A noncommutative extension of Mahler's interpolation theorem

Jean-Éric Pin, Christophe Reutenauer

Contents

A NONCOMMUTATIVE EXTENSION OF MAHLER'S INTERPOLATION THEOREM

June 28, 2021

JEAN-ÉRIC PIN AND CHRISTOPHE REUTENAUER

Abstract. We prove a noncommutative generalisation of Mahler's theorem on interpolation series, a celebrated result of p-adic analysis. Mahler's original result states that a function from $\mathbb{N}$ to $\mathbb{Z}$ is uniformly continuous for the p-adic metric d_p if and only if it can be uniformly approximated by polynomial functions. We prove an analogous result for functions from a free monoid A^* to a free group $F(B)$, where d_p is replaced by the pro-p metric.

1. Introduction

The aim of this paper is to give a noncommutative version of Mahler's theorem on interpolation series [8]. This new version, which applies to functions from a free monoid A^* to a free group $F(B)$, extends a previous extension, due to Silva and the first author [15], for functions from A^* to $\mathbb{Z}$. Several results of our new article were announced in the conference paper [11], in which most proofs were either omitted or just sketched out.

Throughout this paper, p denotes a prime number.

1.1. Original motivation

Our original motivation, described in more details in Section 9.2, seems at first sight to have nothing to do with Mahler's theorem. It is inspired by automata theoretic questions, see [12, 14] for more details. Recall that a subset L of A^* (also called a *language*) is recognized by a monoid M if there exist a monoid morphism $\varphi: A^* \to M$ such that $L = \varphi^{-1}(\varphi(L))$.

Following Eilenberg [6], let $\mathcal{G}_p$ denote the class of all languages recognised by a finite p-group. An elegant description of these languages was given by Eilenberg (see Theorem 9.2) using a noncommutative extension of the binomial coefficients, described in Section 2.1. Our original motivation was to obtain a satisfactory description of the functions $f: A^* \to B^*$ such that, for each language L in $\mathcal{G}_p$, the language $f^{-1}(L)$ is also in $\mathcal{G}_p$.

The connection with Mahler's theorem stems from the fact that these functions are exactly the uniformly continuous functions, when A^* and B^*

2020 *Mathematics Subject Classification.* 68R15 68Q70 20M35 22A20 11S80 54E15 20E18.

Key words and phrases. noncommutative algebra, free group, free monoid, Magnus transformation, subword functions, sequential functions, noncommutative polynomial functions, p-groups, noncommutative interpolation, Mahler's interpolation theorem, p-adic, difference operator, forward difference formula, combinatorics on words.

The first author was supported by the DeLTA project (ANR-16-CE40-0007) and the second author by NSERC (Canada).

are equipped with the pro-p metric, defined in Section 6.4. When $|A| = |B| = 1$, A^* and B^* are isomorphic to the additive monoid $\mathbb{N}$, the pro-p metric is the p-adic metric and our problem amounts to describe the functions from $\mathbb{N}$ to $\mathbb{N}$ which are uniformly continuous for the p-adic metric. As we will see in the next section, this is precisely the goal of Mahler's theorem. To return to the general case, it was therefore natural to look for a noncommutative version of this theorem

1.2. **Mahler's interpolation theorem**

Mahler's interpolation theorem [8] is usually stated for functions of p-adic numbers, but this full version can be easily recovered from the simpler version given in Theorem 1.1 below. Recall that the *difference operator* Δ associates to each function $f:\mathbb{N} \to \mathbb{Z}$, the function $\Delta f:\mathbb{N} \to \mathbb{Z}$ defined by $(\Delta f)(n) = f(n+1) - f(n)$. Let Δ^k be the k-th iteration of Δ. Setting $\delta_k f = (\Delta^k f)(0)$ for every nonnegative integer k and $f_r(n) = \sum_{k=0}^{r} \binom{n}{k} \delta_k f$, Mahler's theorem can be stated as follows:

Theorem 1.1 (Mahler)**.** *Let $f:\mathbb{N} \to \mathbb{Z}$ be a function. The following conditions are equivalent:*

(1) *f is uniformly continuous for the p-adic metric,*
(2) *the functions $\Delta^r f$ tend uniformly to 0 when r tends to ∞,*
(3) *the p-adic norm of $\delta_r f$ tends to 0 when r tends to ∞,*
(4) *f is the uniform limit of the functions f_r when r tends to ∞.*

Just to clarify, $\mathbb{N}$ and $\mathbb{Z}$ are equipped in this statement with the p-adic metric and the uniformity used in conditions (2) and (4) is that of uniform convergence on the space of functions from $\mathbb{N}$ to $\mathbb{Z}$, described in more details in Propositions 6.9 and 6.10.

Mahler's theorem is based on another result of independent interest. *Newton's forward difference formula* states that, for all natural numbers n, $f(n) = \sum_{k=0}^{\infty} \binom{n}{k} \delta_k f$, a sum which is finite for each given n. A remarkable consequence of this formula is that the map $f \to (\delta_k f)_{k \geq 0}$ defines a *bijection* between functions from $\mathbb{N}$ to $\mathbb{Z}$ and integer sequences. We call this bijection the *Newton bijection*.

1.3. **A noncommutative extension**

Our noncommutative extension concerns functions from a free monoid A^* to a free group $F(B)$. Of course, if B is a one-letter alphabet, then $F(B)$ is isomorphic to $\mathbb{Z}$ and one recovers the result of [15]. If, in addition, A is a one-letter alphabet, then A^* is isomorphic to $\mathbb{N}$, and one gets back Mahler's original theorem.

We equip both A^* and $F(B)$ with the pro-p metric, a natural extension of the p-adic metric. A noncommutative version of Newton's forward difference formula and of Newton's bijection was given by the first author in [10]. We give a simpler proof of these results in Section 3. In this noncommutative setting, f is a function from A^* to a group G. For each letter a of A, the *difference operator* Δ^a associates to each function $f:A^* \to G$ the function $\Delta^a f:A^* \to G$ defined by $\Delta^a f(u) = f(u)^{-1} f(ua)$. Next we attach a difference operator Δ^w to each word $w = a_1 \cdots a_n$ of A^* by setting

$\Delta^w f = \Delta^{a_1}(\Delta^{a_2}(\cdots\Delta^{a_n} f)\cdots))$. Setting $\delta_w f = \Delta^w f(1)$, where 1 is the empty word of A^*, the Newton bijection is now the map $f \to (\delta_w f)_{w\in A^*}$. If we just keep the elements $\delta_w f$ such that $|w| \leqslant r$ and replace every other $\delta_w f$ by the identity of G, the inverse of Newton's bijection gives back a function f_r, called the r-th *Newton polynomial function associated to* f.

Our main result offers a noticeable analogy with Mahler's theorem:

Theorem 1.2. *Let $f: A^* \to F(B)$ be a function. The following conditions are equivalent:*

(1) *f is uniformly continuous for the pro-p metric,*
(2) *the functions $\Delta^w f$, where $w \in A^*$, tend uniformly to 1 when $|w|$ tends to ∞,*
(3) *the elements $\delta_w f$, where $w \in A^*$, tend to 1 when $|w|$ tends to ∞,*
(4) *f is the uniform limit of its Newton polynomial functions f_r when r tends to ∞.*

In addition to this theorem, we prove several other results of interest. The first one is a solution to the following question:

Integration problem. *Given an element g of G and a family $(f_a)_{a\in A}$ of functions from A^* to G, find a function f such that $f(1) = g$ and $f_a = \Delta^a f$ for all $a \in A$.*

We show that the integration problem has a unique solution $\mathrm{Seq}(g, (f_a)_{a\in A})$, called the *sequential product* at g of the family $(f_a)_{a\in A}$.

Let us call a function f from $\mathbb{N}$ to $\mathbb{Z}$ a *Newton polynomial function* if $\Delta^k f = 0$ for almost all[1] k. In particular, all polynomial functions are Newton polynomial functions, but the function $n \to \binom{n}{2}$ is also a Newton polynomial function. It is natural to extend this definition as follows:

Definition. *A function $f: A^* \to G$ is a* Newton polynomial function[2] *if $\Delta^w f = \mathbf{1}$ for almost all words $w \in A^*$. In this case, the* degree *of f is the smallest d such that $\Delta^w f = \mathbf{1}$ for all words w of length $> d$.*

In particular, the function f_r introduced earlier is a Newton polynomial function of degree at most r. We show (Proposition 4.2) that a function f is a Newton polynomial function of degre $\leqslant d$ if and only if $\delta_w f = 1$ for all words w of length $d+1$. We also show (Corollary 4.6) that the set of Newton polynomial functions is the smallest set of functions containing the constant functions and closed under sequential product.

1.4. Proof techniques and notation

Our proof techniques are a mixture of algebra, combinatorics and topology. The combinatorial aspects occur already in Section 2.1, where the noncommutative extension of binomial coefficients is introduced, and in Section 3.1, where we define a noncommutative extension of the Magnus transformation (see in particular Propositions 3.1 and 7.5). Algebraic arguments appear in Proposition 6.6 and form the core of Sections 7.3 and 7.4. The topological

[1] Following a standard terminology, we use "almost all" to mean "all but finitely many".

[2] They were called *Mahler polynomial functions* in [15] but *Newton polynomial function* seems to be more appropriate.

aspects are introduced in Section 6. We preferred to place ourselves within the framework of uniform spaces for two reasons: first, it leads to more concise proofs; secondly, it makes it easier to understand when it is mandatory to use a finite alphabet. We come back to metric spaces in the last three subsections of Section 7.

Two applications of our main result are discussed in Section 9. We first study an interpolation problem in the spirit of Mahler's original paper [8] and then come back to our original motivation related to language theory.

We would like to warn the reader of a notation that could lead to confusion. Indeed, starting from Section 3, we use an additive notation for a noncommutative operation. This is not in itself a novelty and is even a standard notation for the sum of ordinals. For our part, we were inspired by Banaschewski and Nelson [1], who use "+" in exactly the same case as we do. Nevertheless, we have sought to replace "+" with another symbol, such as "$\curlywedge$", but we have not found a substitute for $-$ and $\pm$. As this additive notation leads to synthetic formulas, such as 3.2 and 3.3, we finally decided to keep it, while frequently recalling its non-commutative character.

1.5. **Structure of the paper**

Our paper is organised as follows. Basic prerequisites are recalled in Section 2. Newton's Forward Difference Formula is introduced in Section 3 and Newton polynomial functions in Section 4. Newton's bijection is the topic of Section 5. General topological aspects are covered in Section 6 and the special case of free monoids and free groups is treated in Section 7. The proof of our main result is given in Section 8 and applications are presented in Section 9. The article is completed by a short appendix on uniform structures.

2. Prerequisites

As usual, $[n]$ denotes the set $\{1, \dots, n\}$ and $|E|$ the cardinality of a set E.

2.1. **Words and subwords**

Let A be a set called an *alphabet*, whose elements are called *letters*. A *word* on A is a finite sequence of elements of letters, denoted by mere juxtaposition $a_1 \cdots a_n$. If $u = a_1 \cdots a_n$ is a word, then n is the *length* of u and is denoted by $|u|$. The set of words of length n is denoted by A^n.

We let A^* denote the set of words on A. It is a monoid for the *concatenation product*, which associates with two words $u = a_1 \cdots a_p$ and $v = b_1 \cdots b_q$ the word $uv = a_1 \cdots a_p b_1 \cdots b_q$. This product has an identity, the *empty word*, denoted by 1 or by ε when 1 already denotes a letter of the alphabet, as in Example 2.1 below. The empty word is the unique word of length 0. The monoid A^* is actually the *free monoid* on A.

A word $u = a_1 a_2 \cdots a_n$ is a *subword* of a word v if v can be written as $v = v_0 a_1 v_1 \cdots a_n v_n$ for some (possibly empty) words $v_0, v_1, \dots, v_n$. For instance, *aba* is a subword of *caccbca*.

Let u and v be words. Following Eilenberg [6] and Lothaire [7, Chapter 6], let $\binom{v}{u}$ denote the number of distinct ways to write u as a subword of v.

More formally, if $u = a_1a_2\cdots a_n$, then

$$\binom{v}{u} = \mathrm{Card}\{(v_0, v_1, \ldots, v_n) \in (A^*)^{n+1} \mid v_0a_1v_1\cdots a_nv_n = v\}$$

Observe that, if $u = a^n$ and $v = a^m$, then $\binom{v}{u} = \binom{n}{m}$ and hence the numbers $\binom{v}{u}$ constitute a generalization of the classical binomial coefficients. We refer the reader to [7, Chapter 6] for more information on these extended binomial coefficients.

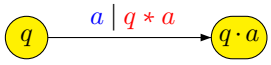
2.2. **Sequential transducers**

We refer the reader to the survey article [5] for more information on sequential transducers. However, we follow Sakarovitch's suggestion [20, p. 651] and use the term *pure sequential* instead of *sequential* and *sequential* instead of *subsequential*.

A *sequential transducer* is an 8-tuple $\mathcal{T} = (Q, A, M, q_0, \cdot, *, m, \rho)$, where Q is the set of *states*, A is a finite alphabet called the *input alphabet*, M is a monoid called the *output monoid*, $q_0 \in Q$ is the *initial state*, the functions $(q, a) \mapsto q\cdot a \in Q$ and $(q, a) \mapsto q * a \in M$ are respectively the *transition function* and the *output function*, $m \in M$ is the *initial prefix* and $\rho\colon Q \to M$ is a function, called the *terminal function*. It is a *pure sequential transducer* if $m = 1$ and $\rho(q) = 1$ for all $q \in Q$. The transducer is called *finite* when Q is finite.

It is convenient to represent a sequential transducer by a labelled graph whose vertices are the states of the transducer. The initial state and the initial prefix are pictured by an incoming arrow, the terminal function by an outcoming arrow, as follows:

For a pure sequential tranducer, we simply give the initial state and ignore the initial prefix and the terminal function. We also represent simultaneously the transition $q \to q\cdot a$ and the output $q * a$ in the following form, where the vertical bar is a separator:



The transition and the output functions can be extended to functions $Q \times A^* \to Q$ (resp. $Q \times A^* \to M$) by setting, for each $u \in A^*$ and each $a \in A$:

$$q\cdot 1 = q \qquad q * 1 = 1$$
$$q\cdot(ua) = (q\cdot u)\cdot a$$
$$q * (ua) = (q * u)((q\cdot u) * a)$$

We also fix some priority rules on the operators. The standard choice is to give highest priority to concatenation, then to "$\cdot$" and then to "$*$". For instance, we write $q\cdot ua$ for $q\cdot(ua)$, $q*ua$ for $q*(ua)$ and $q\cdot u*a$ for $(q\cdot u)*a$.

The function $f\colon A^* \to M$ *realized* by $\mathcal{T}$ is defined by

$$f(u) = m(q_0 * u)\rho(q_0\cdot u)$$

or, when $\mathcal{T}$ is a pure sequential transducer, by

$$f(u) = q_0 * u$$

A *sequential function* is a function that can be realized by a finite sequential transducer.

Example 2.1. For a word $u \in \{0,1,2\}^*$, let $\overline{u}$ denote the nonnegative integer represented by u in base 3. Let $f\colon\{0,1,2\}^* \to \{0,1,2\}^*$ be the Euclidean division by 2 in base 3, that is, the function which associates to a word $u \in \{0,1,2\}^*$ representing an integer n in base 3, the unique word v of the same length as u representing the quotient of the division of n by 2 (in base 3). For instance, $f(1212) = 0221$ since $\overline{1212} = 50$ and $\overline{0221} = 25 = 50/2$. It can also be defined recursively as follows:

$$f(\varepsilon) = \varepsilon$$

$$f(u0) = \begin{cases} f(u)0 & \text{if } \overline{u} \text{ is even} \\ f(u)1 & \text{if } \overline{u} \text{ is odd} \end{cases}$$

$$f(u1) = \begin{cases} f(u)0 & \text{if } \overline{u} \text{ is even} \\ f(u)2 & \text{if } \overline{u} \text{ is odd} \end{cases}$$

$$f(u2) = \begin{cases} f(u)1 & \text{if } \overline{u} \text{ is even} \\ f(u)2 & \text{if } \overline{u} \text{ is odd} \end{cases}$$

As stated in [18], the function f is sequential. Indeed, it is realized by the finite pure sequential transducer represented in Figure 1.

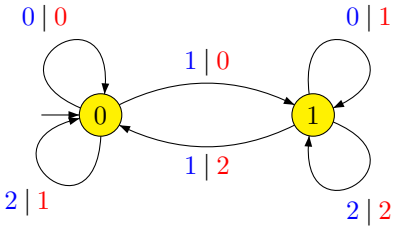
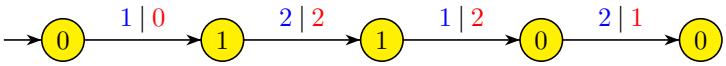


FIGURE 1. Euclidean division by 2 of integers in base 3.

For instance, on the input 1212, the output is 0221, as shown in the figure below:



3. Newton's Forward Difference Formula

Newton's forward difference formula gives an expression of a function in terms of the initial value of the function and the powers of the forward difference operator. The simplest version of this formula states that for each function f from $\mathbb{N}$ to $\mathbb{Z}$,

$$f(n) = \sum_{k=0}^{\infty} \binom{n}{k} \delta_k f, \tag{3.1}$$

a formula which is also the starting point of Mahler's article [8, Theorem 1].

A noncommutative extension of (3.1) for the functions from A^* to $\mathbb{Z}$ was given in [15, Theorem 2.2] and a further extension for the functions from A^* to $F(B)$ was proposed in [10]. In this section, we give a simpler presentation of these results for the functions from A^* to a group G, a slightly more general setting.

Our noncommutative version of Newton's forward difference formula allows to retrieve the function f from the family $(\delta_w f)_{w\in A^*}$. Its precise statement, as given in Theorem 3.6 below, requires some auxiliary definitions and results, as could be expected in this non-commutative framework. To this end, we introduce a noncommutative extension of the Magnus transformation in Section 3.1 and then study the operators Δ^w in more detail in Section 3.2. Section 3.3 is devoted to the proof of Theorem 3.6.

3.1. **Noncommutative Magnus transformation**

Let A^{**} denote the free monoid freely generated by A^*. An element of A^{**} is a finite sequence $(x_1,\dots,x_n)$ of elements of A^*. However, to avoid any confusion between the product in A^* and the product in A^{**}, we adopt an additive notation for A^{**}, although A^{**} is in general noncommutative. This means that we replace the notation $(x_1,\dots,x_n)$ by $x_1+\cdots+x_n$. The addition of two elements $(u_1+\cdots+u_m)$ and $(v_1+\cdots+v_n)$ of A^{**} is also denoted additively, which is coherent, since

$$(u_1+\cdots+u_m)+(v_1+\cdots+v_n)=u_1+\cdots+u_m+v_1+\cdots+v_n.$$

Accordingly, the neutral element of the monoid A^{**} is denoted 0.

For each $u\in A^*$ and $x=x_1+\cdots+x_n\in A^{**}$, let us set

$$x\cdot u=x_1u+\cdots+x_nu. \tag{3.2}$$

We let the reader verify that this defines a *monoid right action* of A^* on A^{**}, which means that the following formulas hold for all $u,u_1,u_2\in A^*$, and for all $x,x_1,x_2\in A^{**}$,

$$\begin{aligned} 0\cdot u&=0\\ (x_1+x_2)\cdot u&=x_1\cdot u+x_2\cdot u\\ x\cdot(u_1u_2)&=(x\cdot u_1)\cdot u_2. \end{aligned}$$

The *noncommutative Magnus transformation* is the mapping μ from A^* into A^{**} defined recursively by setting $\mu(1)=1$ and, for all $w\in A^*$ and $a\in A$,

$$\mu(wa)=\mu(w)+\mu(w)\cdot a. \tag{3.3}$$

Example 3.1. Let $a,b,c,d\in A$. Then

$$\begin{aligned} \mu(a)&=1+a,\\ \mu(ab)&=1+a+b+ab,\\ \mu(abc)&=1+a+b+ab+c+ac+bc+abc,\\ \mu(abcd)&=1+a+b+ab+c+ac+bc+abc+d+ad+bd\\ &\quad+abd+cd+acd+bcd+abcd,\\ \mu(aba)&=1+a+b+ab+a+aa+ba+aba,\\ \mu(abab)&=1+a+b+ab+a+aa+ba+aba+b+ab+bb \end{aligned}$$

$$+ abb + ab + aab + bab + abab. \tag{3.4}$$

Warning. It is tempting to define directly, instead of the right action defined by 3.2, a product on A^{**} given, using the same notation, by the formula

$$(u_1 + \cdots + u_m)(v_1 + \cdots + v_n) = u_1v_1 + \cdots + u_mv_1 + \cdots + u_1v_n + \cdots + u_mv_n \tag{3.5}$$

and then simply write $\mu(a_1a_2\cdots a_n) = (1+a_1)(1+a_2)\cdots(1+a_n)$. This approach using *near-rings* is possible and was used in [10], but it requires special care. Indeed, not only the addition is not commutative, but multiplication only distributes on the left over addition, and not on the right. For instance, $(1+a)(1+b) = (1+a)+(1+a)b = 1+a+b+ab$ is different from $(1+b)+a(1+b) = 1+b+a+ab$.

The function μ is an extension of the classical Magnus transformation M, which is a morphism from the free monoid A^* (and more generally the free group $F(A)$) into the multiplicative monoid of the ring $\mathbb{Z}\langle\langle A\rangle\rangle$ of noncommutative formal power series: it maps each letter a onto $1+a$. For example, since the addition in $\mathbb{Z}\langle\langle A\rangle\rangle$ is commutative, one has

$$\begin{aligned} M(abab) &= (1+a)(1+b)(1+a)(1+b) \\ &= 1 + 2a + 2b + aa + 3ab + ba + bb \\ &\quad + aab + aba + abb + bab + abab, \end{aligned}$$

a formula to be compared with (3.4).

Here is a simple algorithm to obtain $\mu(abcd)$, suggested by Mathieu Guay-Pacquet:

(1) Write $abcd$ backwards to get $dcba$.
(2) Order the subwords of $dcba$ as in a dictionary to obtain the list
$$L = 1, a, b, ba, c, ca, cb, cba, d, da, db, dba, dc, dca, dcb, dcba.$$
(3) Write the words of L backwards to get the list
$$\tilde{L} = 1, a, b, ab, c, ac, bc, abc, d, ad, bd, abd, cd, acd, bcd, abcd.$$

Then $\mu(abcd)$ is the ordered sum of the elements of $\tilde{L}$. To get $\mu(abab)$, it now suffices to replace c by a and d by b in the expression giving $\mu(abcd)$.

This algorithm can be justified as follows. Let $\mathbb{P} = \mathbb{N} - \{0\}$ be the set of positive integers. Define recursively a total order on the set of finite subsets of $\mathbb{P}$ as follows:

(1) for every nonempty finite subset I of $\mathbb{P}$, set $\varnothing < I$;
(2) if I, J are two nonempty subsets of $\mathbb{P}$, then $I < J$ if either $\max(I) < \max(J)$ (for the usual order of natural numbers), or if $\max(I) = \max(J)$ and $I - \{\max(I)\} < J - \{\max(J)\}$.

Example 3.2. One has $\{4,7\} < \{3,4,7\} < \{5,7\}$ and, representing subsets of $\mathbb{P}$ without braces,

$$\varnothing < 1 < 2 < 12 < 3 < 13 < 23 < 123 < 4 < 14 < 24 < 124 < 34 < 134 < 234 < 1234.$$

If $w = a_1\cdots a_n$ and $I = \{i_1 < \cdots < i_k\} \subseteq [n]$, let $w[I]$ denote the word $a_{i_1}\cdots a_{i_k}$. Then the following result holds.

Proposition 3.1. *For each $w \in A^*$,*

$$\mu(w) = \sum_{I \subseteq [|w|]} w[I], \tag{3.6}$$

where the (noncommutative) sum runs, from left to right, over all subsets of $[|w|]$ increasingly ordered by $<$.

Proof. Recall that μ is defined recursively by $\mu(1) = 1$ and by the functional equation (3.3). Therefore, it suffices to show that the function $\nu(w) = \sum_{I \subseteq [|w|]} w[I]$ satisfies the same equations. Since the equality $\nu(1) = w[\varnothing] = 1$ is immediate, it just remains to prove that, for each letter $a \in A$,

$$\nu(wa) = \nu(w) + \nu(w) \cdot a. \tag{3.7}$$

First observe that a subset of $[|w|+1]$ is either a subset of $[|w|]$ or contains $|w|+1$, and every subset of the first category is lower (for the order $<$) than every subset of the second category. Moreover, if $I < J$ are subsets of the second category, then, by definition of the order, $I - \{|w|+1\} < J - \{|w|+1\}$. Since, for I in the second category, $wa[I] = w[I - \{|w|+1\}]a$, one gets

$$\nu(wa) = \sum_{I \subseteq [|w|+1]} wa[I] = \sum_{I \subseteq [|w|]} w[I] + \Big(\sum_{I \subseteq [|w|]} w[I] \Big) \cdot a$$

which proves (3.7). □

3.2. **Difference operators**

Let G be a group and let $f : A^* \to G$ be a function. Following [10], we define the difference operators as follows. For each letter a of A, let $\Delta^a f$ denote the function $A^* \to G$ defined by

$$\Delta^a f(w) = f(w)^{-1} f(wa)$$

for each word w in A^*. We obtain in this way a function $a \mapsto \Delta^a$ from A into the set $\mathcal{M}$ of all mappings from G^{A^*} into itself. We view $\mathcal{M}$ as a monoid under the composition of mappings. Since A^* is the free monoid on A, this function from A to $\mathcal{M}$ extends uniquely to a monoid morphism from A^* into $\mathcal{M}$. Denoting $w \mapsto \Delta^w$ this extension, we get $\Delta^1 f = f$ and, for all words u, v in A^*,

$$\Delta^{uv} f = \Delta^u \Delta^v f. \tag{3.8}$$

Example 3.3. For instance, one gets

$$\begin{aligned}
(\Delta^1 f)(u) &= f(u) \\
(\Delta^a f)(u) &= f(u)^{-1} f(ua) \\
(\Delta^{ab} f)(u) &= f(ub)^{-1} f(u) f(ua)^{-1} f(uab) \\
(\Delta^{abc} f)(u) &= f(ubc)^{-1} f(ub) f(u)^{-1} f(uc) f(uac)^{-1} f(ua) f(uab)^{-1} f(uabc) \\
(\Delta^{abcd} f)(u) &= f(ubcd)^{-1} f(ubc) f(ub)^{-1} f(ubd) f(ud)^{-1} f(u) f(uc)^{-1} f(ucd) \\
&\quad f(uacd)^{-1} f(uac) f(ua)^{-1} f(uad) f(uabd)^{-1} f(uab) f(uabc)^{-1} \\
&\quad f(uabcd) \\
(\Delta^{abab} f)(u) &= f(ubab)^{-1} f(uba) f(ub)^{-1} f(ubb) f(ub)^{-1} f(u) f(ua)^{-1} f(uab) \\
&\quad f(uaab)^{-1} f(uaa) f(ua)^{-1} f(uab) f(uabb)^{-1} f(uab) f(uaba)^{-1}
\end{aligned}$$

$$f(uabab) \tag{3.9}$$

Let us return for a moment to the commutative case, where f is a function from $\mathbb{N}$ to a commutative group $(G,+)$. To do this, we take a one-letter alphabet $A = \{a\}$ and we use the map $a^n \to n$ to identify A^* to $(\mathbb{N},+)$. Writing $\Delta^n f$ for $\Delta^{a^n} f$ and using an additive notation, we get

$$\Delta^0 f(n) = f(n)$$
$$\Delta^1 f(n) = -f(n) + f(n+1),$$

and more generally,

$$\Delta^k f(n) = f(n+k) - \binom{n}{1} f(n+k-1) + \binom{n}{2} f(n+k-2) - \cdots + (-1)^k \binom{n}{k} f(n),$$

which is the standard expression of the k-th power of the difference operator.

The general formula to retrieve the results of Example 3.3 requires some further development and will be presented at the end of Section 3.3.

Difference operators commute with group morphisms, in the following sense:

Proposition 3.2. *Let $f: A^* \to G$ be a function, let $\varphi: G \to H$ be a group morphism and let w be a word. Then*

$$\Delta^w(\varphi \circ f) = \varphi \circ (\Delta^w f). \tag{3.10}$$

Proof. We prove the result by induction on $|w|$. The result is trivial if w is the empty word. If $w = a$ for some letter a, one gets

$$\Delta^a(\varphi \circ f)(x) = (\varphi \circ f(x))^{-1}(\varphi \circ f(xa)) = \varphi(f(x))^{-1}\varphi(f(xa))$$
$$= \varphi(f(x)^{-1} f(xa)) = \varphi(\Delta^a f(x)) = \varphi \circ (\Delta^a f)(x)$$

and thus $\Delta^a(\varphi \circ f) = \varphi \circ (\Delta^a f)$.

If $|w| \geqslant 2$, then $w = ua$ for some word u of length $|w|-1$ and some letter a. Then by (3.8) and by the induction hypothesis applied to u, one gets

$$\Delta^w(\varphi \circ f) = \Delta^u \Delta^a(\varphi \circ f) = \Delta^u(\varphi \circ \Delta^a f) = \varphi \circ \Delta^u(\Delta^a f)$$
$$= \varphi \circ \Delta^{ua} f = \varphi \circ \Delta^w f,$$

which concludes the proof. □

Most of the time, it is difficult to give explicit formulas for the difference operators of a given function. For the convenience of the reader, we present three examples where this computation is not only tractable, but also leads to interesting formulas.

Example 3.4. *The inversion function.*

We first apply the difference operators to the function mapping a word to its inverse in the free group. An auxiliary definition is in order to state this result. The *iterated commutator* $[x_1, x_2, \ldots, x_n]$ of n elements $x_1, x_2, \ldots, x_n$ of a group is defined by induction by setting $[x_1] = x_1$ and for $n \geqslant 2$,

$$[x_1, x_2, \ldots, x_n] = x_1 [x_2, x_3, \ldots, x_n] x_1^{-1} [x_2, x_3, \ldots, x_n]^{-1}.$$

In particular, since $[x_1, x_2] = x_1 x_2 x_1^{-1} x_2^{-1}$, one gets $[x_1, x_2, \ldots, x_n] = [x_1, [x_2, x_3, \ldots, x_n]]$.

Proposition 3.3. *Let $f: A^* \to F(A)$ be the function defined by*

$$f(x) = x^{-1}.$$

Then for every $n > 0$ and for all $a_1, \ldots, a_n \in A$,

$$\Delta^{a_1 a_2 \cdots a_n} f(x) = x[a_1, a_2, \ldots, a_n]^{-1} x^{-1} \tag{3.11}$$

Proof. For $n = 1$, the result follows from the formulas

$$\Delta^a f(x) = (f(x))^{-1} f(xa) = (x^{-1})^{-1}(xa)^{-1} = xa^{-1}x^{-1}$$

Let $n \geqslant 2$ and suppose that the result holds for $n-1$. Thus by (3.8), one has

$$\begin{aligned}(\Delta^{a_1 a_2 \cdots a_n} f)(x) &= \big(\Delta^{a_1}(\Delta^{a_2 \cdots a_n} f)\big)(x) \\ &= \big((\Delta^{a_2 \cdots a_n} f)(x)\big)^{-1} (\Delta^{a_2 \cdots a_n} f)(xa_1).\end{aligned}$$

Applying the induction hypothesis to $\Delta^{a_2 \cdots a_n} f$, one gets

$$(\Delta^{a_2 \cdots a_n} f)(x) = x[a_2, \ldots, a_n]^{-1} x^{-1}$$

and hence

$$\begin{aligned}(\Delta^{a_1 a_2 \cdots a_n} f)(x) &= \big(x[a_2, \ldots, a_n]^{-1} x^{-1}\big)^{-1} \big(xa_1[a_2, \ldots, a_n]^{-1}(xa_1)^{-1}\big) \\ &= x[a_2, \ldots, a_n] x^{-1} xa_1 [a_2, \ldots, a_n]^{-1}(xa_1)^{-1} \\ &= x[a_2, \ldots, a_n] a_1 [a_2, \ldots, a_n]^{-1} a_1^{-1} x^{-1} \\ &= x[a_1, a_2, \ldots, a_n]^{-1} x^{-1}\end{aligned}$$

which proves the induction step. □

Example 3.5. *The Euclidean division by 2 in base 3.*

We come back to the function f considered in Example 2.1. Let us compute the functions $\Delta^x f$. First, we have $\Delta^\varepsilon f = f$ and

$$\Delta^0 f(u) = \begin{cases} 0 & \text{if } \overline{u} \text{ is even} \\ 1 & \text{if } \overline{u} \text{ is odd} \end{cases}$$

$$\Delta^1 f(u) = \begin{cases} 0 & \text{if } \overline{u} \text{ is even} \\ 2 & \text{if } \overline{u} \text{ is odd} \end{cases}$$

$$\Delta^2 f(u) = \begin{cases} 1 & \text{if } \overline{u} \text{ is even} \\ 2 & \text{if } \overline{u} \text{ is odd} \end{cases}$$

The other values of $\Delta^x f$ can be obtained through the following result:

Proposition 3.4. *Let $s, t \in A^*$ and let $g: \{0,1,2\}^* \to A^*$ be the function defined by*

$$g(u) = \begin{cases} s & \text{if } \overline{u} \text{ is even} \\ t & \text{if } \overline{u} \text{ is odd} \end{cases}$$

Then $\Delta^\varepsilon g = g$ and, for each word x,

$$\Delta^x g(u) = \begin{cases} \varepsilon & \text{if } x \notin 1^* \\ (s^{-1}t)^{2^{n-1}(-1)^{n-1+\overline{u}}} & \text{if } x = 1^n \text{ for some } n > 0. \end{cases} \tag{3.12}$$

Proof. (1) Let us first compute $\Delta^0 g$, $\Delta^1 g$ and $\Delta^2 g$. Since $\overline{u}$, $\overline{u0}$ and $\overline{u2}$ have the same parity, one has $g(u) = g(u0) = g(u2)$, so that

$$\Delta^0 g(u) = g(u)^{-1} g(u0) = \epsilon \tag{3.13}$$

and

$$\Delta^2 g(u) = g(u)^{-1} g(u2) = \epsilon. \tag{3.14}$$

Similarly, $\Delta^1 g(u) = g(u)^{-1} g(u1)$, but now, $\overline{u}$ and $\overline{u1}$ have opposite parities. If $\overline{u}$ is even, then $\overline{u1}$ is odd, and therefore

$$\Delta^1 g(u) = g(u)^{-1} g(u1) = s^{-1}t = (s^{-1}t)^{(-1)^{\overline{u}}}. \tag{3.15}$$

The argument is similar when $\overline{u}$ is odd, and leads to the same formula, by noting that $s^{-1}t$ is the inverse of $t^{-1}s$.

(2) Let us prove (3.12) by induction on n. For $n = 1$, the result follows from (3.15). If (3.12) holds for some $n \geqslant 1$, then one has

$$\begin{aligned}
\Delta^{1^{n+1}} g(u) &= \Delta^1(\Delta^{1^n} g(u)) = (\Delta^{1^n} g(u))^{-1} \Delta^{1^n} g(u1) \\
&= ((s^{-1}t)^{2^{n-1}(-1)^{n-1+\overline{u}}})^{-1} (s^{-1}t)^{2^{n-1}(-1)^{n-1+\overline{u1}}} \\
&= (s^{-1}t)^{2^{n-1}(-1)^{n+\overline{u}}} (s^{-1}t)^{2^{n-1}(-1)^{n+\overline{u}}} = (s^{-1}t)^{2^n(-1)^{n+\overline{u}}},
\end{aligned}$$

which concludes the induction step.

(3) Suppose now that x is nonempty and not of the form 1^n. Then we may write $x = ya1^k$ with $a = 0$ or 2 and $k \geqslant 0$. If $k > 0$, then (3.12) shows that $\Delta^{1^k} g(u)$ can take at most two values, which depend on the parity of $\overline{u}$. This is also true if $k = 0$, because in this case, $\Delta^{1^k} g = \Delta^{\varepsilon} g = g$.

It now follows from (3.13) and (3.14), with g replaced by $\Delta^{1^k} g$, that, for all $u \in A^*$, $\Delta^{a1^k} g(u) = \Delta^a \Delta^{1^k} g(u) = \epsilon$ and hence $\Delta^x g(u) = \Delta^y \Delta^{a1^k} g(u) = \epsilon$. □

We already computed $\Delta^x f$ for the words x of length 0 or 1. Next we have for each $n > 0$

$$\begin{aligned}
\Delta^{1^n 0} f(u) &= (0^{-1}1)^{2^{n-1}(-1)^{n-1+\overline{u}}} \\
\Delta^{1^n 1} f(u) &= (0^{-1}2)^{2^{n-1}(-1)^{n-1+\overline{u}}} \\
\Delta^{1^n 2} f(u) &= (1^{-1}2)^{2^{n-1}(-1)^{n-1+\overline{u}}}
\end{aligned}$$

and, for any other word x, $\Delta^x f$ is the constant function to ε.

Example 3.6. *The noncommutative Magnus transformation.*

We now view the noncommutative Magnus transformation defined in Section 3.1 as a function from A^* to $F(A^*)$, the free group freely generated by A^*, for which we also adopt a noncommutative additive notation. The functions $\Delta^w \mu$ are easy to compute:

Proposition 3.5. *The following formula holds for all $u, w \in A^*$:*

$$\Delta^w \mu(u) = \mu(u) \cdot w \tag{3.16}$$

Proof. We prove (3.16) by induction on the length of w. It is trivial if w is the empty word. Suppose that the result holds for w and let a be a letter. Then we have, for all $w \in A^*$ and $a \in A$,

$$\begin{aligned}\Delta^{aw}\mu(u) &= \Delta^a(\Delta^w\mu)(u) = -\Delta^w\mu(u) + \Delta^w\mu(ua)\\ &= -\mu(u)\cdot w + \mu(ua)\cdot w\end{aligned}$$

Now, $\mu(ua) = \mu(u) + \mu(u)\cdot a$ by (3.3), and since $\cdot$ is a right action, one gets

$$\begin{aligned}\Delta^{aw}\mu(u) &= -\mu(u)\cdot w + (\mu(u) + \mu(u)\cdot a)\cdot w\\ &= -\mu(u)\cdot w + \mu(u)\cdot w + \mu(u)\cdot aw = \mu(u)\cdot aw\end{aligned}$$

which proves the induction step. □

3.3. **Newton's Forward Difference Formula**

For each $w \in A^*$, let us set

$$\delta_w f = \Delta^w f(1)$$

and let $\delta_f : A^* \to G$ be the map defined by $\delta_f(w) = \delta_w f$. This map extends to a monoid morphism $\delta_f^* : A^{**} \to G$. Thus $\delta_f^*(w) = \delta_w f$ and if $w_1 + \cdots + w_n$ is an element of A^{**}, then $\delta_f^*(w_1 + \cdots + w_n) = \delta_{w_1} f \cdots \delta_{w_n} f$.

Theorem 3.6 (Newton's Forward Difference Formula)**.** *The equality $f = \delta_f^* \circ \mu$ holds for each function $f : A^* \to G$.*

Before moving on to the proof of this formula, let us illustrate it on a few examples. Let a, b, c, d be letters of A. Then one has

$$\begin{aligned}f(1) &= \delta_1 f\\ f(a) &= (\delta_1 f)(\delta_a f)\\ f(ab) &= (\delta_1 f)(\delta_a f)(\delta_b f)(\delta_{ab} f)\\ f(abc) &= (\delta_1 f)(\delta_a f)(\delta_b f)(\delta_{ab} f)(\delta_c f)(\delta_{ac} f)(\delta_{bc} f)(\delta_{abc} f)\\ f(abcd) &= (\delta_1 f)(\delta_a f)(\delta_b f)(\delta_{ab} f)(\delta_c f)(\delta_{ac} f)(\delta_{bc} f)(\delta_{abc} f)\\ &\qquad (\delta_d f)(\delta_{ad} f)(\delta_{bd} f)(\delta_{abd} f)(\delta_{cd} f)(\delta_{acd} f)(\delta_{bcd} f)(\delta_{abcd} f)\\ f(abab) &= (\delta_1 f)(\delta_a f)(\delta_b f)(\delta_{ab} f)(\delta_a f)(\delta_{aa} f)(\delta_{ba} f)(\delta_{aba} f)\\ &\qquad (\delta_b f)(\delta_{ab} f)(\delta_{bb} f)(\delta_{abb} f)(\delta_{ab} f)(\delta_{aab} f)(\delta_{bab} f)(\delta_{abab} f)\end{aligned}$$

a formula deduced from (3.4) by eliminating the + signs and by replacing each word u by $\delta_u f$.

Theorem 3.6 relies on the following lemma:

Lemma 3.7. *The following formula holds for every $x \in A^{**}$ and every $a \in A$:*

$$\delta_f^*(x\cdot a) = \delta_{\Delta^a f}^*(x). \tag{3.17}$$

Proof. Since the map $x \mapsto x\cdot a$ is a monoid endomorphism of A^{**}, both sides of (3.17), viewed as functions of x, are monoid morphisms from A^{**} into G. Therefore, it suffices to establish (3.17) when x is a generator of A^{**}, that is, $x = u$ for some word $u \in A^*$. Then $x\cdot a = ua$ and hence one has

$$\delta_f^*(u\cdot a) = \delta_{ua} f = \Delta^{ua} f(1) = \Delta^u\Delta^a f(1) = \delta_u(\Delta^a f) = \delta_{\Delta^a f}^*(u).$$ □

Proof of Theorem 3.6. Let us show that, for every word $w \in A^*$,

$$f(w) = \delta_f^* \circ \mu(w) \tag{3.18}$$

We prove (3.18) by induction on $|w|$. If $|w| = 0$, then w is the empty word, and

$$f(1) = \Delta^1 f(1) = \delta_1 f = \delta_f^*(1) = \delta_f^* \circ \mu(1).$$

Suppose that the result holds for all words of length $\leqslant n$ and let u be a word of length $n+1$. Let w be the prefix of length n of u and let a be its last letter, so that $u = wa$. Observing that $\Delta^a f(w) = f(w)^{-1} f(wa)$, one gets

$$f(u) = f(wa) = f(w)\Delta^a f(w). \tag{3.19}$$

Moreover, the induction hypothesis yields

$$\begin{aligned} f(w) &= \delta_f^*(\mu(w)) \\ \Delta^a f(w) &= \delta_{\Delta^a f}^* \circ \mu(w). \end{aligned} \tag{3.20}$$

Applying now Lemma 3.7 with $x = \mu(w)$, one gets

$$\Delta^a f(w) = \delta_{\Delta^a f}^* \circ \mu(w) = \delta_f^*(\mu(w) \cdot a) \tag{3.21}$$

Plugging (3.20) and (3.21) into (3.19) yields

$$\begin{aligned} f(u) = f(w)\Delta^a f(w) &= \delta_f^*\big(\mu(w)\big)\delta_f^*\big(\mu(w)\cdot a\big) \\ &= \delta_f^*\big(\mu(w) + \mu(w)\cdot a\big) \\ &= \delta_f^*\big(\mu(wa)\big) = \delta_f^* \circ \mu(u) \end{aligned}$$

which proves the induction step and concludes the proof. □

Example 3.7. A direct application of Proposition 3.3 and Theorem 3.6 leads to the formula

$$(abc)^{-1} = a^{-1}b^{-1}[a,b]^{-1}c^{-1}[a,c]^{-1}[b,c]^{-1}[a,b,c]^{-1} \tag{3.22}$$

or, equivalently,

$$abc = [a,b,c][b,c][a,c]c[a,b]ba \tag{3.23}$$

Example 3.8. Let us come back to the function f considered in Examples 2.1 and 3.5. Proposition 3.4 shows that $\delta_0 f = 0$, $\delta_1 f = 0$, $\delta_2 f = 1$, and for each $n > 0$,

$$\begin{aligned} \delta_{1^n 0} f &= (0^{-1}1)^{2^{n-1}(-1)^{n-1}} \\ \delta_{1^n 1} f &= (0^{-1}2)^{2^{n-1}(-1)^{n-1}} \\ \delta_{1^n 2} f &= (1^{-1}2)^{2^{n-1}(-1)^{n-1}} \end{aligned}$$

and $\delta_x f = \epsilon$ in all other cases. Applying Newton's Formula Difference Formula, we get for instance

$$\begin{aligned} f(1212) &= \delta_\epsilon f\ \delta_1 f\ \delta_2 f\ \delta_{12} f\ \delta_1 f\ \delta_{11} f\ \delta_{21} f \\ &\qquad \delta_{121} f\ \delta_2 f\ \delta_{12} f\ \delta_{22} f\ \delta_{122} f\ \delta_{12} f\ \delta_{112} f\ \delta_{212} f\ \delta_{1212} f \\ &= \delta_1 f\ \delta_2 f\ \delta_{12} f\ \delta_1 f\ \delta_{11} f\ \delta_2 f\ \delta_{12} f\ \delta_{12} f\ \delta_{112} f \\ &= 01(1^{-1}2)0(0^{-1}2)1(1^{-1}2)(1^{-1}2)(1^{-1}2)^{-2} = 0221, \end{aligned}$$

a somewhat convoluted way to show that 50 divided by 2 equals 25.

Example 3.9. For the noncommutative Magnus transformation, Proposition 3.5 shows that $\delta_w \mu = w$ for all $w \in A^*$. Thus δ_μ^* is the identity, in accordance with Theorem 3.6.

Newton's Forward Difference Formula allows one to recover f from the elements $\delta_u f$. A formula giving $\Delta^w f$ in terms of f, as shown in Example 3.3, was given in [10]. Let us briefly review the steps leading to this formula.

We first consider μ as a function from A^* to $F(A^*)$, the free group freely generated by A^*, for which we keep the noncommutative additive notation already used for A^{**}. This means that every element of $F(A^*)$ is written as

$$\pm x_1 \pm \cdots \pm x_n$$

with $x_1, \ldots, x_n \in A^*$ and no consecutive terms of the form $-x + x$ or $+x - x$ occur in this expression.

Next we extend μ, in the only possible way, to an endomorphism of $F(A^*)$. We show below that it is actually an automorphism, and how to construct its inverse. We also define a right and a left action of A^* on A^{**} as follows. For each element $\pm x_1 + \cdots \pm x_r$ of $F(A^*)$ and each $u \in A^*$, we set

$$(\pm x_1 + \cdots \pm x_n) \cdot u = (\pm x_1 u \pm \cdots \pm x_n u)$$
$$u \cdot (\pm x_1 + \cdots \pm x_n) = \pm u x_1 + \cdots \pm u x_n$$

Note that the right action extends the right action of A^* on A^{**} given by (3.2).

The formula giving $\Delta^w f$ is now easy to obtain. Recall that f is a function from A^* to some group G. Thus f uniquely extends to a group morphism $f^* : F(A^*) \to G$, and the following formula, stated in [10, Proposition 4.5] with a slightly different notation, holds for all u, w in A^*:

$$\Delta^w f(u) = f^*(u \cdot \mu^{-1}(w)) \tag{3.24}$$

It remains to give an explicit formula for the inverse of μ. For this purpose, we introduce a new function $\pi : A^* \to F(A^*)$ defined recursively by setting $\pi(1) = 1$ and, for all $w \in A^*$ and $a \in A$,

$$\pi(wa) = -\pi(w) + \pi(w) \cdot a. \tag{3.25}$$

For instance, if $a, b, c, d \in A$, then we have:

$$\begin{aligned}
\pi(a) &= -1 + a, \\
\pi(ab) &= -a + 1 - b + ab, \\
\pi(abc) &= -ab + b - 1 + a - ac + c - bc + abc, \\
\pi(abcd) &= -abc + bc - c + ac - a + 1 - b + ab - abd + bd - d \\
&\quad + ad - acd + cd - bcd + abcd, \\
\pi(aba) &= -ab + b - 1 + a - aa + a - ba + aba, \\
\pi(abab) &= -aba + ba - a + aa - a + 1 - b + ab - abb + bb - b \\
&\quad + ab - aab + ab - bab + abab.
\end{aligned} \tag{3.26}$$

In the same way as μ, the function π uniquely extends to an endomorphism of $F(A^*)$, also denoted by π. This endomorphism π is not yet the inverse of μ, but we are almost there.

The *reversal* of a word $u = a_1 \cdots a_n$ is the word $\tilde{u} = a_n \cdots a_1$. The reversal map is a permutation on A^* which uniquely extends to a group automorphism of $F(A^*)$. According to [10, Corollary 3.4], the inverse of μ is given by the following formula, for all $x \in F(A^*)$,

$$\mu^{-1}(x) = \widetilde{\pi(\tilde{x})} \tag{3.27}$$

For instance, if $x = ab$, then $\tilde{x} = ba$, whence

$$\pi(ba) = -b + 1 - a + ba \quad \text{and} \quad \mu^{-1}(ab) = \widetilde{\pi(ba)} = -b + 1 - a + ab.$$

For a more complicated example, we let the reader verify that applying (3.24), (3.26) and (3.27), one recovers (3.9).

4. Polynomial functions

In this section, we extend the notion of a Newton polynomial function from $\mathbb{N}$ to $\mathbb{Z}$ to functions from A^* to some group G. The formal definition, as well as a useful characterization, are given in Section 4.1. Next, in Section 4.2, we introduce a new construction, the sequential product, that we use to solve the *integration problem*. In Section 4.3, we associate to each function f from A^* to G a sequence of Newton polynomial functions f_r of degree at most r.

4.1. Polynomial functions and their degree

Let $\mathbf{1}$ denote the constant function from A^* to G that maps every word to 1, the identity element of G.

A function $f\colon A^* \to G$ is called a *Newton polynomial function* if $\Delta^w f = \mathbf{1}$ for almost all words $w \in A^*$. In this case, the *degree* of f, denoted $\deg(f)$, is defined by

$$\begin{aligned}\deg(f) = \min\{d \in \mathbb{N} \cup \{-1\} \mid \Delta^w f = \mathbf{1} \\ \text{for all words } w \in A^* \text{ such that } |w| > d\}.\end{aligned} \tag{4.1}$$

or equivalently, using (3.8) and the fact that $\Delta^w \mathbf{1} = \mathbf{1}$ for all words w,

$$\begin{aligned}\deg(f) = \min\{d \in \mathbb{N} \cup \{-1\} \mid \Delta^w f = \mathbf{1} \\ \text{for all words } w \in A^* \text{ such that } |w| = d + 1\}.\end{aligned} \tag{4.2}$$

Observe that (3.8) and the definition of the degree imply the following inequality, for all words $w \in A^*$,

$$\deg(\Delta^w f) \leqslant \deg(f) - |w|. \tag{4.3}$$

Proposition 4.1.

(1) *The unique Newton polynomial function of degree* -1 *is the function* $\mathbf{1}$.

(2) *A Newton polynomial function has degree* 0 *if and only if it is a constant function different from* $\mathbf{1}$.

(3) *For every Newton polynomial function* f *of nonnegative degree,*

$$\deg(f) = 1 + \max\{\deg(\Delta^a f) \mid a \in A\}. \tag{4.4}$$

Proof. (1) The equality $\deg(\mathbf{1}) = -1$ follows from the definition, and since $\Delta^1 f = f$, the degree of each function $f \neq \mathbf{1}$ is nonnegative.

(2) Let f be a constant function and a be a letter. Then since $\Delta^a f(u) = f(u)^{-1} f(ua)$, one has $\Delta^a f = \mathbf{1}$, and hence $\deg(f) \leqslant 0$ by (4.2). Moreover, if $f \neq \mathbf{1}$, then $\deg(f) = 0$ by (1).

Conversely, if $\deg(f) = 0$, then $f \neq \mathbf{1}$ by (1) and $\Delta^a f = \mathbf{1}$ for each letter a. Since $\Delta^a(u) = f(u)^{-1} f(ua)$, one has $f(ua) = f(u)$, and by an easy induction, f is a constant function.

(3) Let $d = \deg(f)$. If $d = 0$, then (4.4) is immediately verified. If $d > 0$, then $\Delta^w f = \mathbf{1}$ for all words w of length $> d$ and there exists a word w of length d such that $\Delta^w f \neq \mathbf{1}$. Setting $w = uc$, where c is a letter and $|u| = d$, one gets $\Delta^w f = \Delta^u(\Delta^c f)$ by (3.8), whence $\Delta^u(\Delta^c f) \neq \mathbf{1}$ and $\deg(\Delta^c f) \geqslant d - 1$ by (4.2). Therefore

$$d \leqslant 1 + \deg(\Delta^c f) \leqslant 1 + \max\{\deg(\Delta^a f) \mid a \in A\}.$$

Moreover, it follows from (4.3) that, for each letter a, $\deg(\Delta^a f) \leqslant d - 1$ and thus $1 + \max\{\deg(\Delta^a f) \mid a \in A\} \leqslant d$, which proves (4.4). □

The degree of a Newton polynomial function may also be defined by using the functions δ instead of Δ.

Proposition 4.2. *Let $f : A^* \to G$ be a Newton polynomial function and let $d \in \mathbb{N} \cup \{-1\}$. Then $\deg(f) \leqslant d$ if and only if $\delta_w f = 1$ for all words w of length $> d$.*

Proof. Suppose that $\deg(f) \leqslant d$ and let w be a word of length $> d$. Then $\Delta^w f = \mathbf{1}$ by (4.1) and since $\delta_w f = \Delta^w f(1)$, one has $\delta_w f = 1$.

In the opposite direction, we now prove by induction on d that, if $\delta_w f = 1$ for all words w of length $> d$, then $\deg(f) \leqslant d$. If $d = -1$, then $\delta_w f = 1$ for each word w. It follows by Theorem 3.6 that $f = \mathbf{1}$ and thus $\deg(f) = -1$.

Suppose now that $d \geqslant 0$. Since $\deg(f) = 1 + \max\{\deg(\Delta^a f) \mid a \in A\}$ by (4.4), it suffices to show that, for each letter a, $\deg(\Delta^a f) \leqslant d - 1$. Let u be a word of length $> d - 1$. Using (3.8), one gets, since $|ua| > d$,

$$\delta_u(\Delta^a f) = \Delta^u \Delta^a f(1) = \Delta^{ua} f(1) = \delta_{ua} f = 1.$$

It follows by the induction hypothesis that $\deg(\Delta^a f) \leqslant d - 1$, as required. □

Examples of Newton polynomial functions of degree 2 and examples of non-polynomial functions are given at the end of Section 4.2.

4.2. **Integration problem and sequential products**

We now show that f and the functions $\Delta^a f$, for $a \in A$, are related by a functional equation.

Proposition 4.3. *Let $a_1 \cdots a_n$ be a word of A^*. Then the following formula holds:*

$$f(a_1 \cdots a_n) = f(1) \prod_{1 \leqslant i \leqslant n} \Delta^{a_i} f(a_1 \cdots a_{i-1}). \tag{4.5}$$

where the product is evaluated from left to right.

Proof. The result is trivial if $n = 0$. Moreover, if (4.5) holds for n, then, by induction on n

$$f(1)\prod_{1\leqslant i\leqslant n+1}\Delta^{a_i}f(a_1\cdots a_{i-1}) = f(a_1\cdots a_n)\Delta^{a_{n+1}}f(a_1\cdots a_n)$$
$$= f(a_1\cdots a_n)f(a_1\ldots a_n)^{-1}f(a_1\cdots a_n a_{n+1}) = f(a_1\cdots a_n a_{n+1})$$

which proves (4.5). □

The functional equation (4.5) gives an expression of f in terms of $f(1)$ and of the family $(\Delta^a f)_{a\in A}$. We now address the opposite question, which is somewhat similar to the problem of integrating a function from its derivative.

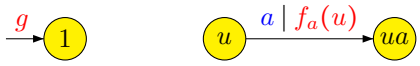
Integration problem. *Given an element g of G and a family $(f_a)_{a\in A}$ of functions from A^* to G, is there a function f such that $f(1) = g$ and $f_a = \Delta^a f$ for all $a \in A$?*

To solve the integration problem, it is convenient to introduce a new definition. Given an element g of G and a family $(f_a)_{a\in A}$ of functions from A^* to G, the *sequential product* $\mathrm{Seq}(g, (f_a)_{a\in A})$ is the function $f\colon A^* \to G$, defined, for each word $a_1\cdots a_n \in A^*$, by

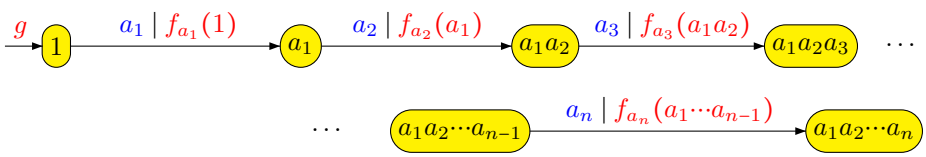
$$f(a_1\cdots a_n) = g\prod_{1\leqslant i\leqslant n} f_{a_i}(a_1\cdots a_{i-1}). \tag{4.6}$$

By abuse of language, a function $f\colon A^* \to G$ is called *a sequential product of a family $(f_a)_{a\in A}$ of functions from A^* to G* if, for some $g \in G$, $f = \mathrm{Seq}(g, (f_a)_{a\in A})$.

This terminology stems from the fact that f can be realized by a sequential transducer with infinitely many states. Indeed, consider the sequential transducer $\mathcal{A} = (A^*, A, G, 1, \cdot, *, g)$, where A^* is the set of states, A the input alphabet, G the output group, 1 the initial state, g the initial prefix. The transition and the output functions are respectively defined by $u\cdot a = ua$ and $u * a = f_a(u)$.



A typical computation in $\mathcal{A}$ looks like this



and hence $\mathcal{A}$ computes the sequential product f defined by (4.6).

We are now ready to solve the integration problem.

Proposition 4.4. *Let $g \in G$ and let $(f_a)_{a\in A}$ be a family of functions from A^* to G. Then the sequential product $\mathrm{Seq}(g, (f_a)_{a\in A})$ is the unique function f such that $f(1) = g$ and $\Delta^a f = f_a$ for all $a \in A$.*

Proof. Let $f = \mathrm{Seq}(g, (f_a)_{a\in A})$. Then $f(1) = g$ by definition. Let $u = a_1 \dots a_n$ be a word and a be a letter. Since $\Delta^a f(u) = f(u)^{-1} f(ua)$, one gets by (4.5)

$$\begin{aligned}\Delta^a f(u) &= \Big(g \prod_{1\leqslant i\leqslant n} f_{a_i}(a_1\cdots a_{i-1})\Big)^{-1} g\Big(\prod_{1\leqslant i\leqslant n} f_{a_i}(a_1\cdots a_{i-1})\Big) f_a(a_1\cdots a_n)\\ &= f_a(a_1\cdots a_n)\end{aligned}$$

whence $\Delta^a f = f_a$.

To prove uniqueness, consider a function f such that $f(1) = g$ and $\Delta^a f = f_a$ for all $a \in A$. Then for each word $a_1\cdots a_n \in A^*$, one gets by (4.5),

$$f(a_1\cdots a_n) = f(1) \prod_{1\leqslant i\leqslant n} \Delta^{a_i} f(a_1\cdots a_{i-1}) = g \prod_{1\leqslant i\leqslant n} f_{a_i}(a_1\cdots a_{i-1}).$$

and thus $f = \mathrm{Seq}(g, (f_a)_{a\in A})$. □

Proposition 4.5. *Let G be a group and let $f\colon A^* \to G$ be a function. The following conditions are equivalent:*

(1) *f is a Newton polynomial function of degree $\leqslant d$,*
(2) *there exists a family $(f_a)_{a\in A}$ of polynomial functions of degree $\leqslant d-1$ such that $f = \mathrm{Seq}(f(1), (f_a)_{a\in A})$.*

In this case, one has $f_a = \Delta^a f$ for every $a \in A$.

Proof. (1) ⇒ (2). Let f be a polynomial function of degree $\leqslant d$. Formula (4.3) shows that, for each letter a, $\Delta^a f$ is a Newton polynomial function of degree at most $d-1$. Moreover, Proposition 4.3 shows that $f(a_1\cdots a_n) = f(1)\prod_{1\leqslant i\leqslant n} \Delta^{a_i} f(a_1\cdots a_{i-1})$, which proves (2).

(2) ⇒ (1). Suppose that (2) holds. Proposition 4.4 shows that, for each letter a, $\Delta^a f = f_a$ and hence $\Delta^a f$ is a Newton polynomial function of degree $\leqslant d-1$. It follows that f is a Newton polynomial function function of degree $\leqslant d$. □

The following characterization of the set of Newton polynomial functions is now an immediate consequence of Proposition 4.5.

Corollary 4.6. *Let G be a group. The set of Newton polynomial functions from A^* to G is the smallest set of functions from A^* to G containing the constant functions and closed under sequential product.*

We now come back to the characterization of the Newton polynomial functions of degree $\leqslant 1$. Let us say that a function $f\colon A^* \to G$ is an *affine morphism* if $f = f(1)g$ for some monoid morphism $g\colon A^* \to G$. Equivalently, conjugating by $f(1)$, one gets $f = hf(1)$ for some monoid morphism $h\colon A^* \to G$.

Proposition 4.7. *A function from A^* to G is a Newton polynomial function of degree $\leqslant 1$ if and only if it is an affine morphism.*

Proof. Proposition 4.5 shows that a function from A^* to G is a polynomial function of degree $\leqslant 1$ if and only if there exists a family $(f_a)_{a\in A}$ of Newton polynomial functions of degree $\leqslant 0$ such that $f = \mathrm{Seq}(f(1), (f_a)_{a\in A})$. Proposition 4.1 shows that these polynomial functions f_a of degree $\leqslant 0$ are constant functions equal to some element g_a of G. It follows that f is a

Newton polynomial function of degree $\leqslant 1$ if and only if there is a family $(g_a)_{a\in A}$ of elements of G such that, for each word $w = a_1\cdots a_n$,

$$f(w) = f(1)\prod_{1\leqslant i\leqslant n} g_{a_i}. \tag{4.7}$$

Defining $g\colon A^* \to G$ as the unique monoid morphism such that $g(a) = g_a$ for each letter a, (4.7) is equivalent to writing $f(w) = f(1)g(w)$, which proves the result. □

Example 4.1. The function $f\colon A^* \to F(A)$ defined by

$$f(a_1\cdots a_n) = a_1(a_1a_2)(a_1a_2a_3)\cdots(a_1\cdots a_n)$$

is a Newton polynomial function of degree 2. Indeed, it is equal to the sequential product $\mathrm{Seq}(1, (f_a)_{a\in A})$ where each f_a is the affine morphism defined by $f_a(u) = ua$.

Proposition 4.8. *Any Newton polynomial function of finite image from A^* to a free group is a constant function.*

Proof. Let $f\colon A^* \to F(B)$ be a Newton polynomial function of finite image. Then for each $a \in A$, the image of $\Delta^a f$ is also finite since $\Delta^a f(u)$, which is equal to $f(u)^{-1}f(ua)$, can only take finitely many values. It follows by induction that, for every word w, the image of $\Delta^w f$ is also finite.

Let d be the degree of f and suppose that $d > 0$. By (4.2), there exists a word w of length d such that $\Delta^w f \neq \mathbf{1}$. Let a be the first letter of w and s its suffix of length $d-1$. By (4.3), $\Delta^s f$ is a Newton polynomial function of degree $\leqslant 1$, and since $\Delta^a(\Delta^s f) \neq \mathbf{1}$, it is actually of degree 1. It follows from Proposition 4.7 that f is an affine morphism. Consequently, there is a monoid morphism $g\colon A^* \to F(B)$ such that $\Delta^s f = \Delta^s f(1)g$. But since $\Delta^s f$ has finite image, $g(A^*)$ has to be a finite submonoid of $F(B)$ and the unique possibility is $g = \mathbf{1}$. But in this case, $\Delta^s f$ is a constant function and hence has degree 0, a contradiction. Thus $d \leqslant 0$ and f is a constant function. □

Just like difference operators, sequential products commute with group morphisms:

Proposition 4.9. *Let $\varphi : G \to H$ be a group morphism and let $(f_a)_{a\in A}$ be a family of functions from A^* to G. Then the following equality holds:*

$$\varphi \circ \mathrm{Seq}(g, (f_a)_{a\in A}) = \mathrm{Seq}(\varphi(g), (\varphi \circ f_a)_{a\in A}). \tag{4.8}$$

Proof. Since φ is a morphism, one has

$$\begin{aligned}\varphi\big(\mathrm{Seq}(g, (f_a)_{a\in A})(a_1\cdots a_n)\big) &= \varphi\Big(g\prod_{1\leqslant i\leqslant n} f_{a_i}(a_1\cdots a_{i-1})\Big)\\ &= \varphi(g)\prod_{1\leqslant i\leqslant n}\varphi\big(f_{a_i}(a_1\cdots a_{i-1})\big)\\ &= \mathrm{Seq}\big(\varphi(g), (\varphi\circ f_a)_{a\in A}\big)(a_1\cdots a_n)\end{aligned}$$

which proves (4.8). □

4.3. **Newton polynomial functions associated to a function**

For each $r \in \mathbb{N}$, let C_r be the set of words of A^* of length at most r. Let ρ_r be the unique monoid endomorphism of A^{**} which maps every element of C_r to itself, and maps any other element of A^* to 0. In other words, if $x = \sum_{1 \leqslant i \leqslant s} u_i$ is an element of A^{**}, where each $u_i \in A^*$ and if $E_r(x) = \{i \in \{1, \dots s\} \mid |u_i| \leqslant r\}$, then

$$\rho_r(x) = \sum_{i \in E_r(x)} u_i. \tag{4.9}$$

For instance $\rho_3(1 + ab + baba + 1 + aba + abaab + b) = 1 + ab + 1 + aba + b$. The function

$$\mu_r = \rho_r \circ \mu$$

from A^* to A^{**} is called the r-th *truncated noncommutative Magnus transformation.*

Example 4.2. Let $a, b, c, d \in A$. Then

$$\begin{aligned}
\mu_2(a) &= 1 + a \\
\mu_2(ab) &= 1 + a + b + ab \\
\mu_2(abc) &= 1 + a + b + ab + c + ac + bc \\
\mu_2(abcd) &= 1 + a + b + ab + c + ac + bc + d + ad + bd + cd \\
\mu_2(abab) &= 1 + a + b + ab + a + aa + ba + b + ab + bb + ab.
\end{aligned}$$

Thus $\mu_2(abab)$, for instance, is obtained by only keeping the words of length $\leqslant 2$ in $\mu(abab)$, as given in (3.4).

Formula (3.3) admits a truncated version:

Lemma 4.10. *The following formula holds for all $w \in A^*$, $a \in A$ and $r > 0$:*

$$\rho_r(\mu(w) \cdot a) = \rho_{r-1}(\mu(w)) \cdot a, \tag{4.10}$$

$$\mu_r(wa) = \mu_r(w) + \mu_{r-1}(w) \cdot a. \tag{4.11}$$

Proof. Formula (4.10) follows from an inspection of the words of $E_r(\mu(w) \cdot a)$: they are exactly the words of the form ua where $u \in E_{r-1}(\mu(w))$.

Let us prove (4.11). By definition, $\mu_r(wa)) = \rho_r \circ \mu(wa))$. Since ρ_r is a monoid endomorphism on A^{**}, applying ρ_r to each side of (3.3) yields:

$$\rho_r(\mu(wa)) = \rho_r(\mu(w) + \mu(w) \cdot a) = \rho_r(\mu(w)) + \rho_r(\mu(w) \cdot a).$$

Moreover, it follows from (4.10) that

$$\rho_r(\mu(w) \cdot a) = \rho_{r-1}(\mu(w)) \cdot a = \mu_{r-1}(w) \cdot a.$$

and hence

$$\rho_r(\mu(wa)) = \rho_r(\mu(w)) + \mu_{r-1}(w) \cdot a = \mu_r(w) + \mu_{r-1}(w) \cdot a. \qquad \square$$

An interesting consequence of Lemma 4.10 is that μ_r is can be expressed as a sequential product of the functions $\mu_{r-1} \cdot a$, for $a \in A$. For this to make sense, we need to consider each μ_r as a function from A^* to the free group of base A^*, written additively like A^{**}.

Corollary 4.11. *For every $r > 0$, $\mu_r = \mathrm{Seq}(1, (\mu_{r-1} \cdot a)_{a \in A})$.*

Proof. On the one hand, it follows from (4.11) that for all $w \in A^*$ and $a \in A$

$$\Delta^a \mu_r(w) = -\mu_r(w) + \mu_r(wa) = \mu_{r-1}(w) \cdot a \tag{4.12}$$

On the other hand, Proposition 4.3 shows that

$$\mu_r = \mathrm{Seq}(1, (\Delta^a \mu_r)_{a \in A}) \tag{4.13}$$

The corollary now follows immediately from (4.12) and (4.13). □

To each function $f: A^* \to G$, we associate, for each $r \geqslant 0$, a function $f_r: A^* \to G$ by setting $f_r = \delta_f^* \circ \mu_r$. We will see in Proposition 4.14 below that f_r is a Newton polynomial function. For this reason, we call f_r the r-th *Newton polynomial function associated to* f.

Lemma 4.12. *The function f_0 is the constant function equal to $f(1)$.*

Proof. Indeed, since $\Delta^1 f = f$, one gets

$$f_0(u) = \delta_f^* \circ \mu_0(u) = \delta_f^*(1) = \delta_1 f = \Delta^1 f(1) = f(1). \qquad \square$$

Example 4.3. Here are a few other examples, in which we write δ_u instead of $\delta_u f$.

$$\begin{aligned}
f_2(a) &= \delta_f^* \circ \mu_2(a) = \delta_f^*(1 + a) = \delta_1 \delta_a \\
f_2(ab) &= \delta_f^* \circ \mu_2(ab) = \delta_f^*(1 + a + b + ab) = \delta_1 \delta_a \delta_b \delta_{ab}, \\
f_2(abc) &= \delta_f^* \circ \mu_2(abc) = \delta_f^*(1 + a + b + ab + c + ac + bc) \\
&= \delta_1 \delta_a \delta_b \delta_{ab} \delta_c \delta_{ac} \delta_{bc} \\
f_2(abcd) &= \delta_f^* \circ \mu_2(abcd) \\
&= \delta_f^*(1 + a + b + ab + c + ac + bc + d + ad + bd + cd) \\
&= \delta_1 \delta_a \delta_b \delta_{ab} \delta_c \delta_{ac} \delta_{bc} \delta_d \delta_{ad} \delta_{bd} \delta_{cd} \\
f_2(abab) &= \delta_1 \delta_a \delta_b \delta_{ab} \delta_a \delta_{aa} \delta_{ba} \delta_b \delta_{ab} \delta_{bb} \delta_{ab}.
\end{aligned}$$

Proposition 4.13. *The formula*

$$\Delta^a(f_r) = \delta_{\Delta^a f}^* \circ \mu_{r-1} = (\Delta^a f)_{r-1} \tag{4.14}$$

holds for all $r > 0$ and $a \in A$.

Proof. Since $f_r = \delta_f^* \circ \mu_r$ and $\mu_r(ua) = \mu_r(u) + \mu_{r-1}(u) \cdot a$ by (4.11), the formula (4.14) follows from the following sequence of equalities:

$$\begin{aligned}
\Delta^a f_r(u) &= f_r(u)^{-1} f_r(ua) = \left[\delta_f^*(\mu_r(u))\right]^{-1} \delta_f^*(\mu_r(ua)) \\
&= \left[\delta_f^*(\mu_r(u))\right]^{-1} \delta_f^*(\mu_r(u) + \mu_{r-1}(u) \cdot a) \\
&= \left[\delta_f^*(\mu_r(u))\right]^{-1} \delta_f^*(\mu_r(u)) \delta_f^*(\mu_{r-1}(u) \cdot a) = \delta_f^*(\mu_{r-1}(u) \cdot a) \\
&= \delta_{\Delta^a f}^*(\mu_{r-1}(u)) \quad \text{by Lemma 3.7} \\
&= (\Delta^a f)_{r-1}(u) \quad \text{by the definition of } (\Delta^a f)_{r-1}.
\end{aligned}$$ □

Proposition 4.14. *For each $r \geqslant 0$, f_r is a Newton polynomial function of degree at most r.*

Proof. We prove the result by induction on r. For $r = 0$, the result follows from Lemma 4.12.

Applying Proposition 4.3 to f_r, one gets, for every word $a_1 \cdots a_k \in A^*$,

$$f_r(a_1 \cdots a_k) = f_r(1) \prod_{1 \leqslant i \leqslant k} \Delta^{a_i} f_r(a_1 \dots a_{i-1}). \tag{4.15}$$

Now, $f_r(1) = \delta_f^* \circ \mu_r(1) = \delta_f^*(1) = f(1)$ and $\Delta^a f_r = (\Delta^a f)_{r-1}$ by Proposition 4.13. It follows that

$$f_r(a_1 \dots a_k) = f(1) \prod_{1 \leqslant i \leqslant k} (\Delta^{a_i} f)_{r-1}(a_1 \dots a_{i-1}). \tag{4.16}$$

By the induction hypothesis applied to $\Delta^a f$, $(\Delta^a f)_{r-1}$ is a Newton polynomial function of degree at most $r-1$. Hence by Proposition 4.5, f_r is a Newton polynomial function of degree at most r. □

Recall that δ_f^* is a monoid morphism from A^{**} to G, but we keep the same notation for its restriction to C_r^*. Theorem 3.6 admits the following counterpart.

Corollary 4.15. *Let $f: A^* \to G$ be a Newton polynomial function of degree at most d. Then $f = \delta_f^* \circ \mu_d$.*

Proof. It suffices to use Theorem 3.6 and to observe that $\delta_f^*(u) = \delta_u f = \Delta^u f(1) = 1$ for each word of length $> d$. □

5. Newton's bijection

Recall that to each function $f: A^* \to G$ is associated the map $\delta_f: A^* \to G$ defined by $\delta_f(w) = \delta_w f$. The *Newton map* is the map $\delta: f \to \delta_f$. We show in this section that δ is a bijection and we explicitly find its inverse.

Let $f^*: A^{**} \to G$ denote the unique monoid morphism extending f and let $\gamma: A^* \to G$ be the map defined by $\gamma(f) = f^* \circ \mu$. Thus, if $u \in A^*$ and $\mu(u) = u_1 + \cdots + u_m$, then

$$\gamma(u) = f^*(u_1 + \cdots + u_m) = f(u_1) \cdots f(u_m).$$

Theorem 5.1 (Newton's bijection)**.** *The Newton map δ is a permutation on the set of functions from A^* to G and its inverse is the permutation γ.*

Proof. Since $f = \delta_f^* \circ \mu$ by Theorem 3.6, $\gamma \circ \delta$ is the identity function. Therefore γ is surjective, δ is injective and it suffices to prove that γ is injective. Let $g, h: A^* \to G$ be such that $g^* \circ \mu = h^* \circ \mu$. Let us show by induction on $|u|$ that $g(u) = h(u)$. If $|u| = 0$, then u is the empty word 1, $\mu(1) = 1$, $g^*(1) = g(1)$, $h^*(1) = h(1)$ and thus $g(1) = h(1)$. Suppose now that $|u| = r+1$. Then $\mu(u) = \mu_r(u) + u$ and since g^* and h^* are monoid morphisms, one gets $g^* \circ \mu(u) = g^*(\mu_r(u) + u) = g^*(\mu_r(u))g(u)$ and similarly $h^* \circ \mu(u) = h^*(\mu_r(u))h(u)$. Since $\mu_r(u)$ is a sum of words of length $\leqslant r$, the induction hypothesis gives $g^*(\mu_r(u)) = h^*(\mu_r(u))$. Now since $g^* \circ \mu(u) = h^* \circ \mu(u)$, one gets $g(u) = h(u)$, which concludes the induction step. □

Theorem 5.1 solves the following interpolation problem.

Corollary 5.2. *For each function $g: A^* \to G$, there exists a unique function $f: A^* \to G$ such that, for all $u \in A^*$, $\delta_u f = g(u)$.*

A function $f: A^* \to G$ is called a *G-polynomial* if $f(w) = 1$ for almost all words $w \in A^*$. The *degree* of a G-polynomial is -1 if $f = \mathbf{1}$; otherwise, it is the smallest d such that $f(w) = 1$ for every word of length $d+1$. One can now enrich Theorem 5.1 as follows.

Theorem 5.3. *For each degree d, the maps δ and γ define mutually inverse bijections between the set of Newton polynomial functions of degree d and the set of G-polynomials of degree d.*

Proof. It suffices to prove that δ and γ define mutually inverse bijections between the set of Newton polynomial functions of degree at most d and the set of G-polynomials of degree at most d. Let f be a Newton polynomial function of degree $\leqslant d$. Then by definition, $\delta(f)$ is a G-polynomial of degree at most d. Let now f be a G-polynomial of degree at most d. Theorem 5.1 shows that $f = \delta \circ \gamma(f) = \delta_{\gamma(f)}$. It follows that for every word w of length $> d$, $1 = f(w) = \delta_{\gamma(f)}(w)$. Thus $\gamma(f)$ is a Newton polynomial function of degree at most d. □

6. Pro-p uniformity and pro-p metric

In order to deal with uniformly continuous function, we have chosen to work with *uniformities* (also called *uniform structures*), following Bourbaki [3], rather than with distances, because it is more natural and makes proofs more fluid. The relevant definitions can be found in Appendix A.

6.1. Residually p-finite monoids

Let p be a prime number and let M be a monoid. Let us say that a finite p-group G *separates* two elements of M is there exists a monoid morphism φ from M onto G such that $\varphi(u) \neq \varphi(v)$.

A monoid M is *residually a finite p-group*, or *residually p-finite* for short, if every pair of elements of M can be separated by a finite p-group. Equivalently, a monoid is residually p-finite if it is a subdirect product (in the category of monoids) of finite p-groups.

Since a monoid morphism from a group to another group is a group morphism, this definition is compatible with the standard definition of a residually p-finite group: a group G is *residually p-finite* if, for each $g \neq 1$ in G, there is some finite p-group H and some morphism $G \to H$ whose kernel does not contain g. Equivalently, a group is residually p-finite if the intersection of all its subgroups of index a power of p is trivial.

The following proposition gathers some known facts about residually p-finite monoids.

Proposition 6.1. *The following properties hold:*

1. *a product of residually p-finite monoids is again residually p-finite;*
2. *a submonoid of a residually p-finite monoid is residually p-finite;*
3. *every free monoid and every free group is residually p-finite;*
4. *a finite monoid is residually p-finite if and only if it is a finite p-group.*

6.2. Pro-p uniformity on a residually p-finite monoid

Let M be a residually p-finite monoid. The *pro-p uniformity* on M is the initial uniformity with respect to all monoids morphisms from M to a finite

p-group, equipped with the discrete uniformity. A base of this uniformity is given by Proposition 6.2 below. For each monoid morphism φ from M onto a finite p-group G, let

$$U_\varphi = \{(u, v) \in M \times M \mid \varphi(u) = \varphi(v)\}$$

Since G is finite, U_φ can be written as a finite union

$$U_\varphi = \bigcup_{g \in G} \left(\varphi^{-1}(g) \times \varphi^{-1}(g)\right) \tag{6.1}$$

Proposition 6.2. *Let M be a residually p-finite monoid. The sets of the form U_φ, where φ runs over the class of all monoid morphisms from M onto a finite p-group, form a base of the pro-p uniformity on M.*

Proof. The sets of the form

$$U_{\varphi,P} = (\varphi \times \varphi)^{-1}(P),$$

where P is an entourage of the discrete uniformity on G, form a subbase of the initial uniformity. Note that $U_\varphi = U_{\varphi,D}$ where D is the diagonal of $G \times G$. Since every entourage of the discrete uniformity on G contains P, every set $U_{\varphi,P}$ contains U_φ. It follows that the sets of the form U_φ form another subbase of the initial uniformity. To prove they do in fact form a base, it suffices to prove that if $\varphi_1 : M \to G_1$ and $\varphi_2 : M \to G_2$ are two morphisms from M onto finite p-groups G_1 and G_2, there exists a morphism φ from M onto a finite p-group G such that $U_\varphi \subseteq U_{\varphi_1} \cap U_{\varphi_2}$. Actually, if φ is the morphism $\varphi_1 \times \varphi_2 : M \to G_1 \times G_2$ and $G = \varphi(M)$, then a simple calculation shows that $U_\varphi = U_{\varphi_1} \cap U_{\varphi_2}$. □

From now on, the term *uniform continuity* will always refer to the pro-p uniformity. We let the reader verify the following straightforward results:

Proposition 6.3. *Let M be a residually p-finite monoid. Then the product on M is uniformly continuous.*

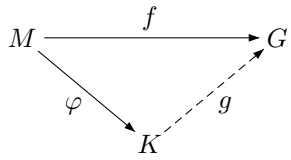
Proposition 6.4. *Let M and N be two residually p-finite monoids. Then every monoid morphism from M to N is uniformly continuous.*

The topology induced by the pro-p uniformity is called the *pro-p topology* on M. It is the initial topology with respect to all monoid morphisms from M onto a discrete finite p-group. Thus the sets of the form $\varphi^{-1}(g)$, where φ is a monoid morphism from M onto a finite p-group G and $g \in G$, form a base of this topology. Since M is residually p-finite, this topology is Hausdorff. It follows from Proposition 6.3 that every residually p-finite monoid is a Hausdorff topological monoid. Applying the standard characterization of initial uniform structures, one gets the following result.

Proposition 6.5. *Let M and N be two residually p-finite monoids. A function $f : M \to N$ is uniformly continuous (respectively continuous) if and only if, for every monoid morphism φ from N onto a finite pro-p group, $\varphi \circ f$ is uniformly continuous (respectively continuous).*

The next proposition gives a purely algebraic characterization of uniformly continuous functions from a residually p-finite monoid to a finite p-group.

Proposition 6.6. *Let M be a residually p-finite monoid. A function f from M to a finite p-group G is uniformly continuous if and only if there exists a monoid morphism φ from M onto a finite p-group K and a map $g\colon K \to G$ such that $f = g \circ \varphi$.*



Proof. Suppose that $f = g\circ\varphi$ for some map $g\colon K \to G$ and some monoid morphism $\varphi\colon M \to K$. Since K and G are finite p-groups, the pro-p uniformity on these groups is the discrete uniformity and thus g is uniformly continuous. Moreover, φ is uniformly continuous by Proposition 6.4, and thus f is uniformly continuous.

Conversely, let $f\colon M \to G$ be a uniformly continuous function. Since the pro-p uniformity on G is the discrete one, there exists an entourage U of M such that for all x, y in M, the condition $(x, y) \in U$ implies $f(x) = f(y)$. It follows that U contains an entourage of the form U_φ, for some morphism φ from M onto a finite p-group K. Consequently, the condition $\varphi(x) = \varphi(y)$, which is equivalent to $(x, y) \in U_\varphi$, implies $f(x) = f(y)$. It follows that f factors through φ. □

6.3. **Sequences and families indexed by A^***

Let X be a topological space. In this paper we use sequences of elements of X, that is, functions from $\mathbb{N}$ to X, but also families of elements of X indexed by A^*, that is, functions from A^* to X. We say that a family $(x_u)_{u\in A^*}$ *converges to x when $|u|$ tends to infinity* if x is a limit point of the map $u \to x_u$ with respect to the filter $\{A^nA^* \mid n \geqslant 0\}$ on A^*. This means that, for each neighborhood V of x, there exists $N > 0$ such that, if $|u| \geqslant N$, then $x_u \in V$.

Let us recall that a sequence $(x_n)_{n\geqslant 0}$ is *ultimately equal* to x if there exists some $r \geqslant 0$ such that, for all $n \geqslant r$, one has $x_n = x$.

Proposition 6.7. *Let M be a residually p-finite monoid. A sequence x_n of elements of M converges to x if and only if, for every monoid morphism φ from N to a finite pro-p group, the sequence $\varphi(x_n)$ is ultimately equal to $\varphi(x)$.*

In particular, one gets the following useful consequence.

Proposition 6.8. *Let M be a residually p-finite monoid. Then for all $x \in M$, $\lim_{n\to\infty} x^{p^n} = 1$.*

Proof. According to Proposition 6.7, it suffices to prove that, for each monoid morphism φ from M to a finite p-group G, the sequence $(\varphi(x))^{p^n}$ tends to $\varphi(1)$, that is, to 1. Since G is a finite p-group of order p^k, $|G|$ divides p^n for all $n \geqslant k$. It follows by Lagrange's theorem that $(\varphi_i(x))^{|G|} = 1$ and hence $(\varphi_i(x))^{p^n} = 1$ for all $n \geqslant |G|$, which proves the result. □

Let us mention a last property, related to function spaces. Let S be a set and let M be a residually p-finite monoid. Let $\mathcal{F}(S, M)$ denote the set of mappings from S to M. For each monoid morphism φ from M to a finite pro-p group, let

$$V_\varphi = \{(f, g) \mid f, g \in \mathcal{F}(S, M) \text{ and } \varphi \circ f = \varphi \circ g\}$$

The sets V_φ form the base of a uniformity on the space $\mathcal{F}(S, M)$, called the *uniformity of uniform convergence*. For a finite p-group G, one can take φ to be the identity and V_φ is the diagonal of $\mathcal{F}(S, G) \times \mathcal{F}(S, G)$. Thus in this case, the uniformity of uniform convergence on $\mathcal{F}(S, G)$ is the discrete one.

Let us recall a classic result; see for instance [2, Chap. X, Remark 3, p. 283].

Proposition 6.9. *If S is a uniform space, then the set of uniformly continuous functions from S to M is closed in $\mathcal{F}(S, M)$. In particular, the unifom limit of a sequence of uniform functions is uniformly continuous.*

The following result is a special case of [2, Chap. X, Proposition 4, p. 278].

Proposition 6.10. *The uniformity of uniform convergence on $\mathcal{F}(S, M)$ is the initial uniformity with respect to the mappings $f \to \varphi \circ f$ from $\mathcal{F}(S, M)$ to $\mathcal{F}(S, G)$, for every monoid morphism φ from M onto a finite p-group G.*

Now, since the uniformity of uniform convergence on $\mathcal{F}(S, G)$ is discrete, one gets the following corollary, which will be used in Section 8.

Corollary 6.11. *Let S be a set and let G be a residually p-finite group.*

1. *A sequence of functions $(f_n : S \to G)_{n \geqslant 0}$ converges uniformly to a function $f : S \to G$ if and only if, for each group morphism from G onto a finite p-group H, the sequence $\varphi \circ f_n$ is ultimately equal to $\varphi \circ f$.*
2. *A family of functions $(f_u : S \to G)_{u \in A^*}$ converges uniformly to the function $f : S \to G$ when $|u|$ tends to infinity if and only if, for each group morphism from G onto a finite p-group H, there exists N such that if $|u| \geqslant N$, then $\varphi \circ f_u = \varphi \circ f$.*

6.4. The metric d_p

Let M be a residually p-finite monoid. One can define a metric d_p on M as follows. Set, for all $u, v \in M$,

$$r_p(u, v) = \max\{n \in \mathbb{N} \cup \{\infty\} \mid \text{no } p\text{-group of order } \leqslant p^n \text{ separates } u \text{ and } v\}.$$

Then, for all u, v in M, the following relations hold:

1. $r_p(u, v) = r_p(v, u)$
2. $r_p(u, w) \geqslant \max\{r_p(u, v), r_p(v, w)\}$

Finally, we put

$$d_p(u, v) = p^{-r_p(u,v)}$$

with the convention $p^{-\infty} = 0$. Then d_p is a metric and even an ultrametric.

For a residually p-finite group H, there is a more convenient way to define d_p. For each $g \in H$, let $v_p(g)$ denote the largest n such that g belongs to the kernel of every morphism from H to a p-group of order p^n. Note that $v_p(g)$ is always finite, except for $g = 1$, in which case it is infinite.

The *pro-p norm* of g is $|g|_p = p^{-v_p(g)}$, with the usual convention $p^{-\infty} = 0$. Note that if $x, y \in H$, then $d_p(x, y) = |x^{-1}y|_p$.

The condition $d_p(x, y) \leqslant p^{-k}$ means that $x^{-1}y$ is in the kernel of each group morphism from G into a p-group of cardinality at most p^k. We leave to the reader to verify that if $G = \mathbb{Z}$, one recovers the usual p-adic valuation, norm and metric.

Another useful example occurs when G is a finite p-group. Recall that the *discrete metric* on a set E is the metric d defined, for all $x, y \in E$ by

$$d(x, y) = \begin{cases} 1 & \text{if } x \neq y, \\ 0 & \text{if } x = y. \end{cases}$$

In this case, the double inequality $d_p(x, y) \leqslant d(x, y) \leqslant |G|\, d_p(x, y)$ shows that the pro-p metric is uniformly equivalent to the discrete metric.

The study of the connection between d_p and the pro-p uniformity brings in some surprises. On one hand, [14, Proposition 3.1] shows that the pro-p uniformity on M is metrizable if and only if, for each finite p-group G, there are only *countably many* morphisms from M onto G. On the other hand, [14, Proposition 3.2][3] shows that the pro-p uniformity on M can be defined by d_p if and only if, for each finite p-group G, there are only *finitely many* morphisms from M onto G.

7. Free monoids and free groups

7.1. **Arbitrary alphabets**

We already mentioned that every free group and every free monoid is residually p-finite. We now identify A^* to a subset of $F(A)$.

Proposition 7.1. *Let A be a set. The pro-p uniformity on A^* is the restriction of the pro-p uniformity on $F(A)$. Furthermore, A^* is dense in $F(A)$.*

Proof. The proof is modelled on that of [16, Proposition 7]. Let G be a finite p-group and let $\varphi\colon F(A) \to G$ be a group morphism. Then the restriction $\varphi_{|A^*}$ of φ to A^* is a monoid morphism and the equality

$$U_\varphi \cap (A^* \times A^*) = U_{\varphi_{|A^*}}$$

shows that the restriction to A^* of the pro-p uniformity of $F(A)$ is a subset of the pro-p uniformity of A^*. To prove the opposite inclusion, it suffices to observe that every monoid morphism $\psi\colon A^* \to G$ extends uniquely to a group morphism $\overline{\psi}\colon F(A) \to G$ for which

$$U_{\overline{\psi}} \cap (A^* \times A^*) = U_{\overline{\psi}_{|A^*}}$$

Let H be the closure of A^* in $F(A)$. Since the closure of a submonoid of a topological monoid is a monoid, H is a submonoid of $F(A)$. Furthermore, Proposition 6.8 implies that for all $x \in H$, $\lim_{n\to\infty} x^{p^n-1} = x^{-1}$. Since H is closed, it follows that $x^{-1} \in H$. Thus H is a subgroup of $F(A)$ containing A, and hence it is equal to $F(A)$. Thus A^* is dense in $F(A)$. □

[3]The definition of d_p given in [14] is actually slightly different, but yields a uniformly equivalent metric.

Proposition 7.2. *Let G be a residually p-finite group. If $f: A^* \to G$ is uniformly continuous, then so is $\Delta^w f$ for every word $w \in A^*$.*

Proof. By induction and by (3.8), it is enough to prove the result for $w = a$, for some letter $a \in A$. In this case, $\Delta^a f: A^* \to G$ is the composition of the following functions:

$$\begin{array}{ccccc} A^* \to A^* \times A^* & & A^* \times A^* \to A^* \times A^* & & A^* \times A^* \to G \times G \\ u \mapsto (u, u) & & (u, v) \mapsto (u, av) & & (u, v) \mapsto (f(u), f(v)) \\ & & G \times G \to G \times G & & G \times G \to G \\ & & (g, h) \mapsto (g^{-1}, h) & & (g, h) \mapsto gh \end{array}$$

as shown by the diagram

$$\begin{aligned} u \mapsto (u, u) \mapsto (u, au) \mapsto (f(u), f(au)) & \\ \mapsto (f(u)^{-1}, f(au)) \mapsto f(u)^{-1} f(au) = \Delta^a f(u). & \end{aligned}$$

Proposition 6.3 shows that the product on A^* (respectively on G) is uniformly continuous. It follows that each of these functions is uniformly continuous and so is their composition. □

7.2. **Finite alphabets, a combinatorial approach**

We will only retain the following consequence of the results stated at the end of Section 6.4.

Proposition 7.3. *For an alphabet A, the following conditions are equivalent:*

(1) *A is finite,*
(2) *the pro-p uniformity on A^* is defined by d_p,*
(3) *the pro-p uniformity on $F(A)$ is defined by d_p.*

When A is finite, the metric d_p can be replaced by a uniformly equivalent metric defined in a purely combinatorial way. Let us define a metric d'_p by setting, for all words $u, v \in A^*$,

$$\begin{aligned} r'_p(u, v) &= \max\Big\{n \;\Big|\; \text{for all } x \in A^n, \binom{u}{x} \equiv \binom{v}{x} \bmod p\Big\} \\ d'_p(u, v) &= p^{-r'_p(u,v)} \end{aligned}$$

It is shown in [9] that d_p and d'_p define the same uniformity. Consequently, one has the following result.

Proposition 7.4. *Let A and B be two finite alphabets. A function $f: A^* \to B^*$ is uniformly continuous if and only if, for all $n > 0$, there exists an $N > 0$ such that, if $\binom{u}{x} \equiv \binom{v}{x} \bmod p$ for all words x of A^* of length $\leqslant N$, then $\binom{f(u)}{z} \equiv \binom{f(v)}{z} \bmod p$ for all words z of B^* length $\leqslant n$.*

Let us illustrate this combinatorial approach by proving that, when A is finite, the truncated noncommutative Magnus transformations introduced in Section 4.3 are uniformly continuous. The proof relies on a combinatorial identity of independent interest, Formula (7.1) below.

We rely on the notation and results introduced at the end of Section 3.1. In particular, recall that the number of occurrences of u as a subword of w, denoted by $\binom{w}{u}$, is also the number of subsets J of $[|w|]$ such that $w[J] = u$.

Given k words $x_1, \dots, x_k$ in A^*, let $E(x_1, \dots, x_k)$ denote the set of all $(k+1)$-tuples $(x, I_1, \dots, I_k)$ such that:

(T_1) x is a word in A^* such that $|x| \leqslant |x_1| + \cdots + |x_k|$;

(T_2) $(I_1 < \cdots < I_k)$ is a chain of subsets of $[|x|]$ whose union is $[|x|]$;

(T_3) $x[I_j] = x_j$ for $1 \leqslant j \leqslant k$.

Remark. The reader may compare this definition with the definition of the *infiltration product* as given in [17, p. 134–135] or in [7, Chap. 6, Section 3]: the infiltration product of k words $x_1, \dots, x_k$ is the sum of all x, the summation being over the set $E(x_1, \dots, x_k)$, except that the condition $I_1 < \cdots < I_k$ is omitted.

Proposition 7.5. *For all words $w, x_1, \dots, x_k$ in A^*, one has*

$$\binom{\mu(w)}{x_1 + \cdots + x_k} = \sum_{(x, I_1, \dots, I_k) \in E(x_1, \dots, x_k)} \binom{w}{x}. \tag{7.1}$$

An example of such a relation is given, for distinct letters a, b, c, by

$$\binom{\mu(w)}{ac + bc} = \binom{w}{acbc} + \binom{w}{abcc} + \binom{w}{bacc} + \binom{w}{abc}.$$

We need in the proof the operation of *standardization*. Let J be the union of a k-chain $(J_1 < \cdots < J_k)$ of finite subsets of $\mathbb{P}$ and let σ be the unique increasing bijection from J to $[|J|]$. The *standardization* of $(J_1 < \cdots < J_k)$ is the k-chain

$$\mathrm{st}(J_1 < \cdots < J_k) = (\sigma(J_1) < \cdots < \sigma(J_k)).$$

For example, if $J = \{1, 2, 4, 7\}$, then $[|J|] = \{1, 2, 3, 4\}$, $\sigma(1) = 1$, $\sigma(2) = 2$, $\sigma(4) = 3$, $\sigma(7) = 4$ and hence $\mathrm{st}(\{1, 7\}, \{2, 4, 7\}) = (\{1, 4\}, \{2, 3, 4\})$. We let the reader verify that the sequence $(\sigma(J_1), \dots, \sigma(J_k))$ is indeed a chain, that is, increasing for $<$, and that the following properties hold:

(P_1) $|\sigma(J_j)| = |J_j|$ for $1 \leqslant j \leqslant k$,

(P_2) if w is a word such that all the sets J_j are subsets of $[|w|]$, then $w[J_j] = (w[J])[\sigma(J_j)]$.

For instance, as a continuation of the previous example, let $w = a_1 \cdots a_7$, $J = \{1, 2, 4, 7\}$, $J_1 = \{1, 7\}$ and $J_2 = \{2, 4, 7\}$. Then $\sigma(J_1) = \{1, 4\}$ and $\sigma(J_2) = \{2, 3, 4\}$. Setting $u = w[J] = a_1 a_2 a_4 a_7$, ($P_2$) states that $w[\{1, 7\}] = u[\{1, 4\}] = a_1 a_7$ and $w[\{2, 4, 7\}] = u[\{2, 3, 4\}] = a_2 a_4 a_7$.

Finally, note that the chain $(J_1 < \cdots < J_k)$ may be recovered from J and $\mathrm{st}(J_1 < \cdots < J_k)$ using the inverse of the bijection σ.

Proof. Proposition 3.1 implies that the left-hand side of (7.1) is equal to $|L|$, where L is the set of k-chains $(J_1 < \cdots < J_k)$ of $[|w|]$ such that $w[J_j] = x_j$ for $1 \leqslant j \leqslant k$. Similarly, the right-hand side of (7.1) is equal to $|R|$, where R is the set of $(k+2)$-tuples $(x, I_1, \dots, I_k, J)$ such that $(x, I_1, \dots, I_k) \in E(x_1, \dots, x_k)$, $J \subseteq [|w|]$ and $w[J] = x$. The rest of the proof consists of finding a bijection from L to R.

Consider the function h which associates to a k-chain $(J_1 < \cdots < J_k)$ of L the $(k+2)$-tuple $(x, I_1, \dots, I_k, J)$ of R defined by

$$J = J_1 \cup \cdots \cup J_k, \ x = w[J] \text{ and } (I_1 < \cdots < I_k) = \mathrm{st}(J_1 < \cdots < J_k).$$

We claim that h is well-defined and is a bijection from L to R.

Setting $h(J_1, \ldots, J_k) = (x, I_1, \ldots, I_k, J)$, one gets $|I_j| = |J_j| = |w[J_j]| = |x_j|$ and by (P_2), $(w[J])[I_j] = w[J_j]$. It follows that (T_3) holds, since, according to the definition of L, $w[J_j] = x_j$. Furthermore, one has

$$|x| = |J| = |J_1 \cup \cdots \cup J_k| \leqslant |J_1| + \cdots + |J_k| = |x_1| + \cdots + |x_k|$$

so that (T_1) is also satisfied. Moreover, (T_2) is a consequence of the definition of standardization. This shows that h is well-defined. It remains to prove that it is bijective.

Suppose that $h(J_1, \ldots, J_k) = (x, I_1, \ldots, I_k, J)$. Since J is known, the standardization may be reversed, and hence $J_1, \ldots, J_k$ are known. Thus h is injective.

To prove the surjectivity of h, consider a $(k+2)$-tuple $(x, I_1, \ldots, I_k, J)$ of R. As previously observed, the standardisation may be reversed since J is known. It follows from (T_2) and from the definition of L that $|J| = |x| = |I_1 \cup \cdots \cup I_k|$. Thus we may find a chain $(J_1 < \cdots < J_k)$ whose standardisation is $(I_1 < \cdots < I_k)$ and such that $J = J_1 \cup \cdots \cup J_k$. Then, for $1 \leqslant j \leqslant k$,

$$w[J_j] = (w[J])[I_j] = x[I_j] = x_j,$$

so that $(I_1, \ldots, I_k)$ belongs to L. Moreover, J is the union of the J_j by construction and $x = w[J]$ since $(x, I_1, \ldots, I_k, J) \in R$. Therefore $h(J_1, \ldots, J_k) = (x, I_1, \ldots, I_k, J)$, which proves that h is surjective. □

Proposition 7.6. *Let A be a finite alphabet. Then, for each $r \geqslant 0$, the function $\mu_r : A^* \to C_r^*$ is uniformly continuous.*

Proof. First recall that C_r is the set of words of A^* of length at most r. Since A is finite, then so is C_r. We claim that the condition

$$\binom{u}{x} \equiv \binom{v}{x} \bmod p \text{ for all words } x \text{ of } A^* \text{ of length } \leqslant rn \tag{7.2}$$

implies

$$\binom{\mu_r(u)}{z} \equiv \binom{\mu_r(v)}{z} \bmod p \text{ for all words } z \text{ of } C_r^* \text{ of length } \leqslant n. \tag{7.3}$$

Let $z = x_1 + \cdots + x_k$, with $k \leqslant n$. By definition of μ_r, one has

$$\binom{\mu_r(u)}{z} = \binom{\mu_r(v)}{z} = 0 \text{ if the length of one of the } x_i\text{'s is larger than } r.$$

Suppose now that, for $1 \leqslant i \leqslant k$, $|x_i| \leqslant r$. If $(x, I_1, \ldots, I_k) \in E(x_1, \ldots, x_k)$, then one has $|x| \leqslant |x_1| + \cdots + |x_k| \leqslant kr \leqslant nr$. It follows by (7.2) that $\binom{u}{x} \equiv \binom{v}{x} \bmod p$. Applying (7.1), one gets

$$\begin{aligned}\binom{\mu_r(u)}{z} &= \sum_{(x, I_1, \ldots, I_k) \in E(x_1, \ldots, x_k)} \binom{u}{x} \\ &\equiv \sum_{(x, I_1, \ldots, I_k) \in E(x_1, \ldots, x_k)} \binom{v}{x} = \binom{\mu_r(v)}{z} \bmod p\end{aligned}$$

It now follows from Proposition 7.4 that μ_r is uniformly continuous. □

7.3. **Sequential product of uniformly continuous functions**

Our goal is to prove Proposition 7.9, the last result of this section. The difficulty is concentrated on an apparently simpler case, which we treat separately.

Proposition 7.7. *Let A be a finite alphabet and let G be a p-finite group. Any sequential product of uniformly continuous functions from A^* to G is uniformly continuous.*

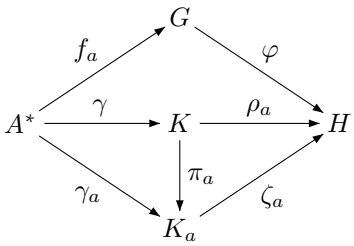
Proof. Let $g \in G$, let $(f_a)_{a\in A}$ be a family of uniformly continuous functions from A^* to G and let $f = \mathrm{Seq}(g, (f_a)_{a\in A})$. According to Proposition 6.6, f is uniformly continuous if and only if, for every morphism φ from G to a finite p-group H, there exists a finite p-group R and a monoid morphism $\rho\colon A^* \to R$ such that $\varphi \circ f$ factors through ρ.

Since each f_a is uniformly continuous, Proposition 6.6 shows that for each $a \in A$, there exists by a finite p-group K_a, a monoid morphism $\gamma_a\colon A^* \to K_a$ and a map $\zeta_a\colon K_a \to H$ such that $\varphi \circ f_a = \zeta_a \circ \gamma_a$. Let γ be the monoid morphism from A^* to $\prod_{a\in A} K_a$ defined by $\gamma(u) = (\gamma_a(u))_{a\in A}$ and let K be the image of γ.

Since A is finite, K is a finite p-group. For each $a \in A$, let $\pi_a\colon K \to K_a$ denote the natural projection, so that $\pi_a \circ \gamma = \gamma_a$. Setting $\rho_a = \zeta_a \circ \pi_a$, one gets

$$\rho_a \circ \gamma = \zeta_a \circ \pi_a \circ \gamma = \zeta_a \circ \gamma_a = \varphi \circ f_a \tag{7.4}$$

The situation is summarized in the following commutative diagram:



Let $H \circ K$ be the wreath product of H by K. Recall that $H \circ K$ is the group with support $H^K \times K$ and product defined, for all $(f_0, u_0), (f_1, u_1) \in H^K \times K$ by

$$(f_0, u_0)(f_1, u_1) = (f, u_0u_1) \text{ where, for all } k \in K,\ f(k) = f_0(k)f_1(ku_0)$$

Let $\rho\colon A \to H \circ K$ be the map defined, for each $a \in A$, by

$$\rho(a) = (\rho_a, \gamma(a)). \tag{7.5}$$

Then ρ extends uniquely to a monoid morphism from A^* to $H \circ K$, also denoted ρ. Since H and K are finite p-groups, then so is $H \circ K$. Now the image F of ρ is a submonoid, and hence a subgroup[4], of $H \circ K$. Thus F is also a finite p-group. We claim that $\varphi \circ f$ factors through ρ. The proof relies on the following lemma:

Lemma 7.8. *Let u be a word of A^* and let $\rho(u) = (\rho_u, \gamma(u))$. Then ρ_u is a map from K to H such that $\varphi(f(u)) = \varphi(g)\rho_u(1)$.*

[4]since for each $x \in M$, $x^{-1} = x^{|F|-1}$.

Proof. Let $u = a_1 \cdots a_n$. According to the definition of the wreath product, and since $\gamma(1) = 1$, $\rho_u : K \to H$ is given by

$$\rho_u(k) = \rho_{a_1}(k\gamma(1))\rho_{a_2}(k\gamma(a_1))\cdots\rho_{a_n}(k\gamma(a_1 \ldots a_{n-1})).$$

It follows that

$$\rho_u(1) = \rho_{a_1}(\gamma(1))\rho_{a_2}(\gamma(a_1))\cdots\rho_{a_n}(\gamma(a_1 \ldots a_{n-1}))$$

Now, since by (7.4), $\rho_a \circ \gamma = \varphi \circ f_a$ for each $a \in A$, one gets

$$\begin{aligned}\rho_u(1) &= (\varphi \circ f_{a_1})(1)\cdots(\varphi \circ f_{a_n})(a_1 \ldots a_{n-1}) \\ &= \varphi\big(f_{a_1}(1)\cdots f_{a_n}(a_1 \ldots a_{n-1})\big)\end{aligned}$$

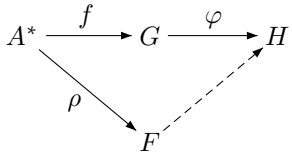
Applying the definition of the sequential product, one obtains

$$f(u) = gf_{a_1}(1)\cdots f_{a_n}(a_1 \ldots a_{n-1})$$

whence

$$\varphi(f(u)) = \varphi\big(gf_{a_1}(1)\cdots f_{a_n}(a_1 \ldots a_{n-1})\big) = \varphi(g)\rho_u(1). \qquad \square$$

We come back to the proof of Proposition 7.7 by proving the claim. Let u and v be words such that $\rho(u) = \rho(v)$. In particular, since $\rho(u) = (\rho_u, \gamma(u))$, one has $\rho_u(1) = \rho_v(1)$, whence $\varphi(g)\rho_u(1) = \varphi(g)\rho_v(1)$. It follows by Lemma 7.8 that $\varphi \circ f(u) = \varphi \circ f(v)$. Thus $\rho(u) = \rho(v)$ implies $\varphi \circ f(u) = \varphi \circ f(v)$ and thus $\varphi \circ f$ factors through ρ. The proposition follows.



$\square$

Proposition 7.9. *Let A be a finite alphabet and let G be a residually p-finite group. Any sequential product of uniformly continuous functions from A^* to G is uniformly continuous.*

Proof. Let $g \in G$, let $(f_a)_{a\in A}$ be a family of uniformly continuous functions from A^* to G and let $f = \mathrm{Seq}(g, (f_a)_{a\in A})$. According to Proposition 6.5, it suffices to prove that, for every morphism φ from G to a finite p-group H, $\varphi \circ f$ is uniformly continuous. Now, Proposition 4.9 shows that $\varphi \circ \mathrm{Seq}(g, (f_a)_{a\in A}) = \mathrm{Seq}(\varphi(g), (\varphi \circ f_a)_{a\in A})$. Thus, by Proposition 7.7, it suffices to prove that each function $\varphi \circ f_a$ is uniformly continuous. But this is clear, since f_a is uniformly continuous by hypothesis and φ is uniformly continuous by Proposition 6.4. $\square$

7.4. Uniform continuity and Newton polynomial functions

The aim of this section is to prove the following theorem.

Theorem 7.10. *Let A be a finite alphabet and let G be a finite p-group. A function $f : A^* \to G$ is uniformly continuous if and only if it is a Newton polynomial function.*

Our proof of Theorem 7.10 is splitted into two halves: one direction is addressed by Proposition 7.11 and the opposite direction by Proposition 7.12.

Proposition 7.11. *Let A be a finite alphabet and let G be a residually p-finite group. Every Newton polynomial function $f\colon A^* \to G$ is uniformly continuous.*

Proof. We prove the result by induction on the degree d of f. If $d \leqslant 0$, then f is a constant function by Proposition 4.1 and hence f is uniformly continuous. Otherwise, Proposition 4.5 shows that f is a sequential product of a family $(f_a)_{a\in A}$ of Newton polynomial functions of degree $\leqslant d-1$. By the induction hypothesis, each f_a is uniformly continuous and hence f is uniformly continuous by Proposition 7.9.

Another possible proof consists in using Corollary 4.15, which states that if f is a Newton polynomial function of degree at most d, then $f = \delta_f^* \circ \mu_d$. Now μ_d is uniformly continuous by Proposition 7.6 and the morphism $\delta_f^* : C_d^* \to F(B)$ is uniformly continuous by Proposition 6.4. Consequently f is uniformly continuous. □

Proposition 7.12. *Let A be a finite alphabet and let G be a finite p-group. If a function $f\colon A^* \to G$ is uniformly continuous, then f is a Newton polynomial function.*

We need several facts about algebras over a field $\mathbb{F}$ (below we use the p-element field $\mathbb{F}_p$). First, if G is a monoid, let $\mathbb{F}[G]$ denote the vector space over $\mathbb{F}$ with basis G. It is an $\mathbb{F}$-algebra, called the *monoid algebra* of G over $\mathbb{F}$. If G is a group, then $\mathbb{F}[G]$ is also called the *group algebra* of G over $\mathbb{F}$. In the particular case where $G = A^*$, it is rather denoted $\mathbb{F}\langle A\rangle$, the algebra of noncommutative polynomials over $\mathbb{F}$. Each monoid morphism from a monoid G_1 into a monoid G_2 extends uniquely, by linearity, to an $\mathbb{F}$-algebra morphism from $\mathbb{F}[G_1]$ into $\mathbb{F}[G_2]$. Similarly, each function from G to $\mathbb{F}$ extends uniquely, by linearity, to a linear form on $\mathbb{F}[G]$.

The vector space of linear forms on an $\mathbb{F}$-algebra R, that is, the *dual* of R, is a left R-module: the action is defined, for all elements x, y in R and each linear form f on R by $(x\cdot f)(y) = f(yx)$. It is indeed a left action: first, $1\cdot f = f$ and $(x_1\cdot(x_2\cdot f))(y) = (x_2\cdot f)(yx_1) = f(yx_1x_2) = ((x_1x_2)\cdot f)(y)$, so that $x_1\cdot(x_2\cdot f) = (x_1x_2)\cdot f$.

Lemma 7.13. *Let f_1, f_2 be linear forms on the $\mathbb{F}$-algebras R_1, R_2 respectively, and let $\zeta\colon R_1 \to R_2$ be an algebra morphism such that $f_2 \circ \zeta = f_1$. Then for each x in R_1, one has $x\cdot f_1 = (\zeta(x)\cdot f_2)\circ\zeta$.*

Proof. For every y in R_1, one has

$$\begin{aligned}(x\cdot f_1)(y) = f_1(yx) = (f_2\circ\zeta)(yx) = f_2\big(\zeta(y)\zeta(x)\big)\\ = \big(\zeta(x)\cdot f_2\big)\big(\zeta(y)\big) = \big((\zeta(x)\cdot f_2)\circ\zeta\big)(y)\end{aligned}$$

and the lemma follows. □

Proof of Proposition 7.12. Let p^r be the order of G. We prove the result by induction on r.

For $r = 1$, G is cyclic of order p and we switch to additive notation. Thus we have to show that $\Delta^w f = \mathbf{0}$ for almost all w. As $G = \mathbb{Z}/p\mathbb{Z}$ is the additive group of the field $\mathbb{F}_p$, we may consider f as a function from A^* to $\mathbb{F}_p$. Since f is uniformly continuous, there exist by Proposition 6.6 a finite p-group H,

a monoid morphism $\zeta\colon A^* \to H$ and a function $\lambda\colon H \to \mathbb{F}_p$ such that $f = \lambda \circ \zeta$, as shown in the left diagram in Figure 2.

We extend by linearity all these functions, as explained previously, and denote these extensions by the same letters. We obtain the diagram on the right hand side of Figure 2. Now ζ is a morphism of $\mathbb{F}_p$-algebra and f, as well as λ, are $\mathbb{F}_p$-linear forms.

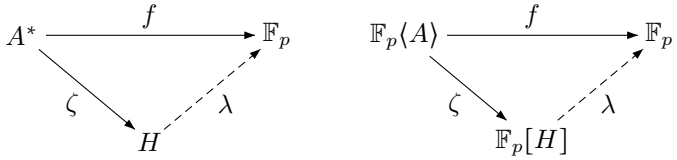


FIGURE 2. Two diagrams.

With these notations, one has $\Delta^a f = (a-1)\cdot f|_{A^*}$, where $\cdot$ denotes the left action of $\mathbb{F}_p\langle A\rangle$ on its dual. Indeed, for each word w in A^*, one has on one hand $\Delta^a f(w) = -f(w) + f(wa)$ and on the other hand

$$((a-1)\cdot f)(w) = f(w(a-1)) = f(wa-w) = f(wa) - f(w). \tag{7.6}$$

Let $w = a_1\cdots a_n$, with $a_1, \ldots, a_n \in A$. Applying Equation (3.8) and the definition of a left action, one gets

$$\Delta^w f = \Delta^{a_1\cdots a_n} f = \big((a_1-1)\cdots(a_n-1)\big)\cdot f|_{A^*}. \tag{7.7}$$

Since $f = \lambda\circ\zeta$ and $\zeta\big((a_1-1)\cdots(a_n-1)\big) = \big(\zeta(a_1)-1\big)\cdots\big(\zeta(a_n)-1\big)$, Lemma 7.13 yields

$$\big((a_1-1)\cdots(a_n-1)\big)\cdot f = \Big(\big((\zeta(a_1)-1)\cdots(\zeta(a_n)-1)\big)\cdot\lambda\Big)\circ\zeta \tag{7.8}$$

Let

$$I_H = \Big\{ \sum_{g\in H} a_g g \mid \sum_{g\in H} a_g = 0 \Big\}$$

be the *augmentation ideal* of $\mathbb{F}_p[H]$. It follows from [6, Proposition VIII.10.4] that if $n \geqslant |H|$, then $I_H^n = 0$. Since every element $\zeta(a_i) - 1$ belongs to I_H, one gets $\big(\zeta(a_1)-1\big)\cdots\big(\zeta(a_n)-1\big) \in I_H^n$ and hence $(\zeta(a_1)-1)\cdots(\zeta(a_n)-1) = 0$. Formulas (7.7) and (7.8) now show that if $n \geqslant |H|$, then $\Delta^w f = \mathbf{0}$, which settles the case $r = 1$.

Suppose now that $r > 1$ and let $f\colon A^* \to G$ be a uniformly continuous function. By a standard result of group theory [19, Theorem 6.5, p. 116], G has a normal subgroup C of order p. Now the quotient map $q\colon G \to G/C$ is uniformly continuous and so is $q\circ f\colon A^* \to G/C$. Since $|G/C| = p^{r-1}$, the induction hypothesis can be applied: there exists n such that $\Delta^v(q\circ f) = \mathbf{1}$ for every word v in A^* of length $\geqslant n$.

Since $\Delta^v(q\circ f) = q\circ(\Delta^v f)$ by Proposition 3.2, one has, for $|v| \geqslant n$, $q\circ(\Delta^v f) = \mathbf{1}$ and hence $\Delta^v f$ maps A^* into C. Note that $\Delta^v f$ is uniformly continuous by Proposition 7.2. Applying the first part of the proof to C, we get the following conclusion: for each v of length $\geqslant n$, there exists n_v such that for each word u of length at least n_v, one has $\Delta^u\Delta^v f = \mathbf{1}$. Let N be the maximum of all n_v taken over the finitely many v of length n. Then for each word w of length at least $N+n$, we may write $w = uv$, with $|v| = n$ and

$|u| \geqslant N \geqslant n_v$. Then $\Delta^w f = \Delta^u \Delta^v f = \mathbf{1}$ and thus f is a Newton polynomial function. □

Note that Proposition 7.11 holds for each residually p-finite group. One may wonder whether Proposition 7.12 can also be extended to this case. As shown Example 8.1 below, this is not the case.

8. Main result

Let us rephrase Theorem 1.2 of the introduction in a sligthly more general setting, which is the main result of this paper.

Theorem 8.1. *Let A be a finite alphabet and let f be a function from A^* to a residually p-finite group G. The following conditions are equivalent:*

1. *f is uniformly continuous for the pro-p uniformity,*
2. *the functions $\Delta^w f$, where $w \in A^*$, tend uniformly to 1 when $|w|$ tends to ∞,*
3. *the elements $\delta_w f$, where $w \in A^*$, tend to 1 when $|w|$ tends to ∞,*
4. *f is the uniform limit of the sequence $(f_r)_{r\geqslant 0}$ of its Newton polynomial functions.*

Proof. (1) ⇒ (2). Let $f\colon A^* \to G$ be a uniformly continuous function and let φ be a group morphism from G onto a finite p-group H. Since φ is uniformly continuous, so is $\varphi \circ f$. It follows by Proposition 7.12 that $\varphi \circ f$ is a Newton polynomial function and thus $\Delta^w(\varphi \circ f) = 1$ for almost all $w \in A^*$. Therefore $\varphi \circ (\Delta^w f) = 1$ by Proposition 3.2 and thus (2) holds by Corollary 6.11.

(2) ⇒ (3) is clear, since $\delta_w f = \Delta^w f(1)$.

(3) ⇒ (4). First, Proposition 4.14 states that the functions f_r are Newton polynomial functions. Let φ be a group morphism from G onto a finite p-group H. By hypothesis there exists N such that for each word u of length $> N$, $\delta_u f \in \mathrm{Ker}(\varphi)$. We show that for every $r \geqslant N$, one has $\varphi \circ f_r = \varphi \circ f$, and then deduce (4) from Corollary 6.11.

Let $w \in A^*$. If $|w| \leqslant N$, then $\mu(w) = \mu_r(w)$, so that $f(w) = f_r(w)$ by Theorem 3.6 and by the definition of f_r, hence $\varphi \circ f_r(w) = \varphi \circ f(w)$.

If $|w| \geqslant N$, then $\mu_N(w) = v_1 + \ldots + v_k$ for some $v_i \in C_N$. Moreover, $\mu_r(w) = x_0 + v_1 + x_1 + \ldots + v_k + x_k$, where each x_i is a sum of words u of length $> N$. For each such word u, one has $\delta_f^*(u) = \delta_u f \in \mathrm{Ker}(\varphi)$ and thus $\delta_f^*(x_i) \in \mathrm{Ker}(\varphi)$. Similarly $\mu(w) = y_0 + v_1 + y_1 + \ldots + v_n + y_n$, where $\delta_f^*(y_i) \in \mathrm{Ker}\,\varphi$. It follows that

$$\begin{aligned}\varphi \circ f_r(w) &= \varphi \circ \delta_f^*(\mu_r(w)) = \varphi \circ \delta_f^*(\mu_N(w)) \\ &= \varphi \circ \delta_f^*(\mu(w)) = \varphi \circ f(w)\end{aligned}$$

and thus $\varphi \circ f_r = \varphi \circ f$.

(4) ⇒ (1) follows from Proposition 6.9, since Newton polynomial functions are uniformly continuous by Proposition 7.11. □

Example 8.1. We come back once again to the function f considered in Examples 2.1, 3.5 and 3.8, except that we now see f as a function from $\{0,1,2\}^*$ to the free group on $\{0,1,2\}$.

Proposition 3.4 shows that f is not a Newton polynomial function, since $\delta_{1^n} f \neq 1$ for all n. However, f is uniformly continuous for d_2. One way to see this is to use the implication (3) $\Rightarrow$ (1) of Theorem 8.1. Indeed, one has $\delta_0 f = 0$, $\delta_1 f = 0$, $\delta_2 f = 1$, $\delta_{1^n 0} f = (0^{-1}1)^{2^{n-1}(-1)^{n-1}}$, $\delta_{1^n 1} f = (0^{-1}2)^{2^{n-1}(-1)^{n-1}}$, $\delta_{1^n 2} f = (1^{-1}2)^{2^{n-1}(-1)^{n-1}}$ and $\delta_u f = \epsilon$ in all other cases. It now follows from Proposition 6.8 that $\delta_w f$ tends to ε when $|w|$ tends to ∞.

Another way to prove this would be to adapt the results of [18]. These results are stated for groups instead of p-groups but can be readily adapted to this latter case. They show that if the transition monoid of the minimal sequential transducer realising a function is a p-group, then this function is uniformly continuous for the metric d_p. In our case, $p = 2$ and the transition monoid of the transducer of f is the cyclic group of order 2. It follows that f is uniformly continuous for the metric d_2.

Example 8.2. Proposition 7.6 shows that, for each $r \geqslant 0$, μ_r is uniformly continuous. However, the function μ is not uniformly continuous. Indeed, consider a function f from A^* to a residually p-finite group G. Newton's Forward Difference Formula shows that $f = \delta_f^* \circ \mu$. Now δ_f^* is a monoid morphism, and hence is uniformly continuous by Proposition 6.4. Thus, if μ was uniformly continuous, then any function f would also be uniformly continuous.

Another way to prove that μ is not uniformly continuous is to use Theorem 8.1. Indeed, we have seen in Example 3.9 that $\delta_w \mu = w$ for all $w \in A^*$. Since these elements do not tend to 1 when $|w|$ tends to infinity, μ is not uniformly continuous.

9. Applications

We conclude this article by giving two consequences of our results. We first consider an interpolation problem in Section 9.1. Section 9.2 is devoted to applications to formal language theory, a topic that originally motivated the authors to study pro-p uniformities [9, 12, 11, 16, 18].

9.1. An interpolation problem

It follows from Theorem 1.1 that for each sequence $(c_n)_{n\geqslant 0}$ of integers, there exists a unique function $f = \mathbb{N} \to \mathbb{Z}$ such that $\delta_n f = c_n$. Moreover, this function is uniformly continuous for d_p if and only if $\lim_{n\to\infty} |c_n|_p = 0$.

Similarly, if A and B are finite alphabets, for each function $c\colon A^* \to F(B)$, there exists a unique function $f\colon A^* \to F(B)$ such that, for all $u \in A^*$, $\delta_u f = c_u$. This function is defined by $f(u) = c(\mu(u))$ for all $u \in A^*$. Moreover, it follows from Theorem 8.1 that this function is uniformly continuous for d_p if and only if $|c(u)|_p$ tends to 1 when $|w|$ tends to ∞.

Mahler's original paper [8] concerned functions of a p-adic variable. His results make it possible, to solve the following interpolation problem: is it possible to extend a function from $\mathbb{N}$ to $\mathbb{Z}$ into a continuous function from $\mathbb{Z}_p$ to $\mathbb{Z}_p$? In our noncommutative setting, we replace $\mathbb{N}$ by A^*, where A is a finite alphabet. In this case, the completion of A^* for the pro-p uniformity is the free pro-p group $F_p(A)$ on A and the problem can be formulated as follows.

Interpolation problem. *Given a function $f\colon A^* \to F(B)$, does there exist a continuous function $F_p(A) \to F_p(B)$ which extends f?*

Theorem 8.1 gives the answer to this question.

Proposition 9.1. *Let $f\colon A^* \to F(B)$ be a function. Then f extends to a continuous function from $F_p(A)$ to $F_p(B)$ if and only if the elements $\delta_w f$, where $w \in A^*$, tend to 1 when $|w|$ tends to ∞. In this case, this extension is unique.*

Proof. Suppose that f admits a continuous extension $\widehat{f}$ from $F_p(A)$ to $F_p(B)$. Since $F_p(A)$ and $F_p(B)$ are not only complete, but also compact, $\widehat{f}$ is uniformly continuous. It follows that f is uniformly continuous, and by Theorem 8.1, the elements $\delta_w f$ tend to 1 when $|w|$ tends to ∞.

Suppose now that the elements $\delta_w f$ tend to 1 when $|w|$ tends to ∞. By Theorem 8.1, f is uniformly continuous and since the embedding map from $F(B)$ into $F_p(B)$ is also uniformly continuous, f can be seen as a uniformly continuous map from A^* to $F_p(B)$. Since A^* is dense in $F_p(A)$ and since $F_p(B)$ is a complete uniform space, f admits a unique uniformly continuous extension from $F_p(A)$ to $F_p(B)$. □

9.2. **Formal languages**

We come back in this section to the problem that originally motivated this research. Let us first recall some definitions.

Let A be a finite alphabet. A subset of A^* is usually called a *language*, as it is a set of words. A language L of A^* is *recognized* by a monoid morphism $\varphi\colon A^* \to M$ if there exists a subset P of M such that $L = \varphi^{-1}(P)$. By extension, one also says that M *recognizes* L if there exists a monoid morphism $\varphi\colon A^* \to M$ that recognizes L.

A language is *recognizable* or *regular* if it can be recognized by a finite monoid. It is a *p-group language* if it can be recognized by a finite p-group. It is not difficult to show that regular languages (respectively p-group languages) are closed under Boolean operations, which comprise finite intersection, finite union and complement. Let $\mathcal{G}_p$ denote the class of p-group languages.

Examples of p-group languages include, for each word v and $0 \leqslant r < p$, the languages

$$L(v, r) = \left\{ w \in A^* \mid \binom{w}{v} \equiv r \bmod p \right\}.$$

first introduced by Eilenberg in [6, p. 239]. It is convenient to call these languages *binomial languages*. Eilenberg proved the following result

Theorem 9.2 (Eilenberg)**.** *A language is a p-group language if and only if it is a Boolean combination of binomial languages.*

A function f from A^* to B^* is *regularity-preserving* if, for each regular language L of B^*, the language $f^{-1}(L)$ is also regular[5]. More generally, if $\mathcal{C}$ is a class of regular languages, f is *$\mathcal{C}$-preserving* if, for each language L of $\mathcal{C}$,

[5]It would probably be more appropriate to say that f^{-1} is regularity-preserving, but we preferred to stick to an already well-established terminology.

the language $f^{-1}(L)$ is also in $\mathcal{C}$. The problem mentioned at the beginning of this section is as follows:

Synthesis problem. *Describe the class of $\mathcal{C}$-preserving functions.*

For instance, although several families of regularity-preserving functions have been identified, the synthesis problem for these functions is still a major open problem. In a series of papers [12, 13, 14], Silva and the first author addressed this problem when $\mathcal{C}$ is a *variety of languages*, in the sense of Eilenberg [6]. In particular, one has:

Proposition 9.3. [15, Prop. 1.3 and Theorem 1.4] *A function is $\mathcal{G}_p$-preserving if and only if it is uniformly continuous for d_p.*

In the case of *sequential* and *rational* functions, $\mathcal{C}$-preserving functions were investigated by Schützenberger and the second author [18]. Another characterization of $\mathcal{G}_p$-functions using profinite equations was obtained in [4, Lemma 4], but it only holds for regular-preserving functions and the next example shows that a $\mathcal{G}_p$-preserving function is not necessarily regular-preserving.

Example 9.1. Let $f{:}\mathbb{N} \to \mathbb{N}$ be the function defined as follows: $f(0) = 0$ and if $n > 0$, the binary representation of $f(n)$ is obtained from that of n by replacing the rightmost bit 1 by 0. For instance, since the binary representation of 26 is 11010, the binary representation of $f(26)$ is 11000, and hence $f(26) = 24$. We let the reader verify that, for all $n, m \in \mathbb{N}$, $d_2(f(n), f(m)) \leqslant d_2(n, m)$ and hence f is uniformly continuous. It follows that f is $\mathcal{G}_2$-preserving. However, it is not regularity-preserving: $\{0\}$ is a regular language, but $f^{-1}(0) = \{0\} \cup \{2^n \mid n \geqslant 0\}$ is not regular.

Our results are of a different nature, since they concern all $\mathcal{G}_p$-preserving functions. Indeed, Theorem 8.1 allows us to solve the synthesis problem for $\mathcal{G}_p$ in the following way:

Theorem 9.4. *The class of $\mathcal{G}_p$-preserving functions is the smallest set of functions containing the constant functions and which is closed under taking sequential products and uniform limits.*

Proof. It follows from Proposition 9.3 that function is $\mathcal{G}_p$-preserving if and only if it is uniformly continuous for d_p. Therefore it suffices to prove that the set of uniformly continuous functions is equal to the smallest set S of functions containing the constant functions and closed under taking sequential products and uniform limits.

Constant functions are uniformly continuous. Furthermore, Proposition 7.9 states that every sequential product of uniformly continuous functions is uniformly continuous and Proposition 6.9 shows that the unifom limit of a sequence of uniform functions is uniformly continuous. It follows that every function of S is uniformly continuous.

Let now f be a uniformly continuous function. By Theorem 8.1, f is the uniform limit of the sequence of its Newton polynomial functions. Moreover, Corollary 4.6 shows that the smallest set of functions containing the constant functions and closed under sequential product is the set of Newton polynomial functions. It follows that f belongs to S. □

There is a counterpart of Proposition 9.3 for regularity preserving functions: a function is regularity preserving if and only if it is uniformly continuous for the profinite uniformity, which is the initial uniformity with respect to all monoid morphisms from A^* onto a finite monoid. However, there is no known counterpart of Theorem 9.4 for these functions.

Appendix A. Uniform spaces

Readers are referred to [3] for an introduction to uniform spaces.

Let X be a set. The subsets of $X \times X$ can be viewed as relations on X. In particular, if U and V are subsets of $X \times X$, we use the notation UV to denote the composition of the two relations, that is, the set

$$UV = \{(x, y) \in X \times X \mid \text{there exists } z \in X,\ (x, z) \in U \text{ and } (z, y) \in V\}.$$

Given a relation U, the *inverse relation* of U is the relation

$$U^{-1} = \{(x, y) \in X \times X \mid (y, x) \in U\}$$

A relation U is *symmetrical* if $U^{-1} = U$. Finally, if $x \in X$ and $U \subseteq X \times X$, we write $U(x)$ for the set $\{y \in X \mid (x, y) \in U\}$.

A *uniformity* (or *uniform structure*) on a set X is a nonempty set $\mathcal{U}$ of subsets of $X \times X$ satisfying the following properties:

- (U_1) if a subset U of $X \times X$ contains an element of $\mathcal{U}$, then $U \in \mathcal{U}$,
- (U_2) the intersection of any two elements of $\mathcal{U}$ contains an element of $\mathcal{U}$,
- (U_3) each element of $\mathcal{U}$ contains the diagonal of $X \times X$,
- (U_4) for each $U \in \mathcal{U}$, $U^{-1} \in \mathcal{U}$,
- (U_5) for each $U \in \mathcal{U}$, there exists $V \in \mathcal{U}$ such that $VV \subseteq U$.

If $\mathcal{U}$ is a uniformity on the set X, the elements of $\mathcal{U}$ are called *entourages*. Note that $X \times X$ is always an entourage. The pair $(X,\mathcal{U})$ (or the set X if $\mathcal{U}$ is understood) is called a *uniform space*.

For each $x \in X$, let $\mathcal{U}(x) = \{U(x) \mid U \in \mathcal{U}\}$. There exists a unique topology on X, called the *topology induced* by $\mathcal{U}$, for which $\mathcal{U}(x)$ is the filter of neighborhoods of x for each $x \in X$. A uniform space $(X,\mathcal{U})$ is *Hausdorff* if the induced topology is Hausdorff. This is equivalent to requiring that the intersection of all the entourages of $\mathcal{U}$ is equal to the diagonal of $X \times X$.

A *base of a uniformity* $\mathcal{U}$ is a subset $\mathcal{B}$ of $\mathcal{U}$ such that each element of $\mathcal{U}$ contains an element of $\mathcal{B}$. In particular, $\mathcal{U}$ consists of all the relations on X containing an element of $\mathcal{B}$. We say that $\mathcal{U}$ is *generated* by $\mathcal{B}$. A set $\mathcal{B}$ of subsets of $X \times X$ is a base of some uniformity if and only if it satisfies properties (U_2), (U_3), (U_5) and (U_6):

- (U_6) for each $U \in \mathcal{B}$, there exists $U' \in \mathcal{B}$ such that $U' \subseteq U^{-1}$.

A *subbase* of a uniformity $\mathcal{U}$ is a subset $\mathcal{B}$ of $\mathcal{U}$ such that the finite intersections of members of $\mathcal{B}$ form a base of $\mathcal{U}$.

The *product* of two uniform spaces $(X_1,\mathcal{U}_1)$ and $(X_2,\mathcal{U}_2)$ is the uniform space $(X_1 \times X_2,\mathcal{U})$, where $\mathcal{U}$ is the uniformity generated by the base consisting of the entourages of $X_1 \times X_2$ of the form

$$\{((x_1, x_2), (y_1, y_2)) \mid (x_1, y_1) \in U_1, (x_2, y_2) \in U_2\}$$

where U_1 is an entourage of X_1 and U_2 is an entourage of X_2.

If $(X,\mathcal{U})$ and $(Y,\mathcal{V})$ are uniform spaces, a function $f\colon X \to Y$ is said to be *uniformly continuous* if, for each entourage V of $\mathcal{V}$, $(f \times f)^{-1}(V)$ is an entourage of $\mathcal{U}$, or, equivalenty, there exists an entourage $U \in \mathcal{U}$ such that the condition $(x, y) \in U$ implies $(f(x), f(y)) \in V$.

Let X be a set, $(X_i, \mathcal{V}_i)_{i\in I}$ a family of uniform spaces, and for each $i \in I$, a function $f_i\colon X \to X_i$. The *initial uniformity* on X with respect to the family $(f_i)_{i\in I}$ is the coarsest uniformity $\mathcal{V}$ on X such that each f_i is uniformly continuous. The sets of the form $(f_i \times f_i)^{-1}(V_i)$, where V_i is an entourage of X_i for some $i \in I$, form a subbase of $\mathcal{V}$.

Acknowledgements

We would like to thank Mathieu Guay-Pacquet for suggesting a simple algorithm to compute our noncommutative version of the Magnus transformation.

References

[1] B. Banaschewski and E. Nelson, On the non-existence of injective near-ring modules, *Canad. Math. Bull.* **20**,1 (1977), 17–23.

[2] N. Bourbaki, *General topology. Chapters 5–10*, *Elements of Mathematics (Berlin)*, Springer-Verlag, Berlin, 1989.

[3] N. Bourbaki, *General topology. Chapters 1–4*, *Elements of Mathematics (Berlin)*, Springer-Verlag, Berlin, 1998.

[4] M. Cadilhac, O. Carton and C. Paperman, Continuity and rational functions, in *44th International Colloquium on Automata, Languages, and Programming*, pp. Art. No. 115, 14, *LIPIcs. Leibniz Int. Proc. Inform.* vol. 80, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2017.

[5] C. Choffrut, Minimizing subsequential transducers: a survey, *Theoret. Comput. Sci.* **292**,1 (2003), 131–143. Selected papers in honor of Jean Berstel.

[6] S. Eilenberg, *Automata, Languages and Machines*, vol. B, Academic Press, New York, 1976.

[7] M. Lothaire, *Combinatorics on words*, *Cambridge Mathematical Library*, Cambridge University Press, Cambridge, 1997.

[8] K. Mahler, An interpolation series for continuous functions of a p-adic variable., *J. Reine Angew. Math.* **199** (1958), 23–34. Correction **208** (1961), 70–72.

[9] J.-É. Pin, Topologie p-adique sur les mots, *Journal de théorie des nombres de Bordeaux* **5** (1993), 263–281.

[10] J.-É. Pin, Newton's forward difference equation for functions from words to words, in *Evolving Computability*, A. Beckmann, V. Mitrana and M. Soskova (eds.), pp. 71–82, *Lect. Notes Comp. Sci.* vol. 9136, Springer International Publishing, 2015.

[11] J.-É. Pin and C. Reutenauer, A Mahler's theorem for word functions, in *46th International Colloquium on Automata, Languages, and Programming*, pp. Art. No. 125, 13, *LIPIcs. Leibniz Int. Proc. Inform.* vol. 132, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2019.

[12] J.-É. Pin and P. V. Silva, A topological approach to transductions, *Theoret. Comput. Sci.* **340** (2005), 443–456.

[13] J.-É. Pin and P. V. Silva, A Mahler's theorem for functions from words to integers, in *25th International Symposium on Theoretical Aspects of Computer Science (STACS 2008)*, S. Albers and P. Weil (eds.), Dagstuhl, Germany, 2008, pp. 585–596, Internationales Begegnungs- Und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.

[14] J.-É. Pin and P. V. Silva, On profinite uniform structures defined by varieties of finite monoids, *Internat. J. Algebra Comput.* **21** (2011), 295–314.

[15] J.-É. Pin and P. V. Silva, A noncommutative extension of Mahler's theorem on interpolation series, *European J. Combin.* **36** (2014), 564–578.
[16] C. Reutenauer, Une topologie du monoïde libre, *Semigroup Forum* **18**,1 (1979), 33–49.
[17] C. Reutenauer, *Free Lie algebras*, *London Mathematical Society Monographs. New Series* vol. 7, The Clarendon Press, Oxford University Press, New York, 1993. Oxford Science Publications.
[18] C. Reutenauer and M.-P. Schützenberger, Variétés et fonctions rationnelles, *Theoret. Comput. Sci.* **145**,1-2 (1995), 229–240.
[19] H. E. Rose, *A course on finite groups*, *Universitext*, Springer-Verlag London, Ltd., London, 2009.
[20] J. Sakarovitch, *Elements of automata theory*, Cambridge University Press, Cambridge, 2009. Translated from the 2003 French original by Reuben Thomas.

Jean-Éric Pin: IRIF, Université de Paris et CNRS - Case 7014 - F-75205 Paris Cedex 13 FRANCE.
Email address: Jean-Eric.Pin@irif.fr
URL: https://www.irif.fr/~jep/indexAnglais.html

Christophe Reutenauer: Mathématiques, Université du Québec à Montréal, Montréal, CP 8888, succ. Centre Ville, Canada H3C 3P8
Email address: reutenauer.christophe@uqam.ca
URL: http://www.lacim.uqam.ca/~christo/