Entropy of Generating Series for Nonlinear Input-Output Systems and Their Interconnections

W. Steven GRAY

Department of Electrical and Computer Engineering, Old Dominion University,
Norfolk, Virginia 23529, USA
E-mail: *sgray@odu.edu*
URL: http://www.ece.odu.edu/~sgray/

Abstract. This paper has two main objectives. The first is to introduce a notion of entropy that is well suited for the analysis of nonlinear input-output systems that have a Chen–Fliess series representation. The latter is defined in terms of its generating series over a noncommutative alphabet. The idea is to assign an entropy to a generating series as an element of a graded vector space. The second objective is to describe the entropy of generating series originating from interconnected systems of Chen–Fliess series that arise in the context of control theory. It is shown that one set of interconnections can never increase entropy as defined here, while a second set has the potential to do so. The paper concludes with a brief introduction to an entropy ultrametric space and some open questions.

Key words: Chen–Fliess series; formal power series; entropy; nonlinear control theory

2020 Mathematics Subject Classification: 68R15; 94A17; 93C10; 16T30

1 Introduction

The concept of *entropy* in the context of dynamical systems has been defined in a variety of ways [7, 56]. Each notion is viewed as a measurement of the rate of increase in dynamical complexity as the system evolves over time. The idea has been adapted and applied in control theory starting with the work of Zames in the 1970s [57] in order to address engineering problems like controlled invariance, estimation, model detection, and stabilization in a finite precision environment due to sampling, quantization, finite bit rates, etc. (see [36] for a more complete overview). More recently, the concept of entropy has been employed in the analysis of networked control systems [32, 39, 45, 51]. The general goal there is to quantify the entropy of a network in terms of the entropies of its subsystems in order to establish some fundamental performance bounds given limited communication between the subsystems. In [36], for example, an explicit bound on the topological entropy of a network of autonomous, continuous-time, nonlinear state space systems is derived in terms of similar upper bounds on its subsystems. The main result is then applied to cascaded systems to produce an upper bound on the entropy of the composite system.

This paper has two main objectives. The first is to introduce a notion of entropy that is well suited for the analysis of input-output systems that have a Chen–Fliess series representation [16, 17]. The latter is defined in terms of its generating series, i.e., a noncommutative formal power series over a finite alphabet. The idea is to assign an entropy to a formal power series as an element of a graded vector space which measures how quickly in an asymptotic sense the homogeneous components of the grading are being occupied by the support of the series. The asymptotic behavior of the Hilbert–Poincaré series for the vector space provides a direct

basis for comparison [1, 2]. This idea is closely related to the concept of entropy appearing in formal language theory using what is called a *generating structure function* [33, 34]. There the grading is always based on word length, and the two notions of entropy coincide if the formal power series is taken to be the characteristic series of the language. As the authors of these works explain, their definition is in turn related to the classical information theoretic notion of entropy introduced by Shannon in terms of channel capacity [47]. Other distinct definitions for the entropy of a language exist. For example, the concept introduced in [46] measures the complexity of a language in terms of the exponential growth rate in the number of equivalence classes of the Nerode congruence relation of the language as the word length increases. In this setting, all rational languages and Dyck languages have zero entropy, which is not the case here or in earlier work. The entropy of a formal power series as defined here is also related in spirit to the concept of entropy appearing in symbolic dynamics [37], namely, the entropy defined for a semigroup under concatenation corresponding to the language of forbidden words in a shift space. It is, however, distinct from this established definition in that there is no underlying set of recurrence inequalities to provide additional structure (see, for example, [37, Lemma 4.1.7]). The proposed entropy concept also coincides in certain special cases (modulo a logarithm) with the notion of entropy defined for graded algebras [40, 41]. But that connection does not appear to be so important for the present work. Finally, it should be stated that the entropy of a generating series can be used to develop notions of entropy for its corresponding input-output map. One such example is *system identification entropy* as defined in [20]. This quantity describes the growth in the number of bits needed to specify the series coefficients in order to approximate the input-output map as the desired accuracy increases.

The second objective of the paper is to describe the entropy of generating series originating from interconnected systems of Chen–Fliess series found in the context of control theory. Such interconnections have been studied extensively by the author and others [12, 13, 16, 18, 21, 22, 23, 24, 25, 26, 27, 29, 50, 52, 53, 55]. It is known that in every case the composite system always has a Chen–Fliess series representation whose generating series can be computed in terms of an induced formal power series product applied to the generating series of the component systems. The main products of interest are addition, the shuffle product, the composition product and the feedback product. The Cauchy product and Hadamard product play a supporting role in this setting. The main question to be addressed here is which system interconnections/formal power series products are capable of increasing entropy and which ones are not. It will turn out that all linear systems have generating series with *zero* entropy, so any interconnection that preserves linearity will not increase entropy. Somewhat surprising, even interconnections like the parallel product of two linear systems, which do not preserve linearity, will still result in a new system having zero entropy. Thus, this new concept of entropy is nontrivial only for nonlinear systems. Finally, this work in many ways is the *complement* of previous work by the author and others regarding the convergence of Chen–Fliess series and their interconnections [25, 26, 28, 50, 52, 53, 54, 55]. The main focus there was on the growth rate of the coefficients of a generating series and the specific types of convergence that are guaranteed for the corresponding Chen–Fliess series. In this paper, the nature of the coefficients is almost entirely irrelevant aside from being either zero or nonzero, i.e., being in the support of the generating series or not. Nevertheless, as shown in [20], the entropy of a generating series can be used to refine the notion of radius of convergence of a Chen–Fliess series as defined in [50].

The paper concludes with a brief introduction to an entropy ultrametric space. Such constructions have appeared in other contexts, for example, [4]. As feedback systems are often analyzed in an ultrametric space setting [12, 21, 25, 29], this framework could have immediate applications in control theory. Some current open problems are posed in this setting.

The paper is organized as follows. In the next section, a few preliminaries regarding formal power series, Chen–Fliess series, and system interconnections are summarized. In Section 3, the

definition of entropy of a generating series is given along with a number of illustrative examples. The entropy of interconnected systems is characterized in Section 4. The entropy ultrametric is introduced in the subsequent section. The conclusions of the paper are summarized in the final section.

2 Preliminaries

An *alphabet* $X = \{x_0, x_1, \ldots, x_m\}$ is any nonempty and finite set of symbols referred to as *letters*. A *word* $\eta = x_{i_1} \cdots x_{i_k}$ is a finite sequence of letters from X. The number of letters in a word η, written as $|\eta|$, is called its *length*. The empty word, $\varnothing$, is taken to have length zero. For a given $x_i \in X$, $|\eta|_{x_i}$ is the number of times x_i appears in η. The collection of all words having length k is denoted by X^k. Define $X^+ = \bigcup_{k \geq 1} X^k$ and $X^* = \bigcup_{k \geq 0} X^k$, the latter of which is a noncommutative monoid under concatenation. Any mapping $c \colon X^* \to \mathbb{R}^\ell$ is called a *formal power series*. It is often written as the formal sum $c = \sum_{\eta \in X^*} (c, \eta) \eta$, where the *coefficient* $(c, \eta) \in \mathbb{R}^\ell$ is the image of $\eta \in X^*$ under c. The *support* of c, $\mathrm{supp}(c)$, is the set of all words having nonzero coefficients. The set of all noncommutative formal power series over the alphabet X is denoted by $\mathbb{R}^\ell \langle\langle X \rangle\rangle$. The subset of series with finite support, i.e., polynomials, is represented by $\mathbb{R}^\ell \langle X \rangle$. As $\mathbb{R}\langle X \rangle$ is dense in $\mathbb{R}\langle\langle X \rangle\rangle$ (under the ultrametric topology [3]), $\mathbb{R}^\ell \langle\langle X \rangle\rangle$ can be viewed as the completion of $\mathbb{R}^\ell \langle X \rangle$. Given any language $L \subseteq X^*$, its *characteristic series* in $\mathbb{R}\langle\langle X \rangle\rangle$ is $\mathrm{char}(L) = \sum_{\eta \in L} \eta$.

2.1 Entropy of graded vector spaces

Let $\mathbb{N} = \{1, 2, \ldots\}$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. An $\mathbb{R}$-vector space V is said to be $\mathbb{N}_0$-*graded* over $\mathbb{R}$ if $V = \bigoplus_{n \in \mathbb{N}_0} V_n$, where each V_n is an $\mathbb{R}$-vector subspace. Elements in V_n are said to have degree n. V is *connected* if $V_0 = \mathbb{R}$ and *locally finite* if each V_n has finite dimension. The *Hilbert–Poincaré series* of a graded locally finite vector space is defined to be the formal power series in indeterminate z [1, 2]

$$V(z) = \sum_{n=0}^\infty \dim(V_n) z^n.$$

Often $V(z)$ is a rational function indicating some type of linear recursion relates the sequence of dimensions $d_n := \dim(V_n)$, $n \geq 0$. The *entropy* of a graded vector space V is taken to be

$$H(V) = \limsup_{n \to \infty} \sqrt[n]{\dim(V_n)},$$

whenever the limit exists [41].

Example 2.1. Consider $\mathbb{R}\langle\langle X \rangle\rangle$ with scalar multiplication and addition defined in the usual way to form an $\mathbb{R}$-vector space. If $\mathbb{R}\langle\langle X \rangle\rangle$ is graded by word length, then it is connected and locally finite with $V_n = \mathrm{span}_{\mathbb{R}}\{\eta \in X^* \colon |\eta| = n\}$ and $d_n = \dim(V_n) = (m+1)^n$ so that

$$V(z) = \sum_{n=0}^\infty (m+1)^n z^n = \frac{1}{1 - (m+1)z}.$$

As the sequence of dimensions is a geometric sequence, clearly, $d_{n+1} = (m+1)d_n$, $n \geq 0$ with $d_0 = 1$. The entropy is $H(\mathbb{R}\langle\langle X \rangle\rangle) = m + 1$.

Consider next an alternative grading of $\mathbb{R}\langle\langle X \rangle\rangle$, where the letter x_0 has twice the degree of the other letters and $\deg(\varnothing) = 1$, that is, $\deg(\eta) = 2|\eta|_{x_0} + \sum_{i=1}^m |\eta|_{x_i} + 1$ for all $\eta \in X^*$

[18, 21]. Let $|A|$ denote the cardinality of set A. Define the formal power series in commuting indeterminates $z_0, z_1, \ldots z_m$

$$W(z_0, z_1, \ldots, z_m) = \sum_{k_0, k_1, \ldots, k_m = 0}^{\infty} |\{\eta \in X^* \colon |\eta|_{x_j} = k_j, j = 0, 1, \ldots, m\}| z_0^{k_0} z_1^{k_1} \cdots z_m^{k_m}$$

$$= \frac{1}{1 - (z_0 + z_1 + \cdots + z_m)}.$$

In light of the assumed grading, it follows that

$$V(z) = 1 + mzW(z^2, z, \ldots, z) = \frac{1 - z^2}{1 - mz - z^2}$$
$$= 1 + mz + m^2 z^2 + (m + m^3) z^3 + (2m^2 + m^4) z^4 + (m + 3m^3 + m^5) z^5$$
$$+ (3m^2 + 4m^4 + m^6) z^6 + (m + 6m^3 + 5m^5 + m^7) z^7 + O(z^8).$$

In which case, $d_0 = 1$, $d_1 = m$, $d_2 = m^2$,

$$d_{n+1} = md_n + d_{n-1}, \qquad n \geq 2$$

and

$$d_n \sim \frac{m}{\sqrt{m^2 + 4}} \left(\frac{m + \sqrt{m^2 + 4}}{2} \right)^n, \qquad n \gg 1.$$

If $m = 1$, then d_n, $n \geq 1$ is the Fibonacci sequence, and $d_n \sim \varphi^n / \sqrt{5}$ with $\varphi = (1 + \sqrt{5})/2$ being the golden ratio. The integer sequences for $m = 2$ and $m = 3$ are also well studied (see A052542 and A052906, respectively, in [48]). Under this alternative grading, the entropy is $H(\mathbb{R}\langle\langle X \rangle\rangle) = (m + \sqrt{m^2 + 4})/2 < m + 1$.

2.2   Elementary products of formal power series

Suppose multiplication on $\mathbb{R}^\ell$ is defined componentwise. Then there are three elementary products on $\mathbb{R}^\ell \langle\langle X \rangle\rangle$ that render it an associative unital $\mathbb{R}$-algebra [14, 16]. The first is the commutative *Hadamard product* defined as

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

The second is the noncommutative *Cauchy product*

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

Lastly, $\mathbb{R}^\ell \langle\langle X \rangle\rangle$ is a commutative $\mathbb{R}$-algebra under the bilinear *shuffle product*

$$c \shuffle d = \sum_{\eta, \xi \in X^*} (c, \eta)(d, \xi)\, \eta \shuffle \xi,$$

where the shuffle product of two words $x_i\eta, x_j\xi \in X^*$ is defined inductively by

$$(x_i\eta) \shuffle (x_j\xi) = x_i(\eta \shuffle (x_j\xi)) + x_j((x_i\eta) \shuffle \xi)$$

with $x_i, x_j \in X$ and $\eta \shuffle \varnothing = \varnothing \shuffle \eta = \eta$. Each product is locally finite. In addition,

$$\mathrm{supp}(c + d) \subseteq \mathrm{supp}(c) \cup \mathrm{supp}(d),$$

$$\text{supp}(c \odot d) = \text{supp}(c) \cap \text{supp}(d),$$
$$\text{supp}(cd) \subseteq \text{supp}(c)\text{supp}(d).$$

The relationship between the support of the shuffle product and that of its arguments is much more complicated, see, for example, [38]. Enumerating the number of distinct words from shuffle products was investigated in [49]. But asymptotic results along these lines, which would be very useful here, appear to be unknown at present.

2.3   Chen–Fliess series

Given any $c \in \mathbb{R}^\ell\langle\langle X \rangle\rangle$, one can associate a causal m-input, ℓ-output operator, F_c, in the following manner. Let $\mathfrak{p} \geq 1$ and $t_0 < t_1$ be given. For a Lebesgue measurable function $u \colon [t_0, t_1] \to \mathbb{R}^m$, define $\|u\|_{\mathfrak{p}} = \max\{\|u_i\|_{\mathfrak{p}} : 1 \leq i \leq m\}$, where $\|u_i\|_{\mathfrak{p}}$ is the usual $L_{\mathfrak{p}}$-norm for a measurable real-valued function, u_i, defined on $[t_0, t_1]$. Let $L_{\mathfrak{p}}^m[t_0, t_1]$ denote the set of all measurable functions defined on $[t_0, t_1]$ having a finite $\|\cdot\|_{\mathfrak{p}}$ norm and $B_{\mathfrak{p}}^m(R)[t_0, t_1] := \{u \in L_{\mathfrak{p}}^m[t_0, t_1] \colon \|u\|_{\mathfrak{p}} \leq R\}$. Assume $C[t_0, t_1]$ is the subset of continuous functions in $L_1^m[t_0, t_1]$. Define inductively for each word $\eta = x_i\bar{\eta} \in X^*$ the map $E_\eta \colon L_1^m[t_0, t_1] \to C[t_0, t_1]$ by setting $E_\varnothing[u] = 1$ and letting

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

where $x_i \in X$, $\bar{\eta} \in X^*$, and $u_0 = 1$. The *Chen–Fliess series* corresponding to $c \in \mathbb{R}^\ell\langle\langle X \rangle\rangle$ is [16]

$$y(t) = F_c[u](t) = \sum_{\eta \in X^*} (c, \eta) E_\eta[u](t, t_0). \tag{2.1}$$

To establish the convergence of this series, assume there exist real numbers $K, M > 0$ such that

$$|(c, \eta)| \leq K M^{|\eta|} |\eta|!, \qquad \forall \eta \in X^*. \tag{2.2}$$

(Define $|z| := \max_i |z_i|$ whenever $z \in \mathbb{R}^\ell$.) It is shown in [28] that under such circumstances the series (2.1) converges uniformly and absolutely so that F_c describes a well defined mapping from $B_{\mathfrak{p}}^m(R)[t_0, t_0 + T]$ into $B_{\mathfrak{q}}^\ell(S)[t_0, t_0 + T]$ for sufficiently small $R, T > 0$, where $\mathfrak{p}, \mathfrak{q} \in [1, \infty]$ satisfy $1/\mathfrak{p} + 1/\mathfrak{q} = 1$. The operator F_c is said to be *locally convergent* and is called a *Fliess operator*. The collection of all generating series c satisfying the growth condition (2.2) is denoted by $\mathbb{R}_{LC}^\ell\langle\langle X \rangle\rangle$.

A special class of Chen–Fliess series consists of those having an input-output map $y = F_c[u]$ with a smooth n dimensional state space realization

$$\dot{z} = g_0(z) + \sum_{i=1}^m g_i(z) u_i, \qquad z(0) = z_0, \qquad y = f(z)$$

on some open set $W \subseteq \mathbb{R}^n$. In this case, the generating series is computed by

$$(c, \eta) = L_\eta f(z_0) := L_{g_{i_1}} \cdots L_{g_{i_k}} f(z_0) \tag{2.3}$$

for any $\eta = x_{i_k} \cdots x_{i_1} \in X^*$, where $z_0 \in \mathbb{R}^n$, and $L_{g_i} f$ is the Lie derivative of the smooth output function $f \colon W \to \mathbb{R}$ with respect to vector field g_i [16, 30]. The *control Lie algebra*, $\mathscr{C}(g)$, for this realization is the smallest Lie algebra containing g_i, $i = 0, 1, \ldots, m$ which is closed under the Lie bracket operation

$$[g_i, g_j](z) = \frac{\partial g_j}{\partial z} g_i(z) - \frac{\partial g_i}{\partial z} g_j(z).$$

Identifying $g_i \sim x_i$ and letting $[x_i, x_j] = x_j x_i - x_i x_j$, $\mathscr{C}(g)$ is isomorphic to a Lie subalgebra $\mathscr{C}(X)$ of the free Lie algebra over X. $\mathscr{C}(X)$ will be referred to as the control Lie algebra over X.

2.4 System interconnections

Given Fliess operators F_c and F_d, where $c, d \in \mathbb{R}_{LC}^{\ell}\langle\langle X\rangle\rangle$, the parallel sum and parallel product connections satisfy $F_c + F_d = F_{c+d}$ and $F_c F_d = F_{c \sqcup\!\sqcup d}$, respectively [16]. When Fliess operators F_c and F_d with $c \in \mathbb{R}_{LC}^{\ell}\langle\langle X\rangle\rangle$ and $d \in \mathbb{R}_{LC}^{m}\langle\langle X\rangle\rangle$ are interconnected in a cascade fashion, the composite system $F_c \circ F_d$ has the Fliess operator representation $F_{c \circ d}$, where the *composition product* of c and d is given by [12, 13]

$$c \circ d = \sum_{\eta \in X^*} (c, \eta)\psi_d(\eta)(\mathbf{1}).$$

Here $\mathbf{1}$ denotes the monomial $1\varnothing$, and ψ_d is the continuous (in the ultrametric sense) algebra homomorphism from $\mathbb{R}\langle\langle X\rangle\rangle$ to the set $\mathrm{End}(\mathbb{R}\langle\langle X\rangle\rangle)$ of vector space endomorphisms on $\mathbb{R}\langle\langle X\rangle\rangle$ uniquely specified by $\psi_d(x_i\eta) = \psi_d(x_i) \circ \psi_d(\eta)$ with $\psi_d(x_i)(e) = x_0(d_i \sqcup\!\sqcup e)$, $i = 0, 1, \ldots, m$ for any $e \in \mathbb{R}\langle\langle X\rangle\rangle$, and where d_i is the i-th component series of d ($d_0 := \mathbf{1}$). By definition, $\psi_d(\varnothing)$ is the identity map on $\mathbb{R}\langle\langle X\rangle\rangle$. A generalized series δ is defined as the unit for the composition product, that is, $\delta \circ c = c \circ \delta = c$ for all $c \in \mathbb{R}\langle\langle X\rangle\rangle$, or, equivalently, F_δ is the identity map on the input space so that $F_\delta[u] = u$ for all admissible inputs u.

Finally, consider the additive feedback connection with F_c in the forward path and F_d in the feedback path. The generating series $e \in \mathbb{R}\langle\langle X\rangle\rangle$ defines a closed-loop system $y = F_e[u]$ when it satisfies the feedback equation $F_e[u] = F_c[u + F_{d \circ e}[u]]$. It is known that there always exists such an $e = c@d$, where the series $c@d$ is the *feedback product* of c and d [21]. This product has an explicit formula in terms of the antipode of a graded connected Hopf algebra [18]. Its construction relies in part on the shuffle algebra.

3 Entropy of generating series

Consider $\mathbb{R}\langle\langle X\rangle\rangle$ as a graded $\mathbb{R}$-vector space with finite dimensional homogeneous components V_n spanned by all the words of degree n, $X(n)$, so that $\dim(V_n) = |X(n)|$. Given a series $c \in \mathbb{R}\langle\langle X\rangle\rangle$ and $n \in \mathbb{N}_0$, let $\mathrm{supp}_n(c) := \mathrm{supp}(c) \bigcap X(n)$. Define the *support sequence* of c to be $S_c = \{n_k \in \mathbb{N}_0 : |\mathrm{supp}_{n_k}(c)| \neq 0, \ k \in \mathbb{N}_0\}$. It is assumed that $|X(n)| \sim K\gamma^n$ for some real numbers $K, \gamma > 0$ when $n \gg 0$.

Definition 3.1. The **entropy** of a series $c \in \mathbb{R}\langle\langle X\rangle\rangle/\mathbb{R}\langle X\rangle$ with support sequence S_c is

$$h(c) = \limsup_{k \to \infty} \frac{1}{n_k} \log_\gamma(|\mathrm{supp}_{n_k}(c)|).$$

If $c \in \mathbb{R}\langle X\rangle$, then $h(c) := 0$.

The entropy of a generating series is roughly a (normalized) measure of the rate at which its support grows with increasing degree relative to the maximum rate it could grow as determined by the grading.

Theorem 3.2. *For every $c \in \mathbb{R}\langle\langle X\rangle\rangle$, $h(c)$ is well defined with $0 \leq h(c) \leq 1$.*

Proof. The only nontrivial case is when $c \in \mathbb{R}\langle\langle X\rangle\rangle/\mathbb{R}\langle X\rangle$. Consider the sequence of real numbers $a_k = \log_\gamma(|\mathrm{supp}_{n_k}(c)|)/n_k$, $n_k \in S(c)$, $k \geq 0$. It is clearly bounded from below by zero. It is also bounded from above by one since no series in $\mathbb{R}\langle\langle X\rangle\rangle$ has a larger support than $\mathrm{char}(X^*) = \sum_{\eta \in X^*} \eta$, and $\mathrm{supp}_k(\mathrm{char}(X^*)) \sim K\gamma^k$, $k \gg 0$. As the infinite sequence a_k, $k \geq 0$ in $\mathbb{R}$ is bounded, it must have at least one cluster point. The largest cluster point uniquely defines $h(c)$. ∎

If the sequence a_k, $k \geq 0$ defined above converges, then clearly $h(c)$ is equivalent to this limit. In addition, it is evident that h is not homogeneous as $h(\alpha c) = h(c)$ for all nonzero $\alpha \in \mathbb{R}$. Unless stated otherwise, the default assumption in most examples will be that $X = \{x_0, x_1\}$, $\mathbb{R}\langle\langle X \rangle\rangle$ is graded by word length so that $K = 1$, $\gamma = 2$, and either $S(c) = \mathbb{N}_0$ or $S(c) = \mathbb{N}$.

Example 3.3. A series (or polynomial) is called a *repeated word series* if the only words in its support are powers of a single word $\xi \in X^+$. For example, $c = x_1 + x_1^2$ and $d = \sum_{n \geq 0}(x_0 x_1 x_0)^n$ are repeated word series. Clearly, $h(c) = 0$. In the latter case, where $\xi = x_0 x_1 x_0$, observe $S_d = \{k|\xi|: k \geq 0\}$ and $\text{supp}_{k|\xi|}(d) = 1$, $k \geq 0$, so that $h(d) = 0$. Therefore, all repeated word series have zero entropy.

Example 3.4. A series $c \in \mathbb{R}\langle\langle X \rangle\rangle$ is said to be *linear* if $|\eta|_{x_1} = 1$ for all $\eta \in \text{supp}(c)$. For example, $c = \sum_{n \geq r} x_0^{n-1} x_1$ is a linear series with relative degree $r \geq 1$ [27]. When $r = 1$, for example, the input-output map F_c is realized by the linear time-invariant system

$$\dot{z} = z + u, \qquad z(0) = 0, \qquad y = z.$$

Observe $\text{supp}_k(c) = 1$, $k \geq r$ so that $h(c) = 0$.

Example 3.5. Consider the linear series $c = \sum_{n_0, n_1 \geq 0} x_0^{n_0} x_1 x_0^{n_1}$. In this case, $\text{supp}_k(c) = k+1$, $k \geq 0$, and thus, $h(c) = \lim_{k \to \infty} \log_2(k+1)/k = 0$. This implies that the generating series for any linear operator F_c has zero entropy.

Example 3.6. A series $c \in \mathbb{R}\langle\langle X \rangle\rangle$ is said to be *input-limited* if for some fixed $N \in \mathbb{N}$, $|\eta|_{x_1} \leq N$ for all $\eta \in \text{supp}(c)$. It is known in general that rationality is not preserved under the composition product. However, $c \circ d$ is rational if c and d are rational, and c is input-limited [12, 13]. Consider, for example, the input-limited series

$$c_N := \sum_{n_0, n_1, \ldots, n_N = 0}^{\infty} x_0^{n_0} x_1 x_0^{n_1} x_1 \cdots x_0^{n_{N-1}} x_1 x_0^{n_N}.$$

The number of compositions (order partitions) of a nonnegative integer $K = n_0 + n_1 + n_2 + \cdots + n_N$ into $N + 1$ parts where $n_i \geq 0$ is $\binom{K+N}{N}$. Therefore, $\text{supp}_k(c) = \binom{k}{N}$, $k \geq N$. Hence,

$$h(c) = \lim_{k \to \infty} \frac{1}{k} \log_2 \left(\frac{k(k-1) \cdots (k-N+1)}{N!} \right) \leq \lim_{k \to \infty} \frac{N}{k} \log_2(k) - \frac{1}{k} \log_2(N!) = 0,$$

so that $h(c) = 0$. This is in fact the case for all input-limited series.

Example 3.7. Given a word $\xi \in X^*$, let $\tilde{\xi}$ denote the word whose letters are written in reverse order. For example, if $\xi = x_0 x_1 x_1$, then $\tilde{\xi} = x_1 x_1 x_0$. Consider the set of all even *palindromes*

$$P = \{\xi\tilde{\xi}: \xi \in X^*\}.$$

Observe if $\eta \in P$ then $\tilde{\eta} \in P$. It can be directly checked that $\left|P \bigcap X^{2n}\right| = 2^n$, $n \geq 0$. Setting $c = \sum_{\eta \in P} \eta$, it follows that $S_c = \{2k: k \geq 0\}$ and $\text{supp}_{2k}(c) = 2^k$, $k \geq 0$. In which case, $h(c) = 1/2$.

Example 3.8. A *word power series* has the form $c_{1/N} = \sum_{\eta \in X^*} \eta^N$, where η^N denotes the Cauchy product power and $N \in \mathbb{N}$ is fixed. Analysis similar to that in the previous example gives $h(c) = 1/N$.

Example 3.9. The word power series $\text{char}(X^*) = \sum_{\eta \in X^*} \eta$ appearing in the proof of Theorem 3.2 has entropy one in light of the previous example, as does the series $d = c - x_0^*$, where $x_0^* := \sum_{k \geq 0} x_0^k$.

Example 3.10. Consider the bilinear state space realization

$$\dot{z} = z + zu, \qquad z(0) = 1, \qquad y = z.$$

Clearly, $\dim(\mathscr{C}(X)) = 1$ since $g_0 = g_1$ and $[g_0, g_1] = 0$. As an algebra, $\mathscr{C}(X)$ has zero entropy [41]. But observe that $(c, \eta) = L_\eta f(z_0) = 1$ for all $\eta \in X^*$. Thus, $c = \sum_{\eta \in X^*} \eta$ so that $h(c) = 1$. That is, the generating series of the input-output map has maximum entropy even though the underlying state space realization has zero entropy in its control Lie algebra. This is possible in light of (2.3) and the fact that $L_{[g_i, g_j]} h = L_{g_i} L_{g_j} h - L_{g_j} L_{g_i} h = 0$ does not imply that $L_{g_i} L_{g_j} h = L_{g_j} L_{g_i} h = 0$. This will be clearer in the next section when this example is revisited using the shuffle algebra, which is playing a hidden role in this problem.

For any letter $x_i \in X$, let x_i^{-1} denote the $\mathbb{R}$-linear *left-shift operator* defined by $x_i^{-1}(\xi) = \xi'$ when $\xi = x_i \xi'$ and zero otherwise. It is defined inductively for higher order shifts via $(x_i \xi)^{-1} = \xi^{-1} x_i^{-1}$, where $\xi \in X^*$. The left-shift operator acts as a derivation on the shuffle product. The *left-augmentation* of $c \in \mathbb{R}\langle\langle X \rangle\rangle$ by $\xi \in X^*$ is simply the Cauchy product $\xi c = \sum_{\eta \in X^*}(c, \eta)\xi\eta = \sum_{\eta \in X^*}(\xi c, \eta)\eta$. Likewise, the *right-augmentation* of $c \in \mathbb{R}\langle\langle X \rangle\rangle$ by $\xi \in X^*$ is $c\xi = \sum_{\eta \in X^*}(c, \eta)\eta\xi = \sum_{\eta \in X^*}(c\xi, \eta)\eta$.

Theorem 3.11. *For any series $c \in \mathbb{R}\langle\langle X \rangle\rangle$ and word $\xi \in X^*$, $h(\xi^{-1}(c)) \leq h(c)$ and $h(\xi c) = h(c\xi) = h(c)$.*

Proof. The first claim is a consequence of the fact that $\text{supp}_k(\xi^{-1}(c)) \subseteq \text{supp}_k(c)$, $k \geq 0$. The equalities follow directly from the identities $|\text{supp}_{k+\deg(\xi)}(\xi c)| = |\text{supp}_{k+\deg(\xi)}(c\xi)| = |\text{supp}_k(c)|$, $k \geq 0$. ∎

In general, entropy is invariant only under augmentations.

Example 3.12. Suppose $c = \sum_{\eta \in X^*} x_0 \eta$. Then $h(c) = 1$, $h(x_0^{-1}(c)) = 1$, and $h(x_1^{-1}(c)) = 0$. On the other hand, $h(x_i c) = h(c x_i) = 1$, $i = 0, 1$.

4 Entropy and interconnected nonlinear systems

Definition 4.1. A magma $(\mathbb{R}\langle\langle X \rangle\rangle, \square)$ with entropy function h is said to be *entropy bounded* if

$$h(c \square d) \leq \max(h(c), h(d)), \qquad \forall c, d \in \mathbb{R}\langle\langle X \rangle\rangle.$$

In essence, an entropy bounded product $\square$ cannot create more entropy than that which is already present in its arguments. The following is the first of two main results in this section.

Theorem 4.2. *Addition, the Hadamard product, and the Cauchy product on $\mathbb{R}\langle\langle X \rangle\rangle$ are all entropy bounded.*

Proof. When necessary, it is sufficient to consider only the case where c and d are *positively supported*. That is, if $\eta \in \text{supp}(c)$, then $(c, \eta) > 0$ and likewise for d. For the products under consideration, the restriction to positively supported series will eliminate the possibility of inter-term cancellations. Such cancellations can reduce the cardinality of the support of the resulting series but never increase it. Without loss of generality, it is assumed below that all the support sequences are equivalent to $\mathbb{N}_0$.

Addition. A slightly stronger result is possible when c and d are positively supported, namely, $h(c + d) = \max(h(c), h(d))$. Otherwise, the possibility of cancellations makes the right-hand side only an upper bound. Suppose c and d are positively supported. Observe then that $\text{supp}(c) \subseteq \text{supp}(c + d)$, $\text{supp}(d) \subseteq \text{supp}(c + d)$, and

$$\text{supp}(c + d) = \text{supp}(c) \cup \text{supp}(d).$$

Therefore, $|\mathrm{supp}_k(c)| \leq |\mathrm{supp}_k(c+d)|$, $|\mathrm{supp}_k(d)| \leq |\mathrm{supp}_k(c+d)|$, and

$$|\mathrm{supp}_k(c+d)| \leq |\mathrm{supp}_k(c)| + |\mathrm{supp}_k(d)|, \qquad \forall k \geq 0. \tag{4.1}$$

Since $\log_2$ is an increasing function, the first two inequalities yield $\max(h(c), h(d)) \leq h(c+d)$. Consider a sequence $a = \{a_k \in \mathbb{R} : a_k \geq 1, k \in \mathbb{N}_0\}$ such that $\lambda(a) := \limsup_{k\to\infty} \log_2(a_k)/k$ is a real number. It can be verified for any two such sequences a and b that

$$\lambda(a+b) = \max(\lambda(a), \lambda(b)). \tag{4.2}$$

Hence, combining (4.1) and (4.2) gives $h(c+d) \leq \max(h(c), h(d))$, and the claim is proved.

Hadamard product. If $\eta \in \mathrm{supp}_k(c \odot d)$, then $\eta \in \mathrm{supp}_k(c) \cap \mathrm{supp}_k(d)$. But since $\mathrm{supp}_k(c) \cap \mathrm{supp}_k(d) \subseteq \mathrm{supp}_k(c)$ and $\mathrm{supp}_k(c) \cap \mathrm{supp}_k(d) \subseteq \mathrm{supp}_k(d)$, it follows that

$$|\mathrm{supp}_k(c \odot d)| \leq \min(|\mathrm{supp}_k(c)|, |\mathrm{supp}_k(d)|), \qquad \forall k \geq 0.$$

Therefore, in general,

$$h(c \odot d) \leq \min(h(c), h(d)).$$

Cauchy product. Assume that $c, d \in \mathbb{R}\langle\langle X \rangle\rangle$ are positively supported. Then $\nu \in \mathrm{supp}_k(cd)$ if and only if

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

where for at least one pair of words (η, ξ) with $|\eta| + |\xi| = k$, the product $(c, \eta)(d, \xi) \neq 0$. Therefore,

$$|\mathrm{supp}_k(cd)| \leq \sum_{j=0}^{k} |\mathrm{supp}_{k-j}(c)||\mathrm{supp}_j(d)|, \qquad \forall k \geq 0.$$

Note that in general this upper bound is conservative since ν can be constructed by concatenation of differently sized words, i.e., $\nu = x_1 x_0 x_1 = (x_1)(x_0 x_1) = (x_1 x_0)(x_1)$. Now define the sequences $a = \{|\mathrm{supp}_k(c)| : k \geq 0\}$ and $b = \{|\mathrm{supp}_k(d)| : k \geq 0\}$. Observe for $k \geq 1$ that

$$\log_2\left(\sum_{j=0}^{k} a_{k-j} b_j\right) \leq \max_{0 \leq j \leq k} \log_2(k a_{k-j} b_j) = \log_2(k) + \max_{0 \leq j \leq k} \log_2(a_{k-j} b_j).$$

Therefore,

$$\limsup_{k\to\infty} \max_{0 \leq j \leq k} \frac{1}{k} \log_2(a_{k-j} b_j) = \limsup_{k\to\infty} \max_{0 \leq j \leq k} \frac{k-j}{k}\left(\frac{1}{k-j} \log_2(a_{k-j})\right) + \frac{j}{k}\left(\frac{1}{j} \log_2(b_j)\right)$$
$$\leq \max(\lambda(a), \lambda(b))$$

since $(k-j/k) + (j/k) = 1$ and the terms $\log_2(a_{k-j})/(k-j)$ and $\log_2(b_j)/j$ are asymptotically bounded by $\lambda(a)$ and $\lambda(b)$, respectively. Thus, $h(cd) \leq \max(h(c), h(d))$ as claimed. ∎

It is assumed in the following set of examples that $X = \{x_0, x_1\}$.

Example 4.3. Define c_0 to be the series in Example 3.3 when $\xi = x_0$. Let c_N denote the series in Example 3.6 when $N \geq 1$. Theorem 4.2 implies that for any integer $M \geq 0$, the input-limited series $d_M = \sum_{N=0}^{M} c_N$ has zero entropy. Note, however, that this property is not true in the limit since $\lim_{M\to\infty} d_M = \mathrm{char}(X^*)$ and $h(\mathrm{char}(X^*)) = 1$. This indicates that $h: \mathbb{R}\langle\langle X \rangle\rangle \to [0, 1]$ is not continuous when convergence on $\mathbb{R}\langle\langle X \rangle\rangle$ is defined in the ultrametric sense [3].

Example 4.4. Reconsider the series of even palindromes $c = \sum_{\eta \in P} \eta$ in Example 3.7, where $h(c) = 1/2$. Let $d = \sum_{k \geq 0} x_0^k$ so that $h(d) = 0$. Then $c \odot d = \sum_{k=0}^{\infty} x_0^{2k}$ so that $h(c \odot d) = 0 = \min(1/2, 0)$ as expected.

Example 4.5. All the series in Examples 3.3–3.6 can all be written as the Cauchy product of repeated words series. For example, the series in Example 3.6 can be written as $c_N = d_0^* x_1 d_1^* x_1 \cdots d_{N-1}^* x_1 d_N^*$, where $d_i^* = \sum_{k \geq 0} x_0^k$. Thus, by Theorem 4.2, every series in these examples must have zero entropy as was demonstrated directly from the definition.

Example 4.6. Reconsider the series of even palindromes $c = \sum_{\eta \in P} \eta$ in Example 3.7 where $h(c) = 1/2$. The support of series cc now contains elements that are *not* palindromes, for example, $x_0^2 x_1^2$. Nevertheless, Theorem 4.2 requires that $h(cc) \leq 1/2$. This can be verified by a direct calculation. Note that every nonempty word in $\mathrm{supp}_k(cc)$ has the form $\eta \tilde{\eta} \xi \tilde{\xi}$, where $2|\eta| + 2|\xi| = k = 2k'$ with $k' := i + j$, $i := |\eta| \geq 1$, and $j := |\xi| \geq 1$. Observe that $k' := i + j \geq 2$ has exactly $\binom{k'-1}{1} = k' - 1$ compositions with two parts i and j. Thus, there are $(k'-1)2^{k'}$ words in $\mathrm{supp}_k(cc)$. In which case, $\log_2(|\mathrm{supp}_k(cc)|)/k = \log_2((k'-1)2^{k'})/2k' = 1/2 + \log_2(k'-1)/2k'$ so that in fact $h(cc) = 1/2$.

The second main result of this section states that the remaining products of interest on $\mathbb{R}\langle\langle X \rangle\rangle$ are not entropy bounded. The claim is demonstrated via counterexamples presented subsequently.

Theorem 4.7. *The shuffle product, composition product, and feedback product on $\mathbb{R}\langle\langle X \rangle\rangle$ are not entropy bounded.*

Example 4.8. This example demonstrates that the shuffle product is not entropy bounded. If $X = \{x_0, x_1\}$, then $x_i^* = \sum_{k=0}^{\infty} x_i^k$ has zero entropy. It is known that [8, 10]

$$\mathrm{char}(X^k) = \sum_{\substack{r_0, r_1 \geq 0 \\ r_0 + r_1 = k}} x_0^{r_0} \shuffle x_1^{r_1}, \qquad k \geq 0.$$

Therefore,

$$\sum_{\eta \in X^*} \eta = \mathrm{char}(X^*) = \sum_{k=0}^{\infty} \mathrm{char}(X^k) = \sum_{k=0}^{\infty} \sum_{\substack{r_0, r_1 \geq 0 \\ r_0 + r_1 = k}} x_0^{r_0} \shuffle x_1^{r_1} = x_0^* \shuffle x_1^*.$$

That is,

$$h(x_0^* \shuffle x_1^*) = 1 > 0 = \max\{h(x_0^*), h(x_1^*)\}.$$

On the other hand, there are specific instances where the entropy bounded property holds, for example, $x_i^* \shuffle x_i^* = (2x_i)^*$ so that $h(x_i^* \shuffle x_i^*) = h((2x_i)^*) = 0$ [19].

Example 4.9. This example illustrates that the composition product is not entropy bounded. Let $c = x_1^*$ so that $h(c) = 0$. It is shown in [24] that

$$\left(c \circ c, x_0^{k_0} x_1^{k_1} \cdots x_0^{k_{l-1}} x_1^{k_l}\right) = (k_0)^{k_1} (k_0 + k_2)^{k_3} \cdots (k_0 + k_2 + k_4 + \cdots + k_{l-1})^{k_l}$$

for all odd $l \geq 1$ and $k_i \geq 0$, $i = 0, 1, \ldots, l$ (assume $0^0 := 1$). Therefore, $\mathrm{supp}(c \circ c) = X^*$ and $h(c \circ c) = 1$.

Example 4.10. It is shown in this example that the feedback product is not entropy bounded. Assume $X = \{x_0, x_1\}$, and let $c = \sum_{k \geq 0} k!\, x_1^k$ so that $h(c) = 0$. Let $d = \delta$, the composition unit. Formally, $\delta \cap X^*$ is empty so that $h(\delta) := 0$. Consider an additive unity feedback interconnection with F_c in the forward path and $F_\delta = I$ in the feedback path. This defines a feedback product $e = c@\delta$. The generating series $c@\delta$ was computed explicitly in [11] using Hopf algebra methods described in [9, 21] and found to have the form $c@\delta = \sum_{n \geq 1} b_n$, where the polynomial sequence $b_n \in \mathbb{R}\langle X \rangle$, $n \geq 1$ satisfies the linear recursion

$$b_n = (n-1)b_{n-1}x_1 + (n-2)b_{n-2}x_0, \qquad n \geq 2 \tag{4.3}$$

with $b_0 = 0$ and $b_1 = 1$. These are the *Devlin polynomials* that appear in a combinatorial characterization of periodic solutions to the Abel equation in the context of Hilbert's 16th problem [5, 6]. The first few polynomials are:

$$b_1 = 1,$$
$$b_2 = x_1,$$
$$b_3 = 2x_1^2 + x_0,$$
$$b_4 = 6x_1^3 + 3x_0x_1 + 2x_1x_0,$$
$$b_5 = 24x_1^4 + 12x_0x_1^2 + 8x_1x_0x_1 + 6x_1^2x_0 + 3x_0^2,$$
$$b_6 = 120x_1^5 + 60x_0x_1^3 + 40x_1x_0x_1^2 + 30x_1^2x_0x_1 + 24x_1^3x_0 + 15x_0^2x_1 + 12x_0x_1x_0 + 8x_1x_0^2.$$

It is not hard to see from (4.3) that $\mathrm{supp}(c@\delta) = X^*$ so that if $\mathbb{R}\langle\langle X \rangle\rangle$ is graded by word length then $h(c@\delta) = 1$. It is shown in [5, Theorem 4.1] that for $n \geq 1$

$$\mathrm{supp}(b_n) = \{\eta \in X^*\colon \deg(\eta) = 2|\eta|_{x_0} + |\eta|_{x_1} + 1 = n\}.$$

Therefore, under the alternative grading in Example 2.1, where $|X(n)| \sim \left(1/\sqrt{5}\right)\varphi^n$ with $\varphi = \left(1 + \sqrt{5}\right)/2$ when $m = 1$, it is evident that $h(c@\delta) = 1$. It is worth noting that this second grading is also the grading for the polynomial algebra of coordinate functions that defines the output feedback Hopf algebra used to compute the general feedback product $c@d$ as discussed in Section 2.4.

The final two examples provide some applications of these results in the context of nonlinear control theory.

Example 4.11. Let c and d be generating series for two linear time-invariant systems as described in Example 3.4. As linearity is preserved under the parallel sum, composition, and feedback interconnections, the resulting generating series in each case must have zero entropy. The parallel product connection does not preserve linearity but nevertheless yields a zero entropy generating series. To see this, observe that every word in the support of $x_0^i x_1 \shuffle x_0^j x_1$ must have the form ξx_1 with $\xi \in X^{i+j-1}$ having the letter x_1 in any position. While it is *not* true that $x_0^* x_1 \shuffle x_0^* x_1 = x_0^* x_1 x_0^* x_1$, it does hold that $\mathrm{supp}((x_0^* x_1) \shuffle (x_0^* x_1)) = \mathrm{supp}(x_0^* x_1 x_0^* x_1)$, and the latter series has zero entropy as discussed in Example 4.5.

Example 4.12. Reconsider the state space system in Example 3.10 after the coordinate transformation $z = \exp(\bar{z})$ has been applied, namely,

$$\dot{\bar{z}} = 1 + u, \qquad \bar{z}(0) = 0, \qquad y = \exp(\bar{z}).$$

The generating series is known to be invariant under a change of coordinates [30]. Observe $\bar{z} = F_{x_0+x_1}[u]$ so that

$$y = \exp(F_{x_0+x_1}[u]) = \sum_{n=0}^{\infty} (F_{x_0+x_1}[u])^n \frac{1}{n!} = \sum_{n=0}^{\infty} (F_{(x_0+x_1)^{\shuffle n}}[u]) \frac{1}{n!},$$

using the shuffle power identity $F_c^n[u] = F_{c^{\sqcup\!\sqcup n}}[u]$. Therefore, $y = F_c[u]$, where

$$c = \sum_{n=0}^{\infty}(x_0 + x_1)^{\sqcup\!\sqcup n}\frac{1}{n!} = \sum_{n=0}^{\infty}\mathrm{char}(X^n) = \sum_{\eta \in X^*}\eta,$$

where the identity $\mathrm{char}(X^n) = (\mathrm{char}(X))^{\sqcup\!\sqcup n}/n!$ has been used above. What is clear in this coordinate system is that the output function is largely responsible for the maximal growth in the entropy of the generating series c due to the appearance of the shuffle product. Observe the vector fields in the state equation satisfy $\bar{g}_0 = \bar{g}_1$ so that $x_0 \sim x_1$ and $[x_0, x_1] = 0$. Hence, the entropy of the control Lie algebra is still zero as in the original coordinate system.

5   Entropy ultrametric space

Two series $c, d \in \mathbb{R}\langle\langle X\rangle\rangle$ are said to be *entropy equivalent*, denoted by $c \sim_h d$, when $h(c-d) = 0$. Using this equivalence relation, let $A = \mathbb{R}\langle\langle X\rangle\rangle/\sim_h$ denote the quotient space. No notational distinction between elements of $\mathbb{R}\langle\langle X\rangle\rangle$ and A will be made except that $\mathbf{0}$ will denote the equivalence class of zero entropy series in A, and $c = d$ is understood to be $c \sim_h d$ when $c, d \in A$.

Theorem 5.1. *The vector space $A = (A, \mathrm{dist}_h)$ with $\mathrm{dist}_h(c, d) := h(c - d)$ for $c, d \in A$ is a bounded ultrametric space.*

Proof. The only nontrivial property is the ultrametric inequality. Observe that for any $c, d, e \in A$:

$$\mathrm{dist}_h(c, d) = h(c - d) = h((c - e) + (e - d))$$
$$\leq \max(h(c - e), h(e - d)) = \max(\mathrm{dist}_h(c, e), \mathrm{dist}_h(e, d)).$$

The boundedness property is clear.   ∎

Note that dist_h is not induced by a semi-norm as it is not homogeneous, recall $\mathrm{dist}_h(\alpha c, \alpha d) = \mathrm{dist}_h(c, d)$ for all nonzero $\alpha \in \mathbb{R}$.

Example 5.2. In the present context, the series in Example 4.3 satisfy $d_M = \mathbf{0}$, $M \geq 0$. So as a sequence in A, it trivially converges to $\mathbf{0}$ in A.

Example 5.3. The sequence $c_{1/N}$, $N \geq 1$ in Example 3.8 defines a nontrivial sequence in A. Since $\lim_{N\to\infty}\mathrm{dist}_h(c_{1/N}, 1) = \lim_{N\to\infty}h\left(\sum_{\eta \in X^+}\eta^N\right) = \lim_{N\to\infty}1/N = 0$, this sequence also converges to $\mathbf{0}$ in A.

An immediate question is in what sense, if any, is A complete? It is known under various notions of completeness that contractions on ultrametric spaces have a unique fixed point or some approximate fixed point [42, 43]. But at present, this is an open problem.

Let X and Y be two arbitrary alphabets. Any $\mathbb{R}$-linear mapping $\tau\colon \mathbb{R}\langle\langle X\rangle\rangle \to \mathbb{R}\langle\langle Y\rangle\rangle$ is called a *transduction* [15, 31, 35, 44]. It is completely specified by

$$\tau(\eta) = \sum_{\xi \in Y^*}(\tau(\eta), \xi)\xi, \qquad \forall\eta \in X^*,$$

if for each $c \in \mathbb{R}\langle\langle X\rangle\rangle$, the sum $\sum_{\eta \in X^*}(\tau(\eta), \xi)(c, \eta)$ is finite for all $\xi \in Y^*$. Otherwise, it is only partially defined on $\mathbb{R}\langle\langle X\rangle\rangle$. One can canonically associate with any transduction τ a series in $\mathbb{R}\langle\langle X \otimes Y\rangle\rangle$, namely

$$\hat{\tau} = \sum_{\eta \in X^*}\eta \otimes \tau(\eta) = \sum_{\eta \in X^*,\, \xi \in Y^*}(\tau(\eta), \xi)\,\eta \otimes \xi.$$

From $\hat{\tau}$ define a second transduction $\tau' \colon \mathbb{R}\langle\langle Y \rangle\rangle \to \mathbb{R}\langle\langle X \rangle\rangle$ via

$$\tau'(\xi) = \sum_{\eta \in X^*} (\tau(\eta), \xi)\eta, \qquad \forall \xi \in Y^*.$$

τ' is called the *inverse* of τ. A transduction τ is called rational if the series $\hat{\tau}$ is a rational series in $\mathbb{R}\langle\langle X \otimes Y \rangle\rangle$. In which case, every rational series in $\mathbb{R}\langle\langle X \rangle\rangle$ is mapped to a rational series in $\mathbb{R}\langle\langle Y \rangle\rangle$ (assuming it is well-defined).

There are a number of interesting open problems regarding transductions and entropy. Suppose $\mathbb{R}\langle\langle X \rangle\rangle$ and $\mathbb{R}\langle\langle Y \rangle\rangle$ are N_0-graded vector spaces. Then $\mathbb{R}\langle\langle X \rangle\rangle \otimes \mathbb{R}\langle\langle Y \rangle\rangle$ is also an N_0-graded vector space with

$$(\mathbb{R}\langle\langle X \rangle\rangle \otimes \mathbb{R}\langle\langle X \rangle\rangle)_n = \bigoplus_{n=i+j} (\mathbb{R}\langle\langle X \rangle\rangle)_i \otimes (\mathbb{R}\langle\langle Y \rangle\rangle)_j.$$

If, as assumed in Section 3, $(\mathbb{R}\langle\langle X \rangle\rangle)_i$ and $(\mathbb{R}\langle\langle Y \rangle\rangle)_j$ are spanned by all the words of degree i and j, respectively, then likewise for the element $(\mathbb{R}\langle\langle X \rangle\rangle \otimes \mathbb{R}\langle\langle X \rangle\rangle)_n$. Therefore, a given $\hat{\tau}$ has a well defined notion of entropy in terms of its formal power series representation as an element in $\mathbb{R}\langle\langle X \otimes Y \rangle\rangle$. If $c \in \mathbb{R}\langle\langle X \rangle\rangle$ and $d \in \mathbb{R}\langle\langle Y \rangle\rangle$, then the first open question is how are $h(c)$ and $h(\tau(c))$ related, and likewise for $h(d)$ and $h(\tau'(d))$? (Here h refers to the appropriate definition of entropy based on its argument.) In the event that $X = Y$, there are corresponding mappings $\tau_A \colon A \to A$ and $\tau'_A \colon A \to A$ between entropy equivalences classes. Such a mapping is said to be *strictly contracting* if

$$\mathrm{dist}_h(\tau_A(c), \tau_A(d)) < \mathrm{dist}_h(c, d), \qquad \forall c \neq d \in A.$$

A second open question is what notion of completeness is available for A that would render a fixed point, i.e., $\tau_A(c) = c$ and/or $\tau'_A(d) = d$?

6 Conclusions

A definition of entropy was introduced for the generating series of a Chen–Fliess series. The concept is most closely related to the notion that appears in formal language theory. It was shown to be trivial in the case of linear systems, i.e., all linear systems have zero entropy. Formal power series products induced or related to system interconnections were then classified in terms of their ability to increase entropy relative to their arguments. In particular, the shuffle product and all the products that utilize it in their definition can increase entropy. This result will likely have applications in future work regarding networks of Chen–Fliess series. Finally, the paper concluded by introducing an entropy ultrametric space as a possible context for the future analysis of feedback systems.

Acknowledgement

References

[1] Anick D.J., Non-commutative graded algebras and their Hilbert series, *J. Algebra* **78** (1982), 120–140.

[2] Atiyah M.F., MacDonald I.G., Introduction to commutative algebra, Addison–Wesley Publishing Co., Reading, Mass. – London – Don Mills, Ont., 1969.

[3] Berstel J., Reutenauer C., Rational series and their languages, *EATCS Monogr. Theoret. Comput. Sci.*, Vol. 12, Springer-Verlag, Berlin, 1988.

[4] Caracciolo S., Radicati L.A., Entropy ultrametric for dynamical and disordered systems, *J. Physique* **50** (1989), 2919–2930.

[5] Devlin J., Word problems related to periodic solutions of a nonautonomous system, *Math. Proc. Cambridge Philos. Soc.* **108** (1990), 127–151.

[6] Devlin J., Word problems related to derivatives of the displacement map, *Math. Proc. Cambridge Philos. Soc.* **110** (1991), 569–579.

[7] Downarowicz T., Entropy in dynamical systems, *New Math. Monogr.*, Vol. 18, Cambridge University Press, Cambridge, 2011.

[8] Duffaut Espinosa L.A., Interconnections of nonlinear systems driven by L_2-Itô stochastic process, Ph.D. Thesis, Old Dominion University, 2009, https://doi.org/10.25777/cngz-5y11.

[9] Duffaut Espinosa L.A., Ebrahimi-Fard K., Gray W.S., A combinatorial Hopf algebra for nonlinear output feedback control systems, *J. Algebra* **453** (2016), 609–643, arXiv:1406.5396.

[10] Duffaut Espinosa L.A., Gray W.S., González O.R., On Fliess operators driven by L_2-Itô random processes, in Proc. 48th IEEE Conf. on Decision and Control, IEEE, Shanghai, China, 2009, 7478–7484.

[11] Ebrahimi-Fard K., Gray W.S., Center problem, Abel equation and the Faà di Bruno Hopf algebra for output feedback, *Int. Math. Res. Not.* **2017** (2017), 5415–5450, arXiv:1507.06939.

[12] Ferfera A., Combinatoire du monoïde libre appliquée à la composition et aux variations de certaines fonctionnelles issues de la théorie des systèmes, Ph.D. Thesis, University of Bordeaux I, 1979.

[13] Ferfera A., Combinatoire du monoïde libre et composition de certains systèmes non linéaires, *Astérisque* **75–76** (1980), 87–93, available at http://www.numdam.org/item/AST_1980__75-76__87_0.

[14] Fliess M., Sur divers produits de séries formelles, *Bull. Soc. Math. France* **102** (1974), 181–191.

[15] Fliess M., Transductions de séries formelles, *Discrete Math.* **10** (1974), 57–74.

[16] Fliess M., Fonctionnelles causales non linéaires et indéterminées non commutatives, *Bull. Soc. Math. France* **109** (1981), 3–40.

[17] Fliess M., Réalisation locale des systèmes non linéaires, algèbres de Lie filtrées transitives et séries génératrices non commutatives, *Invent. Math.* **71** (1983), 521–537.

[18] Foissy L., The Hopf algebra of Fliess operators and its dual pre-Lie algebra, *Comm. Algebra* **43** (2015), 4528–4552, arXiv:1304.1726.

[19] Gray W.S., Formal power series methods in nonlinear control theory, unpublished, 1.2 ed., 2022, available at http://www.ece.odu.edu/~sgray/fps-book.

[20] Gray W.S., System identification entropy for Chen–Fliess series and their interconnections, in Proc. 2022 Allerton Conf. on Communication, Control, & Computing, Allerton Park, Illinois, to appear.

[21] Gray W.S., Duffaut Espinosa L.A., Ebrahimi-Fard K., Faà di Bruno Hopf algebra of the output feedback group for multivariable Fliess operators, *Systems Control Lett.* **74** (2014), 64–73, arXiv:1406.5378.

[22] Gray W.S., Ebrahimi-Fard K., SISO output affine feedback transformation group and its Faà di Bruno Hopf algebra, *SIAM J. Control Optim.* **55** (2017), 885–912, arXiv:1411.0222.

[23] Gray W.S., Ebrahimi-Fard K., Generating series for networks of Chen–Fliess series, *Systems Control Lett.* **147** (2021), 104827, 8 pages, arXiv:2007.00743.

[24] Gray W.S., Herencia-Zapana H., Duffaut Espinosa L.A., González O.R., Bilinear system interconnections and generating series of weighted Petri nets, *Systems Control Lett.* **58** (2009), 841–848.

[25] Gray W.S., Li Y., Generating series for interconnected analytic nonlinear systems, *SIAM J. Control Optim.* **44** (2005), 646–672.

[26] Gray W.S., Thitsa M., A unified approach to generating series for mixed cascades of analytic nonlinear input-output systems, *Internat. J. Control* **85** (2012), 1737–1754.

[27] Gray W.S., Venkatesh G.S., Relative degree of interconnected SISO nonlinear control systems, *Systems Control Lett.* **124** (2019), 99–105.

[28] Gray W.S., Wang Y., Fliess operators on L_p spaces: convergence and continuity, *Systems Control Lett.* **46** (2002), 67–74.

[29] Gray W.S., Wang Y., Formal Fliess operators with applications to feedback interconnections, in Proc. 18th Inter. Symp. on the Mathematical Theory of Networks and Systems, Blacksburg, Virginia, 2008, 12 pages.

[30] Isidori A., Nonlinear control systems, 3rd ed., *Comm. Control Engrg. Ser.*, Springer-Verlag, Berlin, 1995.

[31] Jacob G., Sur un théorème de Shamir, *Inf. Control* **27** (1975), 218–261.

[32] Kawan C., Delvenne J.-C., Network entropy and data rates required for networked control, *IEEE Trans. Control Netw. Syst.* **3** (2016), 57–66, arXiv:1409.6037.

[33] Kuich W., On the entropy of context-free languages, *Inf. Control* **16** (1970), 173–200.

[34] Kuich W., Maurer H., The structure generating function and entropy of tuple languages, *Inf. Control* **19** (1971), 195–203.

[35] Kuich W., Salomaa A., Semirings, automata, languages, *EATCS Monogr. Theoret. Comput. Sci.*, Vol. 5, Springer-Verlag, Berlin, 1986.

[36] Liberzon D., On topological entropy of interconnected nonlinear systems, *IEEE Control Syst. Lett.* **5** (2021), 2210–2214.

[37] Lind D., Marcus B., An introduction to symbolic dynamics and coding, 2nd ed., *Cambridge Math. Lib.*, Cambridge University Press, Cambridge, 2021.

[38] Lothaire M., Combinatorics on words, 2nd ed., *Cambridge Math. Lib.*, Cambridge University Press, Cambridge, 1997.

[39] Matveev A.S., Proskurnikov A.V., Pogromsky A., Fridman E., Comprehending complexity: data-rate constraints in large-scale networks, *IEEE Trans. Automat. Control* **64** (2019), 4252–4259.

[40] McLachlan R.I., Ryland B., The algebraic entropy of classical mechanics, *J. Math. Phys.* **44** (2003), 3071–3087, arXiv:math-ph/0210030.

[41] Newman M.F., Schneider C., Shalev A., The entropy of graded algebras, *J. Algebra* **223** (2000), 85–100.

[42] Priess-Crampe S., Ribenboim P., Ultrametric dynamics, *Illinois J. Math.* **55** (2011), 287–303.

[43] Priess-Crampe S., Ribenboim P., The approximation to a fixed point, *J. Fixed Point Theory Appl.* **14** (2013), 41–53, arXiv:1307.6431.

[44] Salomaa A., Soittola M., Automata-theoretic aspects of formal power series, *Monogr. Comput. Sci.*, Springer-Verlag, New York – Heidelberg, 1978.

[45] Savkin A.V., Analysis and synthesis of networked control systems: topological entropy, observability, robustness and optimal control, *Automatica* **42** (2006), 51–62.

[46] Schneider F.M., Borchmann D., Topological entropy of formal languages, *Semigroup Forum* **94** (2017), 556–581, arXiv:1507.03393.

[47] Shannon C.E., A mathematical theory of communication, *Bell System Tech. J.* **27** (1948), 379–423.

[48] Sloane N.J.A., The on-line encyclopedia of integer sequences, available at https://oeis.org.

[49] Smith C., Enumeration of the distinct shuffles of permutations, Ph.D. Thesis, Harvard University, 2009, available at https://www.proquest.com/docview/304890446.

[50] Thitsa M., Gray W.S., On the radius of convergence of interconnected analytic nonlinear input-output systems, *SIAM J. Control Optim.* **50** (2012), 2786–2813.

[51] Tomar M.S., Zamani M., Compositional quantification of invariance feedback entropy for networks of uncertain control systems, *IEEE Control Syst. Lett.* **4** (2020), 827–832.

[52] Venkatesh G.S., Wiener–Fliess composition of formal power series: additive static feedback and shuffle rational series, Ph.D. Thesis, Old Dominion University, 2021, https://doi.org/10.25777/t2b1-dx91.

[53] Venkatesh G.S., Gray W.S., Formal power series approach to nonlinear systems with static output feedback, *Internat. J. Control*, to appear, arXiv:2110.10034.

[54] Wang Y., Algebraic differential equations and nonlinear control systems, Ph.D. Thesis, Rutgers University, 1990, available at https://www.proquest.com/docview/303872790.

[55] Winter-Arboleda I.M., Gray W.S., Duffaut Espinosa L.A., Fractional Fliess operators: two approaches, in Proc. 49th Conf. on Information Sciences and Systems, IEEE, Baltimore, Maryland, 2015, 6 pages.

[56] Young L.-S., Entropy in dynamical systems, in Entropy, *Princeton Ser. Appl. Math.*, Princeton University Press, Princeton, New Jersey, 2003, 313–327.

[57] Zames G., On the metric complexity of causal linear systems: ε-entropy and ε-dimension for continuous time, *IEEE Trans. Automat. Control* **24** (1979), 222–230.