, x_0^{k_0} x_1^{k_1}) = (k_0)^{k_1}.$$

Setting $k_0 = k_1$ yields the expression

$$(c \circ c, x_0^k x_1^k) = k^k, \quad k \geq 0. \tag{4.12}$$

The key observation is that these coefficients are growing faster than any sequence of coefficients from a rational series can possibly grow. Specifically, from Theorem 4.1 it is known that every rational series has coefficients satisfying a growth rate with Gevrey order $s = 0$. But k^k, $k \geq 0$ is clearly growing faster than this. So $c \circ c$ can not be rational. (See Problem 4.5.6 for additional analysis of this example.) □

There are special conditions under which the composition of two rational series is again rational. The following definition provides one such condition.

Definition 4.7 *A series $c \in \mathbb{R}\langle\langle X \rangle\rangle$ is* ***limited relative to x_i*** *if there exists an integer $\mathcal{N}_i \geq 0$ such that*

$$\max_{\eta \in \text{supp}(c)} |\eta|_{x_i} \leq \mathcal{N}_i.$$

If c is limited relative to x_i for every $i = 1, \ldots, m$ then c is ***input-limited****. In such cases, let $\mathcal{N}_c := \max_i \mathcal{N}_i$.*

Clearly any linear series c has this property (see Example 2.33), specifically $\mathcal{N}_c = 1$, while the series c given in Example 4.12 does not. The following theorem states that the input-limited property is a sufficient condition for preserving rationality under composition. It is easily shown by counterexample that this condition is not necessary (see Problem 4.5.4). Here a series $d \in \mathbb{R}^m\langle\langle X \rangle\rangle$ is said to be rational when each component series d_i, $i = 1, 2, \ldots, m$ is rational.

Theorem 4.9 *Let $c \in \mathbb{R}\langle\langle X\rangle\rangle$ and $d \in \mathbb{R}^m\langle\langle X\rangle\rangle$ be two rational series. If c is input-limited then $c \circ d \in \mathbb{R}\langle\langle X\rangle\rangle$ is rational.*

The proof of this theorem relies on the following lemma which uses an operator monoid reminiscent of the algebra homomorphisms ψ and ϕ utilized in Chapters 2 and 3. Note that the shuffle product on $\mathbb{R}\langle\langle X\rangle\rangle \times \mathbb{R}^{n\times n}\langle\langle X\rangle\rangle$ is defined componentwise.

Lemma 4.2 *Let $c \in \mathbb{R}\langle\langle X\rangle\rangle$ be a rational series with a linear representation (μ, γ, λ). Let $N_i := \mu(x_i) \in \mathbb{R}^{n\times n}$, $i = 0, 1, \ldots, m$. Then for any $d \in \mathbb{R}^m\langle\langle X\rangle\rangle$ it follows that*

$$c \circ d = \sum_{\eta \in \tilde{X}^*} \lambda \Psi_\eta((N_0 x_0)^*)\gamma,$$

where $\tilde{X} := \{x_1, x_2, \ldots, x_m\}$, and the set of operators $\{\Psi_\eta : \eta \in \tilde{X}^*\}$ is the monoid under operator composition involving the following operators

$$\Psi_{x_i} : \mathbb{R}^{n\times n}\langle\langle X\rangle\rangle \to \mathbb{R}^{n\times n}\langle\langle X\rangle\rangle : E \mapsto x_0(N_0x_0)^* N_i(d_i \sqcup\!\sqcup E)$$

with $\Psi_\emptyset$ equivalent to the identity map on $\mathbb{R}^{n\times n}\langle\langle X\rangle\rangle$.

Proof: Observe from the definition of the composition product that

$$\begin{aligned}
c \circ d = &\sum_{k=0}^{\infty} \sum_{i_1,\ldots,i_k=1}^{m} \sum_{n_0,\ldots,n_k=0}^{\infty} \lambda N_0^{n_k} N_{i_k} N_0^{n_{k-1}} N_{i_{k-1}} \cdots N_0^{n_1} N_{i_1} N_0^{n_0}\gamma \cdot \\
&\quad (x_0^{n_k} x_{i_k} x_0^{n_{k-1}} x_{i_{k-1}} \cdots x_0^{n_1} x_{i_1} x_0^{n_0}) \circ d \\
= &\sum_{k=0}^{\infty} \sum_{i_1,\ldots,i_k=1}^{m} \sum_{n_0,\ldots,n_k=0}^{\infty} \lambda N_0^{n_k} N_{i_k} N_0^{n_{k-1}} N_{i_{k-1}} \cdots N_0^{n_1} N_{i_1} N_0^{n_0}\gamma \cdot \\
&\quad x_0^{n_k+1}\left[d_{i_k} \sqcup\!\sqcup \left[x_0^{n_{k-1}+1}\left[d_{i_{k-1}} \sqcup\!\sqcup \cdots x_0^{n_1+1}\left[d_{i_1} \sqcup\!\sqcup x_0^{n_0}\right]\cdots\right]\right]\right].
\end{aligned}$$

From the bilinearity and continuity of the shuffle product (in the ultrametric sense), it follows that

$$\begin{aligned}
c \circ d = &\sum_{k=0}^{\infty} \sum_{i_1,\ldots,i_k=1}^{m} \lambda x_0 \left(\sum_{n_k=0}^{\infty} (N_0x_0)^{n_k}\right) N_{i_k}\left[d_{i_k} \sqcup\!\sqcup \cdot\right. \\
&\quad \left[x_0\left(\sum_{n_{k-1}=0}^{\infty} (N_0x_0)^{n_{k-1}}\right) N_{i_{k-1}}\left[d_{i_{k-1}} \sqcup\!\sqcup \cdots\right.\right.
\end{aligned}$$

$$
\begin{aligned}
& x_0\left(\sum_{n_1=0}^{\infty}(N_0x_0)^{n_1}\right)N_{i_1}\left[d_{i_1}\sqcup\!\sqcup\left(\sum_{n_0=0}^{\infty}(N_0x_0)^{n_0}\right)\right]\cdots\Big]\Big]\Big]\gamma \\
&= \sum_{k=0}^{\infty}\sum_{i_1,\ldots,i_k=1}^{m}\lambda x_0(N_0x_0)^*N_{i_k}\left[d_{i_k}\sqcup\!\sqcup\left[x_0(N_0x_0)^*N_{i_{k-1}}\left[d_{i_{k-1}}\sqcup\!\sqcup\right.\right.\right.\cdots \\
&\qquad x_0(N_0x_0)^*N_{i_1}\left[d_{i_1}\sqcup\!\sqcup(N_0x_0)^*\right]\cdots]]]\,\gamma.
\end{aligned}
$$

Finally, applying the definition of Ψ_η,

$$
\begin{aligned}
c\circ d &= \sum_{k=0}^{\infty}\sum_{x_{i_k}\cdots x_{i_1}\in\tilde{X}^k}\lambda\Psi_{x_{i_k}}\Psi_{x_{i_{k-1}}}\cdots\Psi_{x_{i_1}}((N_0x_0)^*)\gamma \\
&= \sum_{\eta\in\tilde{X}^*}\lambda\Psi_\eta((N_0x_0)^*)\gamma,
\end{aligned}
$$

and the lemma is proved. ■

Proof of Theorem 4.9: Since c is input-limited, it follows from Lemma 4.2 that

$$
c\circ d = \sum_{k=0}^{\mathcal{N}_c}\sum_{\eta\in\tilde{X}^k}\lambda\Psi_\eta((N_0x_0)^*)\gamma.
$$

Clearly each operator Ψ_η is mapping a rational series to another rational series as it involves only a finite number of rational operations. Therefore, for any integer $k\geq 0$ the formal power series

$$
\sum_{\eta\in\tilde{X}^k}\lambda\Psi_\eta((N_0x_0)^*)\gamma
$$

is again rational since the summation is finite. Thus, $c\circ d$ must be rational. ■

Example 4.13 Let $X=\{x_0,x_1\}$ and consider two rational linear series $c,d\in\mathbb{R}\langle\langle X\rangle\rangle$ with linear representations as in Example 4.11. In light of the linearity of the series and Theorem 4.9, $c\circ d$ must also be rational. Specifically, for any $k\geq 0$ observe

$$
(c\circ d, x_0^kx_1) = \sum_{j=0}^{k-1}(c, x_0^{k-1-j}x_1)(d, x_0^jx_1)
$$

$$
\begin{aligned}
&= \sum_{j=0}^{k-1} (C_c A_c^{k-1-j} B_c)(C_d A_d^j B_d) \\
&= C_c \left[\sum_{j=0}^{k-1} A_c^{k-1-j} B_c C_d A_d^j \right] B_d \\
&= \left[\begin{array}{cc} C_c & 0 \end{array}\right] \left[\begin{array}{cc} A_c & B_c C_d \\ 0 & A_d \end{array}\right]^k \left[\begin{array}{c} 0 \\ B_d \end{array}\right].
\end{aligned}
$$

Thus, a linear representation for $c \circ d$ is immediately evident. □

It is noted in closing that the input-limited condition does not apply to the feedback product. That is, if c and d are rational with c input-limited, it is not the case that $c@d$ is necessarily rational.

Example 4.14 Suppose $c = 1 + x_1 x_0^*$, which is clearly input-limited and rational. Consider the system F_c in a feedback interconnection with itself. It can be readily checked that $c@c$ and $c@\delta$ satisfy the same fixed point equation, and thus are equivalent series. Now define $e = c@\delta|_{x_1 \to 0}$ in $\mathbb{R}[[X_0]]$. This corresponds to the generating series for the self-excited unity feedback system, and in this case the fixed point equation reduces to

$$
\begin{aligned}
e &= (1 + x_1 x_0^*) \circ e \\
&= 1 + \sum_{k=0}^{\infty} x_1 x_0^k \circ e = 1 + \sum_{k=0}^{\infty} x_0 (e \sqcup\!\sqcup\, x_0^k) \\
&= 1 + x_0 (e \sqcup\!\sqcup\, x_0^*).
\end{aligned}
$$

Since $x_0^{-1}(e) = e \sqcup\!\sqcup\, x_0^*$ and $(e, \emptyset) = 1$, it follows that $y = F_e[0]$ satisfied the differential equation

$$
y'(t) = y(t)\, \mathrm{e}^t, \quad y(0) = 1,
$$

which has the solution

$$
y(t) = \mathrm{e}^{\mathrm{e}^t - 1}, \quad t \geq 0. \tag{4.13}
$$

Therefore, the coefficients of $(c@c, x_0^k) = (e, x_0^k)$ correspond to the Bell numbers, which first appear in Example 3.8 and are known to be growing at a rate faster than Gevrey order $s = 0$. Hence, in light of Theorem 4.1, $c@c$ can not be rational. □

Problems

Section 4.1

Problem 4.1.1 Let $c \in \mathbb{R}\langle\langle X\rangle\rangle$.

(a) Prove c is Cauchy invertible if and only if it is not proper.
(b) Show that if c is Cauchy invertible, then the inverse is unique.

Problem 4.1.2 Compute the Cauchy inverse of each of the following series in $\mathbb{R}\langle\langle X\rangle\rangle$ assuming $X = \{x_0, x_1\}$.

(a) $c = x_0 + x_1$
(b) $c = 1 + x_0 x_1$
(c) $c = 1 + x_0 + x_1 + x_0x_1 + x_1^2 + x_0x_1^2 + x_1^3 + x_0x_1^3 + x_1^4 + \cdots$

Problem 4.1.3 Show that $(\mathrm{char}(X))^* = \mathrm{char}(X^*)$.
Remark: See Problems 2.4.6 and 4.5.2.

Problem 4.1.4 Consider the mapping $\mathrm{dist}(\mathbf{A}, \mathbf{B}) = \max_{ij} \mathrm{dist}(a_{ij}, b_{ij})$, where $(\mathbf{A}, \mathbf{B}) \in \mathbb{R}\langle\langle X\rangle\rangle^{n\times n}$.

(a) Verify that dist is an ultrametric on $\mathbb{R}\langle\langle X\rangle\rangle^{n\times n}$.
(b) Show that if $\mathbf{C} \in \mathbb{R}\langle\langle X\rangle\rangle^{n\times n}$ is proper then $\mathbf{C}^*$ is well defined.
(c) Is the converse of the statement in (b) true? Explain.

Problem 4.1.5 Show that the matrix $\mathbf{C}^*$ defined in Lemma 4.1 is the unique solution to the matrix equations $(I_n - \mathbf{C})\mathbf{C}^* = I_n$ and $\mathbf{C}^*(I_n - \mathbf{C}) = I_n$.

Problem 4.1.6 Verify the expression for $\mathbf{C}^*$ in equation (4.2). The following identities are useful: $C_1^* C_4 \Delta_2^* = \Delta_1^* C_4 C_2^*$ and $C_2^* C_3 \Delta_1^* = \Delta_2^* C_3 C_1^*$.
Remark: Δ_1 and Δ_2 are related to the *Schur complements* found in block matrix inversion formulas.

Section 4.2

Problem 4.2.1 Show that a subset $V \subset \mathbb{R}\langle\langle X\rangle\rangle$ is stable if and only if $x^{-1}(c) \in V$ for all $c \in V$ and $x \in X$.

Problem 4.2.2 Let V_c and V_d be two stable subsets in $\mathbb{R}\langle\langle X\rangle\rangle$. Their *inner sum* is $V_c + V_d := \{\hat{c} + \hat{d} \in \mathbb{R}\langle\langle X\rangle\rangle : \hat{c} \in V_c,\ \hat{d} \in V_d\}$.

(a) Show that if V_c and V_d are stable then so is $V_c + V_d$.
(b) Likewise, show that if V_c and V_d are finite dimensional $\mathbb{R}$-vector subspaces of $\mathbb{R}\langle\langle X\rangle\rangle$ then the same is true of $V_c + V_d$.

Problem 4.2.3 Suppose V is a finite dimensional $\mathbb{R}$-vector subspace of $\mathbb{R}\langle\langle X\rangle\rangle$.

(a) Show that if V is stable then *every* series in V is rational, and in particular, every series in any given basis for V is rational. Is the converse true?
(b) A specific series c in V is said to be *stable with respect to* V if $\xi^{-1}(c) \in V$ for all $\xi \in X^*$. Show that each series $\bar{c}_i$ in a given basis $\{\bar{c}_i\}_{i=1}^n$ for V is stable with respect to V if and only if V is stable.
(c) Let $\{\bar{c}_i\}_{i=1}^n$ be a basis for V. Show that $x^{-1}(\bar{c}_i) \in V$ for all $x \in X$ and $i = 1, 2, \ldots, n$ if and only if V is stable.

Section 4.3

Problem 4.3.1 Let $c, d \in \mathbb{R}\langle\langle X\rangle\rangle$. Show that if V_c and V_d are stable subspaces of $\mathbb{R}\langle\langle X\rangle\rangle$ then so is

$$V_{cd} = \{\hat{c}d + \hat{d} : \hat{c} \in V_c,\ \hat{d} \in V_d\}.$$

Section 4.4

Problem 4.4.1 For any $c \in \mathbb{R}\langle\langle X\rangle\rangle$ and $p \in \mathbb{R}\langle X\rangle$, show that

(a) $\mathcal{H}_c(p) = \sum_{\nu,\eta \in X^*} (c, \nu\eta)(p, \eta)\, \nu$
(b) $\xi^{-1}(\mathcal{H}_c(p)) = \mathcal{H}_{\xi^{-1}(c)}(p)$, $\xi \in X^*$.

Problem 4.4.2 In Section 2.2, it was stated that the set of series $\mathbb{R}\langle\langle X\rangle\rangle$ forms an $\mathbb{R}\langle X\rangle$-module, where the product of a polynomial and a series is defined via the Cauchy product. Equation (4.9) introduces another $\mathbb{R}\langle X\rangle$-module on $\mathbb{R}\langle\langle X\rangle\rangle$, which is more natural for extracting representations of a series c from its Hankel mapping, $\mathcal{H}_c$.

(a) Completely describe this new $\mathbb{R}\langle X\rangle$-module on $\mathbb{R}\langle\langle X\rangle\rangle$.

(b) Contrast the definition in part (a) to that for the $\mathbb{R}\langle X\rangle$-module defined on $\mathbb{R}\langle\langle X\rangle\rangle$ by the Cauchy product.

Problem 4.4.3 Verify the rational factorization of c given in Example 4.8 by:

(a) Computing $H(s) = C(sI - A)^{-1}B_1$ and then applying (1.8).
(b) Computing the first two coefficients of the linear series $ba^{-1}x_1 = \sum_{i\geq 1} h_i x_0^{i-1}x_1$.

Problem 4.4.4 Let $X = \{x_0, x_1\}$. Consider the series $c \in \mathbb{R}\langle\langle X\rangle\rangle$ with coefficients

$$(c,\eta) = \begin{cases} CA^kB_1 & : \ \eta = x_0^k x_1, \ \ k \geq 0 \\ CA^k z_0 & : \ \eta = x_0^k, \ \ k \geq 0 \\ 0 & : \ \text{otherwise}, \end{cases}$$

where (A, B_1, C) is an n-dimensional linear state space realization in controllability canonical form and $z_0 \in \mathbb{R}^n$ is arbitrary.

(a) Show that c is a rational series.
(b) Derive a linear representation of c when $n = 2$ by choosing the same basis columns indexed by η_1, η_2, and η_3 from the Hankel matrix $\mathcal{H}_c$ as in Example 4.8.
(c) Is this representation of c minimal ? Explain.
(d) Show that the change of state space coordinates

$$T = \begin{bmatrix} 1 & 0 & 0 \\ z_{01} & 1 & 0 \\ z_{02} & 0 & 1 \end{bmatrix}$$

(that is, $\tilde{N}_i = TN_iT^{-1}$, $\tilde{\gamma} = T\gamma$ and $\tilde{\lambda} = \lambda T^{-1}$) applied to the representation found in part (b) produces the linear representation of c:

$$N_0 = \left[\begin{array}{c|cc} 0 & 0 & 0 \\ \hline 0 & 0 & -(\tilde{a},\emptyset) \\ 0 & 1 & -(\tilde{a},x_0) \end{array}\right], \quad N_1 = \left[\begin{array}{c|cc} 0 & 0 & 0 \\ \hline 1 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

$$\lambda = \left[\begin{array}{c|cc} 0 & (c,x_1) & (c,x_0x_1) \end{array}\right], \quad \gamma = \left[\begin{array}{c} 1 \\ \hline z_{01} \\ z_{02} \end{array}\right].$$

Problem 4.4.5 Consider the linear representation

$$N_0 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \ N_1 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \ \lambda = \begin{bmatrix} 1 & 0 \end{bmatrix}, \ \gamma = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

(a) Compute the corresponding rational series c.
(b) Verify that c is rational without using the fact that it has a linear representation.
(c) Determine whether the given linear representation of c is minimal.

Remark: See Problem 1.4.2.

Problem 4.4.6 Give a specific example of a rational series $c \in \mathbb{R}\langle\langle X \rangle\rangle$ and two corresponding representations, one which is minimal and the other which is not.

Problem 4.4.7 Provide an example, if possible, for each scenario described below. If no such example exists, give a justification.

(a) A series c which is linear but not rational.
(b) A series d which is rational but not linear.
(c) A series e which is both rational and linear.
(d) A series f which is rational but not globally convergent.
(e) A series g which is globally convergent but not rational.

Section 4.5

Problem 4.5.1 Show that the shuffle inverse as defined in Problem 2.4.11 is not a rational operation.
Remark: See Problem 3.5.4.

Problem 4.5.2 Let $X = \{x_0, x_1\}$. Verify the following identities:

(a) $(x_i^*)^{\sqcup\!\sqcup\, k} = (kx_i)^*, \ \ k \geq 1$
(b) $x_0^* \sqcup\!\sqcup\, x_1^* = \text{char}(X^*)$
(c) $(x_0x_1)^* \sqcup\!\sqcup\, (-x_0x_1)^* = (-4x_0^2x_1^2)^*$
(d) $x_0(x_1^* \sqcup\!\sqcup\, (x_0(x_1^* \sqcup\!\sqcup\, 1))) = x_0x_1^*x_0(2x_1)^*$
(e) $x_0(x_1^* \sqcup\!\sqcup\, (x_0(x_1^* \sqcup\!\sqcup \cdots (x_0(x_1^* \sqcup\!\sqcup\, 1)) \cdots)))$
$$= x_0x_1^*x_0(2x_1)^* \cdots x_0(kx_1)^*.$$

The series x_1^* appears k times on the left-hand side of the last equation.

Problem 4.5.3 Consider a linear series $c = x_0^*x_1$.

(a) Verify that $c \sqcup\!\sqcup\, c \neq c^2$.
(b) Show that $\mathrm{supp}(c \sqcup\!\sqcup\, c) = \mathrm{supp}(c^2)$.

Problem 4.5.4 Provide an example of two rational series c and d where c is not input-limited, but $c \circ d$ is still rational.
Remark: See Lemma 2.5.

Problem 4.5.5 Let $X = \{x_0, x_1\}$. Show that the case where $c = (1 - x_1)^{-1} = x_1^*$ and $d = x_1$ provides an example where the *right* argument of the composition product is input-limited but rationality is not preserved.
Remark: See Problem 3.6.4.

Problem 4.5.6 In this problem, the analysis of the series $c \circ c$, where $c = (1 - x_1)^{-1} = (x_1)^*$, presented in Example 4.12 is expanded.

(a) Establish the identity

$$(c \circ c, x_0^{k_0} x_1^{k_1} \cdots x_0^{k_{l-1}} x_1^{k_l}) = (k_0)^{k_1}(k_0 + k_2)^{k_3} \cdots (k_0 + k_2 + \cdots + k_{l-1})^{k_l} \tag{4.14}$$

for all odd $l \geq 1$ and $k_i \geq 0$, $i = 0, 1, \ldots, l$. (Assume $0^0 := 1$.)

(b) Using the solution from part (a) verify that:

$$(c \circ c, x_0^{n_0} x_1 x_0^{n_1} x_1 \cdots x_0^{n_{j-1}} x_1 x_0^{n_j}) = n_0(n_0 + n_1) \cdots (n_0 + n_1 + \cdots + n_{j-1})$$

$$(c \circ c, x_1^{m_0} x_0 x_1^{m_1} \cdots x_0 x_1^{m_k}) = 0^{m_0} 1^{m_1} 2^{m_2} \cdots k^{m_k}$$

for all $j \geq 0$ and $n_i \geq 0$, $i = 0, 1, \ldots, j$; and all $k \geq 0$ and $m_i \geq 0$, $i = 0, 1, \ldots, k$.

(c) Using a result from part (b), show that

$$c \circ c = 1 + \sum_{k=1}^{\infty} x_0 x_1^* x_0 (2x_1)^* \cdots x_0 (kx_1)^*.$$

(d) Demonstrate that the identity above can also be written in the form

$$c \circ c = \sum_{k=0}^{\infty} \Psi_{x_1^k}(\mathbf{1}), \tag{4.15}$$

where the operator $\Psi_{x_1^k} : \mathbb{R}\langle\langle X \rangle\rangle \to \mathbb{R}\langle\langle X \rangle\rangle$ is defined by applying the following operator k times

$$\Psi_{x_1}(e) = x_0(x_1^* \sqcup\!\sqcup\, e)$$

when $k \geq 1$ and $\Psi_\emptyset(\mathbf{1}) = \mathbf{1}$.

Problem 4.5.7 Suppose c and d are two globally convergent series, not necessarily rational, where c is input-limited. Is this latter condition sufficient to guarantee that $c \circ d$ is also globally convergent? Provide either a proof or a counterexample.

Problem 4.5.8 For any $c, d \in \mathbb{R}\langle\langle X\rangle\rangle$ define the *Hadamard product* to be

$$c \odot d = \sum_{\eta \in X^*} (c, \eta)(d, \eta)\eta.$$

(a) Show that the Hadamard product preserves rationality.
(b) Show that the projection operator $P_0 : \mathbb{R}\langle\langle X\rangle\rangle \to \mathbb{R}[[X_0]] : c \mapsto c_0 = c \odot x_0^*$ preserves rationality.
(c) Prove that $c@d$ is rational only if $(c@d)_0 := P_0(c@d)$ is rational. Specifically, there must exist a linear representation $(\lambda_0, N_0, \gamma_0)$ such that $(c@d)_0 = \sum_{k \geq 0} \lambda_0 N_0^k \gamma_0$, or equivalently, $F_{c@d}[0](t) = \lambda_0 \exp(N_0 t)\gamma_0$, $t \geq t_0$.

Bibliographic Notes

Section 4.1 The theory of rational and recognizable formal power series was initiated by Schützenberger in the context of automata theory in [174, 175, 176]. His treatment was largely inspired by the work of Kleene in [131], who provided the central algebraic framework for automata theory. Lemma 4.1, which is used in Section 4.3 to prove Schützenberger's Theorem, appeared in the book by Conway on Kleene algebras [46]. A general introduction to rational series is the concise monograph by Berstel and Reutenauer [8] and the book by Salomaa and Soittola [169].

Section 4.2 There are several equivalent ways to define a linear representation of a formal power series. The original approach of Schützenberger in [174] involves the matrix trace operator. The one employed in this chapter is due to Fliess [61, 69]. System theorists will likely prefer this approach as it is closer in flavor to concepts from linear system theory like Markov parameters [125]. Theorem 4.1 appears in [95]. Theorem 4.2 is given in the Ph.D. dissertation of Jacob [115].

Section 4.3 Schützenberger's Theorem (Theorem 4.3) first appeared in [174]. The proof provided here follows that given in [8], which in part comes from [46] (i.e, Lemma 4.1).

Section 4.4

The Hankel rank condition for rationality/recognizability given in Theorem 4.4 is due to Fliess [61] and Isidori [112]. See also Chapter 3 of Isidori's book [113] for another treatment of the subject. The proof given here was adapted from that given in [169]. The description of the canonical factorization of the Hankel matrix follows the approach in [113]. Minimality of linear representations of a generating series is equivalent to the notion of minimality for bilinear state space realizations whose corresponding Chen-Fliess series share such generating series (see Chapter 6). The latter topic has been pursued by a number of different authors. For a treatment based solely on formal power series methods, see [169]. For analyzes in the context of bilinear realization theory, see [19, 24, 48, 61, 112, 113, 114, 164, 183]. An interesting combination of both approaches appears in [77]. Theorem 4.5 establishing the minimal dimension of a linear representation is largely based on [113], as is Theorem 4.6 relating minimal linear representations by a change of coordinates.

Section 4.5

Theorems 4.7 and 4.8 concerning the rationality of the shuffle product appear in [69]. The proof given for Theorem 4.7 was suggested in [8]. The counterexample showing that the composition product is not rational is due to Ferfera [58, 59]. The exact calculations for this example along with other related identities appeared in [90]. Theorem 4.9 introducing the input-limited condition is also the work of Ferfera in [58, 59], but the proof given here using Lemma 4.2 is quite different and can be found in [90]. Example 4.14 and Problem 4.5.8 addressing the rationality of the feedback connection both appeared in [82]. (The rationality of the Hadamard product is due to Fliess [69].)

References

1. E. Abe, *Hopf Algebras*, Cambridge University Press, Cambridge, United Kingdom, 1980.
2. L. V. Ahlfors, *Complex Analysis: An Introduction to the Theory of Analytic Functions of One Complex Variable, Third Edition*, McGraw-Hill, New York, 1979.
3. W. Balser, *From Divergent Power Series to Analytic Functions*, Springer-Verlag, Berlin, 1994.
4. R. G. Bartle, *The Elements of Real Analysis*, John Wiley & Sons, Inc., New York, 1976.
5. F. Baudoin, *An Introduction to the Geometry Stochastic Flows*, Imperial College Press, London, 2004.
6. G. Baumslag, *Topics in Combinatorial Group Theory*, Birkhäuser, Basel, 1993.
7. J. Berstel, *Transductions and Context-Free Languages*, Teubner, Stuttgart, 1979.
8. J. Berstel and C. Reutenauer, *Rational Series and Their Languages*, Springer-Verlag, Berlin, 1988.
9. ———, *Noncommutative Rational Series with Applications*, Cambridge University Press, Cambridge, United Kingdom, 2010.
10. W. M. Boothy, *An Introduction to Differentiable Manifolds and Riemannian Geometry*, Academic Press, Orlando, Florida, 1975.
11. É. Borel, *Leçons sur les séries divergentes, deuxième édition*, Gauthier-Villars, Paris, 1928. (republished by Jacques Gabay, Sceaux, 1988).
12. N. Bourbaki, *Elements of Mathematics, Lie Groups and Lie Alegbras, Chapters 1-3*, Springer-Verlag, Berlin, 1989.
13. ———, *Elements of Mathematics, Lie Groups and Lie Algebras, Chapters 7-9*, Springer-Verlag, Berlin, 2005.
14. S. Boyd, L. O. Chua, and C. A. Desoer, Analytical foundations of Volterra series, *IMA Journal of Mathematical Control and Information*, 1 (1984), 243–282.
15. J. W. Brewer, Kronecker products and matrix calculus in system theory, *IEEE Transactions on Circuits and Systems*, CAS-25 (1978), 772–781.
16. ———, Matrix calculus and the sensitivity analysis of linear dynamic systems, *IEEE Transactions on Automatic Control*, AC-23 (1978), 748–751.
17. ———, Correction to 'Kronecker products and matrix calculus in system theory', *IEEE Transactions on Circuits and Systems*, CAS-26 (1979), 360.
18. ———, Derivatives of the characteristic polynomial, trace, and determinant with respect to a matrix, *IEEE Transactions on Automatic Control*, AC-24 (1979), 787–790.
19. R. W. Brockett, On the algebraic structure of bilinear systems, in *Theory and Applications of Variable Structure Systems,* R. Mohler and R. Ruberti, Eds., Academic Press, New York, 1972, pp. 153–168.

20. R. W. Brockett, The Lie groups of simple feedback systems, in *Proceedings of the 1976 IEEE Conference on Decision and Control including the 15th Symposium on Adaptive Processes*, Clearwater, Florida, 1976, pp. 1189–1193.
21. T. J. I. Bromwich, *An Introduction to the Theory of Infinite Series, Second Edition*, Macmillan & Company, Ltd., London, 1965. zzz: Needs a custom edit: T. J. I'A.
22. C. Brouder, A. Frabetti, and C. Krattenthaler, Non-commutative Hopf algebra of formal diffeomorphisms, *Advances in Mathematics*, 200 (2006), 479–524.
23. J. W. Brown and R. V. Churchill, *Complex Variables and Applications, Ninth Edition*, McGraw-Hill Education, New York, 2014.
24. C. Bruni, G. DiPillo, and G. Koch, Bilinear systems: An appealing class of "nearly linear" systems in theory and applications, *IEEE Transactions on Automatic Control*, AC-19 (1974), 334–348.
25. F. M. Callier and C. A. Desoer, *Linear System Theory*, Springer-Verlag, New York, 1991.
26. E. Cartan, Sur la structure des groupes infinis de transformations, *Annales Scientifiques de l'École Normale Supérieure*, 21-22 (1904-1905), 153–206, 219–308.
27. P. Cartier, A primer of Hopf algebras, Tech. Rep. IHES/M/06/40, Institut des Hautes Études Scientifiques, Bures-sur-Yvette, France, September 2006.
28. C. G. Cassandras and S. Lafortune, *Introduction to Discrete Event Systems*, Springer-Verlag, New York, 1999.
29. D. P. L. Castrigiano and S. A. Hayes, *Catastrophe Theory*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1993.
30. C.-T. Chen, *Linear System Theory and Design*, Holt, Rinehart and Winston, New York, 1984.
31. K.-T. Chen, Commutator calculus and link invariants, *Proceedings of the American Mathematical Society*, 3 (1952), 44–55. Errata: 3 (1952) 993.
32. ———, Iterated integrals and exponential homomorphisms, *Proceedings of the London Mathematical Society*, 4 (1954), 502–512.
33. ———, Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula, *Annals of Mathematics, Second Series*, 65 (1957), 163–178.
34. ———, On the composition functions of nilpotent Lie groups, *Proceedings of the American Mathematical Society*, 8 (1957), 1158–1159.
35. ———, Integration of paths - a faithful representation of paths by noncommutative formal power series, *Transactions of the American Mathematical Society*, 89 (1958), 395–407.
36. ———, Algebraization of iterated integration along paths, *Bulletin of the American Mathematical Society*, 73 (1967), 975–978.
37. ———, Iterated path integrals and generalized paths, *Bulletin of the American Mathematical Society*, 73 (1967), 935–938.
38. ———, Algebraic paths, *Journal of Algebra*, 10 (1968), 8–36.
39. ———, Algebras of iterated integrals and fundamentals groups, *Transactions of the American Mathematical Society*, 156 (1971), 359–379.
40. ———, Differential forms and homotopy groups, *Journal of Differential Geometry*, 6 (1971-72), 231–246.
41. ———, Iterated integrals of differential forms and loop space homology, *Annals of Mathematics, Second Series*, 97 (1973), 217–246.
42. ———, Iterated path integrals, *Bulletin of the American Mathematical Society*, 83 (1977), 831–879.
43. ———, Pullback path fibrations, homotopies and iterated integrals, *Bulletin of the Institute of Mathematics, Academi Sinica*, 8 (1980), 263–275.

44. K.-T. Chen, R. H. Fox, and R. C. Lyndon, Free differential calculus, IV. the quotient groups of the lower central series, *The Annals of Mathematics, Second Series*, 68 (1958), 81–95.
45. P. M. Cohn, Sur le critère de Friedrichs pour les commutateurs dans une algèbre associative libre, *Comptes Rendus de l'Académie des Sciences de Paris*, 239 (1954), 743–745.
46. J. H. Conway, *Regular Algebra and Finite Machines*, Chapman and Hall, Ltd., London, 1971.
47. P. E. Crouch, Dynamical realizations of finite Volterra series, *SIAM Journal on Control and Optimization*, 19 (1981), 177–202.
48. P. D'Alessandro, A. Isidori, and A. Ruberti, Realization and structure theory of bilinear dynamical systems, *SIAM Journal on Control*, 12 (1974), 517–535.
49. J. Devlin, Word problems related to periodic solutions of a non-autonomous system, *Mathematical Proceedings of the Cambridge Philosophical Society*, 108 (1990), 127–151.
50. G. Doetsch, *Handbuch der Laplace–Transformation*, vol. II, Birkhäuser, Basel, 1956.
51. S. Dăscălescu, C. Năstăsescu, and Ş. Raianu, *Hopf Algebras - An Introduction*, Marcel Dekker, Inc., New York, 2001.
52. L. A. Duffaut Espinosa, *Interconnections of Nonlinear Systems Driven by L_2-Itô Stochastic Processes*, Ph.D. dissertation, Old Dominion University, 2009.
53. L. A. Duffaut Espinosa, K. Ebrahimi-Fard, and W. S. Gray, A combinatorial Hopf algebra for nonlinear output feedback control systems, *Journal of Algebra*, 453 (2016), 609–643.
54. L. A. Duffaut Espinosa, W. S. Gray, and O. R. González, On Fliess operators driven by L_2-Itô random processes, in *Proceedings of the 48th IEEE Conference on Decision and Control*, Shanghai, China, 2009, pp. 7478–7484.
55. E. B. Dynkin, Calculation of the coefficients in the Campbell-Hausdorff formula, *Doklady Akademii Nauk SSSR (N.S.)*, 57 (1947), 323–326. English translation by A. Kleshchev (1997).
56. K. Ebrahimi-Fard and W. S. Gray, Center problem, Abel equation and the Faà di Bruno Hopf algebra for output feedback, *International Mathematics Research Notices*, 2017 (2017), 5415–5450.
57. ———, The Faà di Bruno Hopf algebra for multivariable feedback recursions in the center problem for higher order Abel equations, in *Computation and Combinatorics in Dynamics, Stochastics and Control, The Abel Symposium, Rosendal, Norway, August 2016,* E. Celledoni, G. Di Nunno, K. Ebrahimi-Fard, and H. Z. Munthe-Kaas, Eds., vol. 13, Springer Nature Switzerland AG, Cham, Switzerland, 2018, pp. 265–296.
58. A. Ferfera, *Combinatoire du Monoïde Libre Appliquée à la Composition et aux Variations de Certaines Fonctionnelles Issues de la Théorie des Systèmes*, Ph.D. dissertation, University of Bordeaux I, 1979.
59. ———, Combinatoire du monoïde libre et composition de certains systèmes non linéaires, *Astérisque*, 75-76 (1980), 87–93.
60. H. Figueroa and J. M. Gracia-Bondía, Combinatorial Hopf algebras in quantum field theory I, *Reviews in Mathematical Physics*, 17 (2005), 1–96.
61. M. Fliess, Matrices de Hankel, *Journal de Mathématiques Pures et Appliquées*, 53 (1974), 197–224. Erratum: 54 (1975) 481.
62. ———, Sur divers produits de séries formelles, *Bulletin de la Société Mathématique de France*, 102 (1974), 181–191.
63. ———, Séries de Volterra et séries formelles non commutatives, *Comptes rendus de l'Académie des Sciences, Série A*, A-280 (1975), 965–967.

64. ———, Un outil algébrique: Les séries formelles non commutatives, in *Mathematical Systems Theory,* G. Marchesini and S. K. Mitter, Eds., Lecture Notes in Economics and Mathematical Systems, Vol. 131, Springer-Verlag, Berlin, 1976, pp. 122–148.
65. ———, Nonlinear realization theory and abstract transitive Lie algebra, *Bulletin of the American Mathematical Society*, 2 (1980), 444–446.
66. ———, A note on Volterra series expansions for nonlinear differential systems, *IEEE Transactions on Automatic Control*, AC-25 (1980), 116–117.
67. ———, Realization of non-linear systems via E. Cartan's third fundamental theorem, in *Proceedings of the 19th IEEE Conference on Decision and Control*, Albuquerque, New Mexico, 1980, pp. 922–925.
68. ———, An algebraic approach to functional expansions, application to a singular optimal control problem, in *Control Science and Technology for the Progress of Society, Proceedings of the 8th IFAC World Congress*, Kyoto, Japan, 1981, pp. 331–336.
69. ———, Fonctionnelles causales non linéaires et indéterminées non commutatives, *Bulletin de la Société Mathématique de France*, 109 (1981), 3–40.
70. ———, On the concept of derivatives and Taylor expansions for nonlinear input-output systems, in *Proceedings of the 22nd IEEE Conference on Decision and Control*, San Antonio, Texas, 1983, pp. 643–646.
71. ———, On the relationship between the local realization of nonlinear systems, filtered, transitive Lie algebras and noncommutative indeterminates, in *Differential Geometric Control Theory,* R. W. Brockett, R. S. Millman, and H. J. Sussmann, Eds., Birkhäuser, Boston, 1983, pp. 226–236.
72. ———, Réalisation locale des systèmes non linéaires, algèbres de Lie filtrées transitives et séries génératrices non commutatives, *Inventiones Mathematicae*, 71 (1983), 521–537.
73. ———, Vers une notion de dérivation fonctionnelle causale, *Annales de l'Institut Henri Poincaré, Analyse Non Linéaire*, 3 (1986), 67–76.
74. M. Fliess, M. Lamnabhi, and F. Lamnabhi-Lagarrigue, An algebraic approach to nonlinear functional expansions, *IEEE Transactions on Circuits and Systems*, CAS-30 (1983), 554–570.
75. L. Foissy, Faà di Bruno subalgebras of the Hopf algebra of planar trees from combinatorial Dyson-Schwinger equations, *Advances in Mathematics*, 218 (2008), 136–162.
76. L. Foissy, The Hopf algebra of Fliess operators and its dual pre-Lie algebra, *Communications in Algebra*, 43 (2015), 4528–4552.
77. E. Fornasini and G. Marchesini, Algebraic realization theory of bilinear discrete-time input-output maps, *Journal of the Franklin Institute*, 301 (1976), 143–159.
78. M. V. Foursov and C. Hespel, Weighted Petri nets and polynomial dynamical systems, in *Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems*, Kyoto, Japan, 2006, pp. 1539–1546.
79. K. O. Friedrichs, Mathematical aspects of the quantum theory of fields. part V. Fields modified by linear homogeneous forces, *Communications in Pure and Applied Mathematics*, 6 (1953), 1–72.
80. E. Gehrig, *Hopf Algebras, Projections, and Coordinates of the First Kind in Control Theory*, Ph.D. dissertation, Arizona State University, 2007.
81. W. S. Gray, A unified approach to generating series for nonlinear cascade systems, in *Proc. 48^{th} IEEE Conference on Decision and Control*, Shanghai, China, 2009, pp. 7478–7484.

82. ———, On the rationality of the feedback connection, in *Proceedings of the 19th International Symposium on the Mathematical Theory of Networks and Systems*, Budapest, Hungary, 2010, pp. 1035–1040.
83. W. S. Gray and L. A. Duffaut Espinosa, A Faà di Bruno Hopf algebra for a group of Fliess operators with applications to feedback, *Systems and Control Letters*, 60 (2011), 441–449.
84. W. S. Gray and L. A. Duffaut Espinosa, Feedback transformation group for nonlinear input-output systems, in *Proceedings of the 52nd IEEE Conference on Decision and Control*, Florence, Italy, 2013, pp. 2570–2575.
85. ———, Dynamic output feedback invariants of full relative degree nonlinear SISO systems, in *Proceedings of the 23rd International Symposium on the Mathematical Theory of Networks and Systems*, Hong Kong, 2018, pp. 762–768.
86. W. S. Gray, L. A. Duffaut Espinosa, and K. Ebrahimi-Fard, Faà di Bruno Hopf algebra of the output feedback group for multivariable Fliess operators, *Systems & Control Letters*, 74 (2014), 64–73.
87. ———, Recursive algorithm for the antipode in the SISO feedback product, in *Proceedings of the 21st International Symposium on the Mathematical Theory of Networks and Systems*, Groningen, The Netherlands, 2014, pp. 1088–1093.
88. W. S. Gray, L. A. Duffaut Espinosa, and M. Thitsa, Left inversion of analytic nonlinear SISO systems via formal power series methods, *Automatica*, 50 (2014), 2381–2388.
89. W. S. Gray and K. Ebrahimi-Fard, SISO output affine feedback transformation group and its Faà di Bruno Hopf algebra, *SIAM Journal on Control and Optimization*, 55 (2017), 885–912.
90. W. S. Gray, H. Herencia-Zapana, L. A. Duffaut Espinosa, and O. R. González, Bilinear systems interconnections and generating series of weighted Petri nets, *Systems and Control Letters*, 58 (2009), 841–848.
91. W. S. Gray and Y. Li, Generating series for interconnected analytic nonlinear systems, *SIAM Journal on Control and Optimization*, 44 (2005), 646–572.
92. W. S. Gray, M. Palmstrøm, and A. Schmeding, Continuity of formal power series products in nonlinear control theory, https://arxiv.org/abs/2102.05732, 2021.
93. W. S. Gray and J. M. A. Scherpen, Hankel operators and Gramians for nonlinear systems, in *Proceedings of the 37th IEEE Conference on Decision and Control*, Tampa, Florida, 1998, pp. 1416–1421.
94. W. S. Gray and M. Thitsa, A unified approach to generating series for mixed cascades of analytic nonlinear input-output systems, *International Journal of Control*, 85 (2012), 1737–1754.
95. W. S. Gray and Y. Wang, Fliess operators on L_p spaces: Convergence and continuity, *Systems and Control Letters*, 46 (2002), 67–74.
96. ———, Formal Fliess operators with applications to feedback interconnections, in *Procceedings of the 18th International Symposium on the Mathematical Theory of Networks and Systems*, Blacksburg, Virginia, 2008.
97. D. Grinberg and V. Reiner, Hopf algebras in combinatorics, https://arxiv.org/abs/1409.8356, 2014.
98. W. Gröbner, *Die Lie-Reihen und ihre Anwendungen*, VEB Deutscher Verlag der Wissenschaften, Berlin, 1967.
99. M. Gross and A. Lentin, *Introduction to Formal Grammars*, Springer-Verlag, Berlin, 1970.
100. R. Grossman and R. G. Larson, The realization of input-output maps using bialgebras, *Forum Mathematicum*, 4 (1992), 109–121.

101. R. L. Grossman and R. G. Larson, Bialgebras and realizations, in *Hopf Algebras,* J. Bergen, S. Catoiu, and W. Chin, Eds., Marcel Dekker, New York, 2004, pp. 157–166.
102. L. Grunenfelder, Algebraic aspects of control systems and realizations, *Journal of Algebra*, 165 (1994), 446–464.
103. V. W. Guillemin and S. Sternberg, An algebraic model of transitive differential geometry, *Bulletin of the American Mathematical Association*, (1964), 16–47.
104. B. C. Hall, *Lie Groups, Lie Algebras, and Representations*, Springer-Verlag, New York, 2003.
105. P. Hall, *Theory of Groups, Second Edition*, Chelsea Publishing Company, New York, 1976.
106. M. A. Harrison, *Introduction to Formal Language Theory*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1978.
107. S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces*, Academic Press, Orlando, Florida, 1978.
108. V. Hoang Ngoc Minh, Evaluation transform, *Theoretical Computer Science*, 79 (1991), 163–177.
109. L. Hörmander, *An Introduction to Complex Analysis in Several Variables, Third Edition*, Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1990.
110. J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory*, Springer-Verlag, New York, 1972.
111. M. A. Hurwitz, Sur un théorème de M. Hadamard, *Comptes rendus de l'Académie des Sciences*, 128 (1899), 350–353.
112. A. Isidori, Direct construction of minimal bilinear realizations from nonlinear input-output maps, *IEEE Transactions on Automatic Control*, 18 (1973), 626–631.
113. ———, *Nonlinear Control Systems, Third Edition*, Springer-Verlag, London, 1995.
114. A. Isidori and A. Ruberti., Realization theory of bilinear systems, in *Geometric Methods in System Theory,* D. Q. Mayne and R. W. Brockett, Eds., Springer, Dordrecht, 1973, pp. 83–130.
115. M. G. Jacob, *Représentations et Substitutions Matricielles dans la Théorie Algébrique des Transductions*, Ph.D. dissertation, University of Paris VII, 1975.
116. N. Jacobson, *Lie Algebras*, Dover Publications, Inc., New York, 1962.
117. B. Jakubczyk, Existence and uniqueness of realizations of nonlinear systems, *SIAM Journal on Control and Optimization*, 18 (1980), 445–471.
118. ———, Construction of formal and analytic realizations of nonlinear systems, in *Feedback Control of Linear and Nonlinear Systems,* D. Hinrichsen and A. Isidori, Eds., Springer-Verlag, Berlin, 1982, pp. 147–156.
119. ———, Local realization of nonlinear causal operators, *SIAM Journal on Control and Optimization*, 24 (1986), 230–242.
120. ———, Realization theory for nonlinear systems: Three approaches, in *Algebraic and Geometric Methods in Nonlinear Control Theory,* M. Fliess and M. Hazewinkel, Eds., D. Reidel Publishing Company, Dordrecht, The Netherlands, 1986, pp. 3–31.
121. ———, Convergence of power series along vector fields and their commutators; a Cartan-Kähler type theorem, *Annales Polonici Mathematici*, 74 (2000), 117–132.
122. ———, Existence theorems in nonlinear realization theory and a Cauchy-Kowalewski type theorem, in *Contemporary Trends in Nonlinear Geometric Control Theory and Its Applications,* A. Anzaldo-Meneses, B. Bonnard, J.

P. Gauthier, and F. Monroy-Pérez, Eds., World Scientific, Singapore, 2002, pp. 347–357.
123. S. A. Joni and G. Rota, Coalgebras and bialgebras in combinatorics, *Studies in Applied Mathematics*, 61 (1979), 93–139.
124. ———, Coalgebras and bialgebras in combinatorics, *Contemporary Mathematica*, 6 (1982), 1–47.
125. T. Kailath, *Linear Systems*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1980.
126. R. Kalman, On partial realizations, transfer functions, and canoncial forms, *Acta polytechnica Scandinavica*, 31 (1979), 9–32.
127. R. E. Kalman, P. L. Falb, and M. A. Arbib, *Topics in Mathematical System Theory*, McGraw-Hill, Inc., New York, 1969.
128. M. Kawski, Calculating the logarithm of the Chen-Fliess series, in *Proceeedings of the 2000 International Symposium on Mathematical Theory of Networks and Systems*, Perpignan, France, 2000.
129. ———, Chronological algebras: Combinatorics and control, *Journal of Mathematical Sciences*, 103 (2001), 725–744.
130. H. K. Khalil, *Nonlinear Systems, Third Edition*, Prentice-Hall, Inc., Upper Saddle River, New Jersey, 2002.
131. S. C. Kleene, Representation of events in nerve nets and finite automata, in *Automata Studies,* C. E. Shannon and J. McCarthy, Eds., Princeton University Press, Princeton, New Jersey, 1956, pp. 3–41.
132. K. Knopp, *Infinite Sequences and Series*, Dover Publications, Inc., New York, 1956.
133. E. Kreyszik, *Introductory Functional Analysis with Applications*, John Wiley & Sons, New York, 1978.
134. W. Kuich and A. Salomaa, *Semirings, Automata, Languages*, Springer-Verlag, Berlin, 1986.
135. M. Lamnabhi, A new symbolic calculus for the response of nonlinear systems, *Systems and Control Letters*, 2 (1982), 154–162.
136. F. Lamnabhi-Lagarrigue, Volterra and Fliess series expansions for nonlinear systems, in *The Control Handbook,* William S. Levine, Ed., CRC Press, Boca Raton, Florida, 1996, pp. 879–888.
137. F. Lamnabhi-Lagarrigue and P. E. Crouch, A formula for iterated derivatives along trajectories of nonlinear systems, *Systems & Control Letters*, 11 (1988), 1–7.
138. M. Lazard, Sur les groupes nilpotents et les anneaux de Lie, *Annales Scientifiques de l'École Normale Supérieure, Troisième Série*, 71 (1954), 101–190.
139. P. Leroux and X. G. Viennot, A combinatorial approach to nonlinear functional expansions: an introduction with an example, *Theoretical Computer Science*, 79 (1991), 179–193.
140. Y. Li, *Generating Series of Interconnected Nonlinear Systems and the Formal Laplace–Borel Transform*, Ph.D. dissertation, Old Dominion University, 2004.
141. Y. Li and W. S. Gray, The formal Laplace–Borel transform of Fliess operators and the composition product, *International Journal of Mathematics and Mathematical Sciences*, 2006 (2006), 1–14.
142. M. Lothaire, *Combinatorics on Words*, Cambridge Univesity Press, Cambridge, United Kingdom, 1983.
143. ———, *Algebraic Combinatorics on Words*, Cambridge University Press, Cambridge, United Kingdom, 2002.
144. ———, *Applied Combinatorics on Words*, Cambridge University Press, Cambridge, United Kingdom, 2005.

145. R. C. Lyndon, On the Burnsides problem, *Transactions of the American Mathematical Society*, 77 (1954), 202–215.
146. R. C. Lyndon and P. E. Schupp, *Combinatorial Group Theory*, Springer-Verlag, Berlin, 1977.
147. S. MacLane and G. Birkhoff, *Algebra*, The MacMillan Company, New York, 1967.
148. W. Magnus, A. Karrass, and D. Solitar, *Combinatorial Group Theory*, Dover Publications, Inc., Mineola, New York, 1976.
149. A. I. Malcev, Nilpotent groups without torsion, *Izvestiya Akademii Nauk SSSR, Mathematics Series*, 13 (1949), 213–220. In Russian.
150. D. Manchon, Hopf algebras, from basics to applications to renormalization, http://arxiv:math/0408405, 2006.
151. J. W. Milnor and J. C. Moore, On the structure of Hopf algebras, *The Annals of Mathematics, Second Series*, 81 (1965), 211–264.
152. A. W. Naylor and G. R. Sell, *Linear Operator Theory in Engineering and Science*, Springer-Verlag, New York, 1982.
153. H. Nijmeijer and A. J. van der Schaft, *Nonlinear Dynamical Control Systems*, Springer-Verlag, New York, 1990.
154. J. L. Peterson, *Petri Net Theory Theory and the Modeling of Systems*, Prentice-Hall, Inc., Englewoods Cliffs, New Jersey, 1981.
155. S. Priess-Crampe and P.Ribenboim, Fixed point and attractor theorems for ultrametric spaces, *Forum Mathematicum.*, 12 (2000), 53–64.
156. R. Ree, Lie elements and an algebra associated with shuffles, *Annals of Mathematics, Second Series*, 68 (1958), 210–220.
157. C. Reutenauer, The local realization of generating series of finite Lie rank, in *Algebraic and Geometric Methods in Nonlinear Control Theory,* M. Fliess and M. Hazewinkel, Eds., D. Reidel Publishing Company, Dordrecht, The Netherlands, 1986, pp. 33–43.
158. ———, *Free Lie Algebras*, Oxford University Press, New York, 1993.
159. G. E. Rèvèsz, *Introduction to Formal Languages*, McGraw-Hill Book Company, New York, 1983.
160. E. Riccomagno, Calcolo simbolico per processi stocastici, Master's thesis, Università di Genova, 1993.
161. E. A. M. Rocha, *An Algebraic Approach to Nonlinear Control Theory*, Ph.D. dissertation, University of Aveiro, 2003.
162. E. M. Rocha, On computation of the logarithm of the Chen-Fliess series for nonlinear systems, in *Nonlinear and Adaptive Control,* Zinober, Alan and Owens, David, Eds., Springer-Verlag, Berlin, Heidelberg, 2003, pp. 317–326.
163. G. Rozenberg and E. A. Salomaa, *Handbook of Formal Languages, Volume 1: Word, Language, Grammar*, Springer-Verlag, Berlin, 1997.
164. A. Ruberti, A. Isidori, and P. D'Alessandro, *Theory of Bilinear Dynamical Systems*, Springer-Verlag Wien, New York, 1972.
165. W. Rudin, *Function Theory in the Unit Ball of* $\mathbb{C}^n$, Springer-Verlag, New York, 1980.
166. W. J. Rugh, *Nonlinear System Theory, The Volterra/Wiener Approach*, The Johns Hopkins University Press, Baltimore, 1981.
167. J. M. Ruiz, *The Basic Theory of Power Series*, Viewweg, Braunschweig/Wiesbaden, Germany, 1993.
168. A. Salomaa, *Formal Languages*, Academic Press, New York, 1973.
169. A. Salomaa and M. Soittola, *Automata-Theoretic Aspects of Formal Power Series*, Springer-Verlag, New York, 1978.

170. J. M. A. Scherpen and W. S. Gray, On singular value functions and Hankel operators for nonlinear systems, in *Proceedings of the 1999 American Control Conference*, San Diego, 1999, pp. 2360–2364.
171. ———, State dependent matrices and balanced energy functions for nonlinear systems, in *Proceedings of the 2000 International Symposium on Mathematical Theory of Networks and Systems*, Perpignan, France, 2000.
172. M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*, John Wiley, New York, 1980.
173. E. Schörner, Ultrametric fixed point theorems and applications, in *Valuation Theory and Its Applications, Vol. II,* F.-V. Kuhlmannm, S. Kuhlmann, and M. Marshall, Eds., American Mathematical Society, Providence, Rhode Island, 2003, pp. 353–359.
174. M. P. Schützenberger, On the definition of a family of automata, *Information and Control*, 4 (1961), 245–270.
175. ———, Certain elementary families of automata, in *Proceedings of the Symposium of Mathematical Theory of Automata*, New York, 1962, pp. 139–153.
176. ———, On a theorem of R. Jungen, *Proceedings of the American Mathematical Society*, 13 (1962), 885–890.
177. B. Shawyer and B. Watson, *Borel's Methods of Summability*, Clarendon Press, Oxford, United Kingdom, 1994.
178. R. A. Silverman, *Introductory Complex Analysis*, Dover Publications, Inc., New York, 1972.
179. I. M. Singer and S. Sternberg, The infinite groups of Lie and Cartan. I. the transitive groups, *Journal d'Analyse Mathématique*, 15 (1965), 1–114.
180. E. D. Sontag, Spaces of observables in nonlinear control, in *International Congress of Mathematicians, 1994*, Birkhäuser Verlag, Basel, 1995, pp. 1532–1545.
181. ———, *Mathematical Control Theory: Deterministic Finite Dimensional Systems, Second Edition*, Springer-Verlag, New York, 1998.
182. E. D. Sontag and Y. Wang, Input/output equations and realizability, in *Realization and Modelling in System Theory,* M. A. Kaashoek, J. H. van Schuppen, and A. C. M. Ran, Eds., Birkhäuser Boston, Boston, Massachusetts, 1990, pp. 125–132.
183. H. Sussmann, Minimal realizations and canonical forms for bilinear systems, *Journal of the Franklin Institute*, 301 (1976), 593–604.
184. H. J. Sussmann, Existence and uniqueness of minimal realizations of nonlinear systems, *Mathematical Systems Theory*, 10 (1977), 263–284.
185. ———, Lie brackets and local controllability: A sufficient condition for scalar-input systems, *SIAM Journal on Control and Optimization*, 21 (1983), 686–713.
186. ———, A product expansion for the Chen series, in *Theory and Applications of Nonlinear Control Systems,* C. I. Byrnes and A. Lindquist, Eds., Elsevier Science Publishers B.V. (North-Holland), New York, 1986, pp. 323–335.
187. ———, A proof of the realization theorem for convergent generating series of finite Lie rank, internal report SYCON-90-02, Rutgers Center for Systems and Control, 1990.
188. M. E. Sweedler, *Hopf Algebras*, W. A. Benjamin, Inc., New York, 1969.
189. M. Thitsa, *On the Radius of Convergence of Interconnected Analytic Nonlinear Systems*, Ph.D. dissertation, Old Dominion University, 2011.
190. M. Thitsa and W. S. Gray, On the radius of convergence of interconnected analytic nonlinear systems, *SIAM Journal on Control and Optimization*, 50 (2012), 2786–2813.

191. E. C. Titchmarsh, *The Theory of Functions, Second Edition*, Oxford University Press, London, 1939.
192. P. Tondeur, *Collected Papers of K.-T. Chen*, Birkhäuser Boston, Boston, Massachusetts, 2001.
193. E. U. Heckmanns, *Aspects of Ultrametric Spaces*, Queen's University, Kingston, Ontario, Canada, 1998.
194. V. S. Varadarajan, *Lie Groups, Lie Algebras, and Their Representations*, Springer-Verlag, New York, 1984.
195. G. S. Venkatesh, *Wiener-Fliess Composition of Formal Power Series: Additive Static Feedback and Shuffle Rational Series*, Ph.D. dissertation, Old Dominion University, 2021.
196. M. Vidyasagar, *Nonlinear Systems Analysis, Second Edition*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1993.
197. V. Volterra, Sopra le funzioni che dipendono da altre funzioni, *Rendiconti della R. Accademia dei Lincei, Secondo Semestre*, 3 (1887), 97–105, 141–146, 153–158.
198. ———, *Theory of Functionals and of Integral and Integro-Differential Equations*, Dover Publications, Inc., New York, 1959.
199. Y. Wang, *Differential Equations and Nonlinear Control Systems*, Ph.D. dissertation, Rutgers University, 1990.
200. Y. Wang and E. D. Sontag, On two definitions of observation spaces, *Systems and Control Letters*, 13 (1989), 279–289.
201. ———, Algebraic differential equations and rational control systems, *SIAM Journal on Control and Optimization*, 30 (1992), 1126–1149.
202. ———, Generating series and nonlinear systems: analytic aspects, local realizability and I/O representations, *Forum Mathematicum*, 4 (1992), 299–322.
203. ———, Orders of input/output differential equations and state-space dimensions, *SIAM Journal on Control and Optimization*, 33 (1995), 1102–1126.
204. H. S. Wilf, *Generatingfunctionology, Second Edition*, Academic Press, San Diego, California, 1994.
205. I. M. Winter Arboleda, *On Analytic Nonlinear Input-output Systems: Expanded Global Convergence and System Interconnections*, Ph.D. dissertation, Old Dominion University, 2019.
206. I. M. Winter Arboleda, W. S. Gray, and L. A. Duffaut Espinosa, Fractional Fliess operators: Two approaches, in *Proceedings of the 49th Conference on Information Sciences and Systems*, Baltimore, Maryland, 2015.
207. W. Wong, Volterra series, universal bilinear systems, and Fock representations, in *Proceedings of the 1979 Conference on Information Sciences and Systems*, Baltimore, Maryland, 1979, pp. 207–213.