

A Representation Theorem for (q-)Holonomic Sequences

T. Kotek¹, J. A. Makowsky²

*Department of Computer Science
Technion–Israel Institute of Technology
3200 Haifa, Israel
{tkotek,janos}@cs.technion.ac.il*

Abstract

Chomsky and Schützenberger showed in 1963 that the sequence $d_L(n)$, which counts the number of words of a given length n in a regular language L , satisfies a linear recurrence relation with constant coefficients for n , i.e., it is C-finite. It follows that every sequence $s(n)$ which satisfies a linear recurrence relation with constant coefficients can be represented as $d_{L_1}(n) - d_{L_2}(n)$ for two regular languages. We view this as a representation theorem for C-finite sequences.

Holonomic or P-recursive sequences are sequences which satisfy a linear recurrence relation with polynomial coefficients. q -holonomic sequences are the q -analog of holonomic sequences.

In this paper we prove representation theorems of holonomic and q -holonomic sequences based on position specific weights on words, and for holonomic sequences, without using weights, based on sparse regular languages.

Keywords: Holonomic sequences, q -holonomic sequences, positional weights on words, Regular languages, Monadic Second Order Logic
05A15, 05A05, 68R15

1. Introduction

1.1. Holonomic sequences

In this paper we study sequences $a(n)$ of natural numbers or integers which arise in combinatorics. Many such sequences satisfy linear recurrence relations with constant coefficients, or with coefficients which are polynomials in n . The former are called C-finite, and the latter are called holonomic (or P-recursive). There is also a q -analog for holonomic sequences, the q -holonomic sequences, which arose first in the 19th century in the context of hypergeometric sequences. Note that hypergeometric sequences are holonomic.

There is a substantial theory of how to verify and prove identities among the terms of $a(n)$, see [PWZ96, Koo93].

¹Partially supported by grants of the Graduate School of the Technion–Israel Institute of Technology of the Fein Foundation

²Partially supported by a grant of the Fund for Promotion of Research of the Technion–Israel Institute of Technology and grant ISF 1392/07 of the Israel Science Foundation (2007-2011)

We are interested in the case where $a(n)$ admits a *combinatorial* or a *logical* interpretation, i.e., $a(n)$ counts the number of some relations or functions on the set $[n] = \{1, \dots, n\}$ or on the ordered structure $\langle [n], <_{nat} \rangle$ which have a certain property possibly definable in some logical formalism (with or without its natural order). Here $<_{nat}$ denotes the natural order on integers. In this paper we are mostly interested in the case where the underlying structure is a fixed linear order on $[n]$. E. Specker asked in [Spe88] whether a combinatorial interpretation can be found for special holonomic sequences, i.e., holonomic sequences where the leading coefficient in the recurrence relation is 1. We shall discuss Specker's question in detail in Section 6.

A simple example for a C-finite sequence is $a(n) = 2^n$ with recurrence $a(n) = 2 \cdot a(n-1)$ and $a(0) = 1$. $a(n)$ counts the number of subsets of the set $[n]$, and it also counts the number of unary predicates on $\langle [n], <_{nat} \rangle$, which can be viewed as the number of binary words of size n . Similarly, the Catalan numbers satisfy the recurrence relation $(n+2)C(n+1) = 2(2n+1) \cdot C(n)$, hence $C(n)$ is holonomic. They are known to count the number of words of size $2n$ of the Dyck language of balanced parentheses. An example of a q -holonomic sequence is given by the q -derangement numbers, which count the number of automorphisms of a vector space V which do not fix any vector in V except for 0. Here V is an n -dimensional vector space over the finite field of q elements, cf. [CR92].

The exact framework needed for the paper is described in Sections 2 and 3.

1.2. Representation theorems

It was observed in the literature [CS63, SS78] that sequences $a_L(n)$ which count the number of words of size n in a regular language L are C-finite, i.e. they satisfy a linear recurrence relation with constant coefficients. Conversely, every C-finite sequence $a(n)$ can be obtained as the difference $a_{L_1}(n) - a_{L_2}(n)$ where L_1, L_2 are two regular languages.

Similarly, sequences $b_L(n)$ which count the number of words of size n in a context-free language L are P-recursive or holonomic, i.e., they satisfy a linear recurrence relation with coefficients in $\mathbb{Z}[x]$. However, using a growth argument one can easily see that there are P-recursive sequences which cannot be obtained by counting words in any (not necessarily context-free) language.

In [KM10], a characterization of P-recursive sequences was given counting special classes of lattice paths.

In this paper we want to put the result from [KM10] into a more general context which allows us also to characterize q -holonomic sequences.

We augment languages over a fixed alphabet by introducing positionally dependent weight for occurrences of letters.

We want to count weighted words in a language L . The weight is defined by a position specific scoring matrix, widely used in computational biology to search DNA and protein databases for sequence similarities, cf. [SSGE82, AMS⁺97]. This approach is also reminiscent to counting weighted homomorphisms, cf. [BCL⁺06]. Position-independent weights on words were used in [NZ99] for the extension of the powerful (and so far under-utilized) Goulden-Jackson Cluster method for finding the generating function for the number of words avoiding, as factors, the members of a prescribed set. Our main result shows that by suitably choosing the weights one can characterize both holonomic and q -holonomic sequences. The exact statement of these characterizations are given in Sections 4 and 5.

The characterization of holonomic sequences given in [KM10] does not use weights, but counts combinations of regular languages with lattice paths. Our second main result shows how

to replace lattice paths by more general structures. We look at structures $\langle [n], <_{nat}, R_1, \dots, R_k \rangle$ where each R_i is a binary relation.

We show in Section 5 that for every holonomic sequence $a(n)$ there is a k and a formula ϕ in $\mathbf{MSOL}(R_1, \dots, R_k)$ such that $a(n)$ counts the number of relations R_i on $\langle [n], <_{nat} \rangle$ such that $\langle [n], <_{nat}, R_1, \dots, R_k \rangle \models \phi$. The converse is easily seen not to be true using a growth argument, see Remark 16. However, we can identify a subset $HOL - MSOL$ of $\mathbf{MSOL}(R_1, \dots, R_k)$ which captures exactly holonomicity.

1.3. Outline

In Section 2 we give the background on recurrence relations. In Section 3 we give the background on logic and regular languages. In Section 4 we state and prove our first representation theorem for holonomic and q -holonomic sequences using positionally weighted languages. In Section 5 we state and prove our second representation theorem for holonomic sequences using sparse regular languages instead of positionally weighted languages. Finally, in Section 6 we draw our conclusions and state open problems.

2. Linear Recurrences

We are in particular interested in linear recurrence relations which may hold over \mathbb{Z} or \mathbb{Z}_m .

Definition 1 (Recurrence relations). *Given a sequence $a(n)$ of integers we say $a(n)$ is*

- (i) *C-finite or rational if there is a fixed $r \in \mathbb{N} \setminus \{0\}$ for which $a(n)$ satisfies for all $n > r$*

$$a(n+r) = \sum_{i=0}^{r-1} p_i a(n+i)$$

where each $p_i \in \mathbb{Z}$.

- (ii) *P-recursive or holonomic if there is a fixed $r \in \mathbb{N} \setminus \{0\}$ for which $a(n)$ satisfies for all $n > r$*

$$p_r(n) \cdot a(n+r) = \sum_{i=0}^{r-1} p_i(n) a(n+i)$$

where each p_i is a polynomial in $\mathbb{Z}[x]$ and $p_r(n) \neq 0$ for any n . We call it simply P-recursive or SP-recursive, if additionally $p_r(n) = 1$ for every $n \in \mathbb{Z}$.

- (iii) *hypergeometric if $a(n)$ satisfies for all $n > 2$*

$$p_1(n) \cdot a(n+1) = p_0(n) a(n)$$

where each p_i is a polynomial in $\mathbb{Z}[x]$ and $p_1(n) \neq 0$ for any n . In other words, $a(n)$ is P-recursive with $q = 1$.

- (iv) *Let q be a formal parameter. $a_q(n)$ is q -holonomic if there is a fixed $r \in \mathbb{N} \setminus \{0\}$ for which $a_q(n)$ satisfies for all $n > r$*

$$e_r(n) \cdot a_q(n+r) = \sum_{i=0}^{r-1} e_i(n) a_q(n+i)$$

where each e_i is a polynomial in $\mathbb{Q}[q^x]$, cf. [WZ92].

The terminology C-finite and holonomic are due to [Zei90]. P-recursive is due to [Sta80]. P-recursive sequences were already studied in [Bir30, BT33]. We use both terms P-recursive and holonomic interchangeably. P-recursive sequences also arise in the literature as the sequences of coefficients of *differentiably finite (D-finite)* power series, cf. [Sta80].

The following are well known, see [FS09, EvPSW03].

- Lemma 2.** (i) *Let $a(n)$ be C-finite. Then there is a constant $c \in \mathbb{Z}$ such that $a(n) \leq 2^{cn}$.*
(ii) *Furthermore, for every holonomic sequence $a(n)$ there is a constant $\gamma \in \mathbb{N}$ such that $|a(n)| \leq n!^\gamma$ for all $n \geq 2$.*
(iii) *The sets of C-finite, SP-recursive, P-recursive (holonomic) and q-holonomic sequences are closed under addition, subtraction and point-wise multiplication.*

In general, the bound on the growth rate of holonomic sequences is best possible, since $a(n) = n!^m$ is easily seen to be holonomic for integer m , [Ger04].

Proposition 3. *Let $a(n)$ be a function $a : \mathbb{N} \rightarrow \mathbb{Z}$.*

- (i) *If $a(n)$ is C-finite then $a(n)$ is SP-recursive.*
(ii) *If $a(n)$ is SP-recursive then $a(n)$ is P-recursive.*
(iii) *If $a(n)$ is hypergeometric then $a(n)$ is P-recursive.*
(iv) *Every holonomic sequence $b(n)$ can be obtained from a q-holonomic sequence $a_q(n)$ by a suitable limit process.*

Moreover, the converses of (i), (ii), (iii) and (iv) do not hold.

Proof. The implications follow from the definitions.

$n!$ is SP-recursive, but not C-finite, as, by Lemma 2 it grows too fast. The Catalan numbers are P-recursive (holonomic) by the recurrence relation $(n+2)C(n+1) = 2(2n+1) \cdot C(n)$. They are not SP-recursive, because SP-recursive sequences satisfy modular recurrence relations, but the Catalan numbers fail to do so already modulo 2, cf. [Spe88].

The derangement numbers $D(n)$ count the number of permutations of $[n]$ without fixed points. They satisfy the recurrence relation $D(n) = (n-1)(D(n-1) + D(n-2))$, so they are SP-recursive. They are not hypergeometric, cf. [PWZ96]. For (iv), see [Ber09]. \square

3. Some Background From Logic and Regular Languages

In this section we introduce the necessary background from logic and regular languages.

3.1. Definability in MSOL

MSOL, an extension of First Order Logic, **FOL**, is defined as follows. We denote by ρ a vocabulary, i.e., a set of relation symbols. The formulas of **MSOL**(ρ) are defined like the ones of **FOL**, with the addition that we allow countably many variables for unary relation symbols U_i for $i \in \mathbb{N}$ called *set variables* and quantification over these. We refer the reader to [EFT94] for further details. We say a class \mathcal{K} of structures of vocabulary ρ is *definable* in **MSOL** if there exists $\phi \in \mathbf{MSOL}(\rho)$ such that $\mathcal{K} = \{\mathcal{A} : \mathcal{A} \models \phi\}$ i.e., \mathcal{K} is the set of ρ -structures which satisfy ϕ .

For $t \in \mathbb{N}$, let ρ_t be the vocabulary which consists of t unary relations and one binary relation. The binary relation will always be interpreted as a linear order of the universe. The class of

structures of vocabulary ρ_t for which the unary relations in ρ_t form a partition of the universe is **MSOL**-definable by a sentence $\phi_{\text{Partition}}$.

For a unary relation U over $[n]$ we can interpret $\langle [n], <_{\text{nat}}, U \rangle$ as a binary word where position i is occupied by letter 1 if $i \in U$ and by letter 0 if $i \in [n] - U$. Similarly, for a vocabulary ρ_t consisting of t unary relation symbols and $<_{\text{nat}}$, where the relations $\bar{U} = (U_1, \dots, U_t)$ over $[n]$ form a partition of $[n]$, we can interpret $\langle [n], <_{\text{nat}}, U_1, \dots, U_t \rangle$ as a word over an alphabet of size t .

We state here precisely the connection between regular languages and **MSOL**, due to R. Büchi, C. Elgot and B. Trakhtenbrot, cf. [Lib04, EF95]:

Theorem 4. *Let \mathcal{L} be a language. Then \mathcal{L} is regular iff \mathcal{L} is definable in **MSOL**.*

3.2. Ordered disjoint unions

For a set C of natural numbers and a natural number c , we denote $C + c = \{x + c : x \in C\}$. Let $\mathcal{A}_i = \langle [n_i], <_{\text{nat}_i}, U_{1,i}, \dots, U_{t,i} \rangle$ for $i = 1, 2$ be ρ_t -structures where $<_{\text{nat}_i}$ is the natural order on $[n_i]$. Then the *ordered disjoint union* $\mathcal{B} = \mathcal{A}_1 \sqcup_{<} \mathcal{A}_2$ is a ρ_t -structure $\mathcal{B} = \langle [n], <_{\text{nat}_B}, U_{1,B}, \dots, U_{t,B} \rangle$ where $n = n_1 + n_2$, $U_{j,B} = U_{j,1} \cup (U_{j,2} + n_1)$ for $j = 1, \dots, t$, and the relation $<_{\text{nat}_B}$ is the natural order on $[n]$.

The following is well known [EF95]:

Proposition 5 (Hintikka sentences). *Let ρ be a vocabulary. For every $\xi \in \mathbb{N}$ there is a finite set*

$$\Theta^\xi = \{\theta_1, \dots, \theta_\alpha\}$$

*of **MSOL**(ρ)-sentences of quantifier rank ξ such that*

- *Every $\theta \in \Theta^\xi$ has a model.*
- *The conjunction of any two sentences $\theta_1, \theta_2 \in \Theta^\xi$ is unsatisfiable.*
- *every **MSOL**(ρ) sentence ψ of quantifier rank ξ is equivalent to exactly one finite disjunction of sentences in Θ^ξ .*

For fixed quantifier rank ξ and a ρ -structure \mathcal{A} we denote by $\theta(\xi, \mathcal{A})$ the unique sentence $\theta \in \Theta^\xi$ such that $\mathcal{A} \models \theta$. If the quantifier rank is clear from the context, we omit it and write $\theta(\mathcal{A})$. $\theta(\xi, \mathcal{A})$ is called the Hintikka sentence of \mathcal{A} of quantifier rank ξ .

Proposition 6. *Let $\mathcal{A}_1, \mathcal{A}_2$ be ρ_t -structures given by*

$$\mathcal{A} = \langle [n_i], <_{\text{nat}_i}, U_{1,i}, \dots, U_{t,i} \rangle$$

where $<_{\text{nat}_i}$ is a linear order of $[n_i]$. Then $\theta(\xi, \mathcal{A}_1 \sqcup_{<} \mathcal{A}_2) \in \mathbf{MSOL}(\rho_t)$ depends only on $\theta(\xi, \mathcal{A}_1) \in \mathbf{MSOL}(\rho_t)$ and $\theta(\xi, \mathcal{A}_2) \in \mathbf{MSOL}(\rho_t)$.

3.3. Bounded and sparse regular languages

Definition 7. *A language L is bounded if there exist $e \in \mathbb{N}$ and words v_1, \dots, v_e such that $L \subseteq v_1^* \cdots v_e^*$. A language L is sparse if the number of words in L of size n , $a_L(n)$, is bounded from above by a polynomial.*

It is well-known and often rediscovered that a regular language L is bounded iff L is *sparse*. [GKRS08]. The earliest reference may be [Tro81]. F. D'Alessandro, B. Intrigila and S. Varrichio, [DIV06], proved the following theorem:

Theorem 8. *Let L be a regular language which is bounded. There exist polynomials $p_0(x), \dots, p_{y-1}(x) \in \mathbb{Q}[x]$ and $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$,*

$$a_L(n) = p_{(n \bmod y)}(n),$$

where $(n \bmod y)$ stands for the number ℓ in $\{0, \dots, y-1\}$ for which it holds that $\ell \equiv n \pmod{y}$.

In fact, Theorem 8 holds also for context-free languages.

4. Positionally Weighted Words and Our Main Theorem

4.1. Characterizing (q) -holonomic sequences using positional weights

Let Σ be an alphabet of size $t = |\Sigma|$. We denote by $w[i]$ the i -th letter, and by $|w|$ the length of the word $w \in \Sigma^*$. For each $s \in \Sigma$, let $\alpha_s(i)$ be a function $\alpha_s : \mathbb{N} \rightarrow \mathbb{Q}$. We define the *weight of a word* $w \in \Sigma^*$ by

$$\text{weight}_{\bar{\alpha}}(w) = \prod_{i=1}^{|w|} \alpha_{w[i]}(i) \quad (1)$$

and for each language $L \subseteq \Sigma^*$ its *positionally weighted counting* by

$$d_{L, \bar{\alpha}}(n) = \sum_{w \in L, |w|=n} \text{weight}_{\bar{\alpha}}(w), \quad (2)$$

where the summation is over all words of length n in L .

Definition 9. *Let $\bar{\alpha} = (\alpha_s(x) : s \in \Sigma)$ be a tuple where each $\alpha_s(x)$ is a function $\alpha_s : \mathbb{N} \rightarrow \mathbb{Q}$. A sequence of integers $a(n)$ has a PW-interpretation for $\bar{\alpha}$ if there is a regular language $L \subseteq \Sigma^*$, such that $a(n) = d_{L, \bar{\alpha}}(n)$.*

If $\alpha_s(n) = 1$ for all $s \in \Sigma$ and $n \in \mathbb{N}$ then the function $d_{L, \bar{\alpha}}(n) = d_L(n)$ is the usual counting function for L .

Theorem 10 (Main Theorem). *Let $a(n)$ be a sequence of integers.*

- (i) *$a(n)$ has a PW-interpretation for polynomials $\alpha_s(x) \in \mathbb{Z}[x]$ iff $a(n)$ is SP-recursive.*
- (ii) *$a(n)$ has a PW-interpretation for rational functions $\alpha_s(x) \in \mathbb{Q}(x)$ iff $a(n)$ is P-recursive.*
- (iii) *$a(n)$ has a PW-interpretation for exponential polynomials $\alpha_s(x) \in \mathbb{Q}[q^x]$ iff $a(n)$ is q -Holonomic.*

The proof of Theorem 10 is given in Subsections 4.2 and 4.3.

4.2. Proof of Theorem 10: From recurrences to PW-interpretations

Let p_0, \dots, p_r be elements of $\mathbb{Z}[x]$ or $\mathbb{Q}[q^x]$ and let $a(n)$ satisfy the recurrence

$$p_0(n) \cdot a(n+q) = \sum_{i=1}^q p_i(n) a(n+i). \quad (3)$$

The initial conditions of the recurrence are $a(1), \dots, a(q)$. If $a(n)$ is SP-recursive then the recurrence in Equation (3) is such that $p_0(x)$ is identically 1. We will show that $a(n)$ has a PW-interpretation as indicated in the theorem.

We will prove this direction by looking at the recurrence tree of the recurrence given in Equation (3). The recurrence tree is a rooted tree in which every vertex is labeled with an element from $[n]$. Every vertex in the recurrence tree has degree 0 or r . The root is labeled n and the leafs are labeled elements of $[r]$ (and only the leafs are labeled as such). The difference between the label of a vertex v and its i -th child is i .

Let us look at a path $\bar{v} = (v_0, \dots, v_y)$ in the recurrence tree of $a(n)$ from the root to a leaf. Let $\bar{\pi} = (\pi_1, \dots, \pi_y)$ be a tuple of $[r]$ elements such that the sequence of labels on the path is $(n, n - \pi_1, \dots, n - (\pi_1 + \dots + \pi_y))$. In other words, v_i is the π_i -th child of v_{i-1} . Each such path \bar{v} corresponds to successive application of the recurrence formula $a(n) \rightarrow a(n - \pi_1) \rightarrow a(n - (\pi_1 + \pi_2)) \rightarrow \dots \rightarrow a(n - (\pi_1 + \dots + \pi_y))$, where the last step is an initial condition. At each step $a(n - (\pi_1 + \dots + \pi_{z-1})) \rightarrow a(n - (\pi_1 + \dots + \pi_z))$ of the successive application of the recurrence formula we multiply by the coefficient $p_{\pi_z}(n - (\pi_1 + \dots + \pi_{z-1}))$ and we divide by $p_0(n - (\pi_1 + \dots + \pi_{z-1}))$. Finally we multiply by an initial condition, which is the label of the leaf v_y . The value of the path is the product of the relevant coefficients in the recurrence formula, given as

$$val(\bar{v}) = a(n - (\pi_1 + \dots + \pi_y)) \prod_{z=1}^y \frac{p_{\pi_z}(n - (\pi_1 + \dots + \pi_{z-1}))}{p_0(n - (\pi_1 + \dots + \pi_{z-1}))}$$

$a(n)$ is then the sum of the values $val(\bar{v})$ over all the paths \bar{v} from the root to a leaf in the recurrence tree of $a(n)$.

We will use unary relations $U_1, \dots, U_r \subseteq [n] - [r]$ to encode the path and $U_{r+1}, \dots, U_{2r} \subseteq [r]$ to encode the initial condition. The relations U_i for $i \in [r]$ will correspond to the choices in the path. That is, if the path contains the step $a(n') \rightarrow a(n' - i)$ then n' will belong to U_i , and $n' - 1, \dots, n' - (i - 1)$ will not belong to any of the U_i , $i > 0$. Note that $j \in U_i$ implies $j \notin U_k$ for $k \in [r] - \{i\}$. If the path ends in the initial condition $a(i)$ for $i \in [r]$ then $U_{r+i} = \{i\}$ and $U_{r+k} = \emptyset$ for every $k \in [r] - \{i\}$. The relation U_0 will be used to contain those elements of $[n]$ not in $\bigcup_{i \in [2r]} U_i$. Thus, the tuple (U_0, \dots, U_{2r}) will form a partition of $[n]$. The sequence $a(n)$ is given by

$$a(n) = \sum_{\langle [n], \langle_{nat}, U_0, \dots, U_{2r} \rangle \rangle} \prod_{i=0}^r \prod_{j \in U_i} \frac{p_i(j)}{p_0(j)} \cdot \prod_{i=q+1}^{2r} \prod_{j \in U_i} a(i)$$

where the summation is over all τ_k -structures $\langle [n], \langle_{nat}, U_0, \dots, U_{2r} \rangle \rangle$ which satisfy the following conditions:

- (i) (U_0, \dots, U_{2r}) form a partition of $[n]$
- (ii) for $i \in [r]$, $U_i \cap [r] = \emptyset$
- (iii) for $i \in [2r] - [r]$, $U_i \cap ([n] - [r]) = \emptyset$
- (iv) if $n' \in U_i$ for some $i \in [2r] - [r]$, then $(\bigcup_{i=r+1}^{2r} U_i) \cap [n' - 1] = \emptyset$

(v) for $n' \in [n] - [r]$, if $n' \in U_i$ for $i \in [r]$ then $\{n' - 1, \dots, n' - (i - 1)\} \subseteq U_0$ and $n' - i \notin U_0$

Condition (v) requires that for a recursive application $a(n') \rightarrow a(n' - i)$, all of the intermediate values between n' and $n' - i$ cannot appear in U_1, \dots, U_{2r} , and that the next recursion step $n' - i$ must appear in U_1, \dots, U_{2r} . Condition (iv) says the path stops once we reach an initial condition.

The above conditions on (U_0, \dots, U_{2r}) are easily seen to be definable in **MSOL** in the presence of the natural order $<_{nat}$ on $[n]$ by some sentence ϕ . Let $\alpha_i(x) = \frac{p_i(x)}{p_0(x)}$ for $i \in [r]$. Let $\alpha_{r+i}(x) \in \mathbb{Z}[x]$ for $i \in [r]$ be the polynomials defined by $p_{r+i}(n) = a(i)$ for all n .

Recall that we denote by τ_r the vocabulary which consists of a binary relation and r' unary relation symbols. By changing the order of the products $a(n)$ can be written equivalently as

$$a(n) = \sum_{\langle [n], <_{nat}, U_0, \dots, U_{2r} \rangle : \phi} \prod_{j=0}^n \left(\prod_{i: j \in U_i} \alpha_i(j) \right). \quad (4)$$

The inner product in Equation (4) is over those $i \in [r]$ such that j belongs to U_i . The summation in Equation (4) is over all τ_{2r+1} -structures $\mathcal{M} = \langle [n], <_{nat}, U_0, \dots, U_{2r} \rangle$ with universe $[n]$ in which the binary relation is interpreted as the natural order $<_{nat}$ on $[n]$ and which satisfy the sentence ϕ , $\mathcal{M} \models \phi$. By Theorem 4, there exists a regular language such that

$$a(n) = d_{L, \bar{\alpha}}(n).$$

If Equation (3) is an SP-recurrence we have that $p_0(x)$ is identically 1. Thus, $\frac{p_i(j)}{p_0(j)}$ in Equation (4.2) are polynomials in j over \mathbb{Z} . If Equation (3) is a P-recurrence, then $\frac{p_i(j)}{p_0(j)}$ in Equation (4.2) are rational functions in j over \mathbb{Q} . Finally, if Equation (3) is a q-holonomic recurrence, then $\frac{p_i(j)}{p_0(j)}$ in Equation (4.2) are elements of field of functions $\mathbb{Q}(q^j)$. Setting $t = 2r + 1$ the first direction of the theorem is done.

4.3. Proof of Theorem 10: From PW-interpretations to recurrences

We now want to show how to convert a PW-interpretation into a P-recurrence or a q-holonomic recurrence. We use the decomposition properties of **MSOL** to compute a scheme of recurrence relations for a finite number of sequences, and then extract a P-recurrence for the desired sequence.

Let \mathcal{R} be one of $\mathbb{Z}[x]$, $\mathbb{Q}(x)$ and $\mathbb{Q}(q^x)$. Let $a(n)$ be a sequence of integers which has a PW-interpretation for $\bar{\alpha} = (\alpha_s : s \in \Sigma)$ where all the $\alpha_s(x)$ are in \mathcal{R} .

Let $r = |\Sigma|$. For any formula $\psi \in \mathbf{MSOL}(\tau_r)$, let the sequence $a_\psi(n)$ be

$$a_\psi(n) = \sum_{\langle [n], <_{nat}, U_1, \dots, U_r \rangle : \psi} \prod_{i=1}^r \prod_{j \in [n]: j \in U_i} \alpha_i(j)$$

where the sum is over all τ_r -structures $\langle [n], <_{nat}, U_1, \dots, U_r \rangle$ which satisfy ψ , where $<_{nat}$ is the natural order on $[n]$. By theorem 4 there exists a formula $\phi \in \mathbf{MSOL}(\tau_r)$ such that Equation (2) can be rewritten as $a(n) = a_\phi(n)$. Let ρ be the quantifier rank of ϕ and let $\Theta = \{\theta_1, \dots, \theta_{|\Theta|}\}$ be the finite set of Hintikka sentences of quantifier rank ρ from Proposition 5. As ϕ is equivalent to the disjunction

$$\bigvee_{\theta \in \Theta, \theta \models \phi} \theta$$

and the formulas in Θ are pairwise not satisfiable, we have

$$a(n) = \sum_{\theta \in \Theta, \theta \neq \phi} a_\theta(n)$$

By the closure of \mathcal{R} to finite sum, we need only show that each a_θ is P-recursive, SP-recursive or q-holonomic (depending on whether $\mathcal{R} = \mathbb{Z}[x]$, $\mathcal{R} = \mathbb{Q}(x)$ or $\mathcal{R} = \mathbb{Q}(q^x)$).

Denote by STR_1 the set of τ_r structures with universe $[1] = \{1\}$. Every τ_r -structure \mathcal{A} with universe $[n]$ is given uniquely as the ordered disjoint union of two structures \mathcal{B} and \mathcal{C} of size $n-1$ and 1 respectively. Thus, we now write $a_\theta(n)$ as

$$a_\theta(n) = \sum_{C \in STR_1} \prod_{i \in [r]: 1 \in U_i^C} \alpha_i(n) \sum_{\mathcal{B}: \mathcal{B} \sqcup_{<} C = \theta} \prod_{i=1}^r \prod_{j \in [n-1]: j \in U_i^{\mathcal{B}}} \alpha_i(j)$$

where $C = \langle [1], <_{nat}, U_1^C, \dots, U_r^C \rangle$ and $\mathcal{B} = \langle [n-1], <_{nat}, U_1^{\mathcal{B}}, <_{nat}, \dots, U_r^{\mathcal{B}} \rangle$, and the inner summation is over those τ_r -structures \mathcal{B} of size $n-1$ such that $\mathcal{B} \sqcup_{<} C$ satisfies θ . Notice

$$\hat{a}_C(n) = \prod_{i \in [r]: 1 \in U_i^C} \alpha_i(n)$$

belongs to \mathcal{R} .

Proposition 6, for the case where \mathcal{A}_2 is a one element structure now states:

There exists a function $\gamma : \Theta \times STR_1 \rightarrow \Theta$ such that for every τ_r -structure \mathcal{B} with Hintikka sentence $\theta(\mathcal{B})$ and every $C \in STR_1(\tau_r)$, $\mathcal{B} \sqcup_{<} C$ satisfies $\gamma(\theta(\mathcal{B}), C)$.

Therefore, we may use recursively the values $a_{\theta_i}(n)$ for $\theta(\mathcal{B})$,

$$a_\theta(n) = \sum_{(\chi \in \Theta, C): \gamma(\chi, C) = \theta} \hat{a}_C(n) a_\chi(n)$$

where the summation is over pairs $(\chi, C) \in \Theta \times STR_1(\tau_r)$ such that $\gamma(\chi, C) = \theta$. So we have

$$a_\theta(n) = \sum_{\chi \in \Theta} \sum_{C: \gamma(\chi, C) = \theta} \hat{a}_C(n) a_\chi(n-1)$$

Let $p_{\theta, \chi}(n) = \sum_{C: \gamma(\chi, C) = \theta} \hat{a}_C(n)$. Since STR_1 is of fixed cardinality, $p_{\theta, \chi}(x)$ belongs to \mathcal{R} . Thus, $a_\theta(n)$ is given by the recurrence

$$a_\theta(n) = \sum_{\chi \in \Theta} p_{\theta, \chi}(n) a_\chi(n-1) \quad (5)$$

Since Θ is finite, Equation (5) can be written in matrix form for a finite $|\Theta| \times |\Theta|$ -matrix D as follows. Let $D = (d_{i,j})$ be the matrix defined as $d_{i,j} = p_{\theta_i, \theta_j}(n)$ and $\bar{b}(n) = (a_{\theta_1}(n), \dots, a_{\theta_{|\Theta|}}(n))^T$. Then

$$\bar{b}(n) = D\bar{b}(n-1)$$

Let $char(x) = \sum_{k=0}^{|\Theta|} \delta_k(n) x^k$ be the characteristic polynomial of D . The Cayley-Hamilton theorem holds $\mathbb{Z}[x]$, for $\mathbb{Q}(x)$ and for $\mathbb{Q}(q^x)$. Hence the matrix D over \mathcal{R} satisfies

$$\sum_{k=0}^{|\Theta|} \delta_k(n) D^k = 0$$

where each $\delta_k(x)$ is in \mathcal{R} . Multiplying by $\bar{b}(n)$ we get

$$\sum_{k=0}^{|\Theta|} \delta_k(n) \bar{b}(n+k) = 0$$

Thus, each $b_i(n) = a_{\theta_i}(n)$ satisfies a linear recurrence relation with coefficients in \mathcal{R} . Finally, for the case of SP-recursive $a(n)$ we need also note that $\delta_{|\Theta|}(x)$ is identically 1.

5. A Weightless Representation Theorem For Holonomic Sequences

5.1. Characterizing holonomic sequences using sparse structures

Let τ_k be the vocabulary consisting of k binary relation symbols.

Definition 11. Let $\mathcal{A} = \langle [n], <_{nat}, R_1, \dots, R_k \rangle$ be a τ_k -structure. We say \mathcal{A} is diagonal if R_1, \dots, R_k form a partition of

$$\{(i, j) \in [n] \times [n] \mid i \leq j\}.$$

Moreover, we say a diagonal \mathcal{A} is sparse if for every $i \in [k]$, $j \in [n]$ and $m < j$, if $m \in R_i$ then $m+1$ belongs to exactly one of R_i, \dots, R_k .

Definition 12. Let ϕ and ψ be $\mathbf{MSOL}(\rho_k)$ sentences. A class of sparse diagonal τ_k -structures C is $\langle \phi, \psi \rangle$ -expressible if C is the class of all τ_k -structures $\langle [n], <_{nat}, R_1, \dots, R_l \rangle$ which satisfy the following:

- It holds that

$$\langle [n], <_{nat}, R_1(1, -), \dots, R_k(1, -) \rangle \models \phi,$$

where $R_i(1, -) = R_i \cap \{(1, j) \mid j \in [n]\}$.

- For each j , it holds that

$$\langle [j], <_{nat}, R_1(-, j), \dots, R_k(-, j) \rangle \models \psi,$$

where $R_i(-, j) = R_i \cap \{(\ell, j) \mid \ell \in [n]\}$.

If there exist such ϕ and ψ , we say C is \mathbf{MSOL} -expressible.

Theorem 15 below states that P-recursive sequences can be characterized in terms of counting sequences of $\langle \phi, \psi \rangle$ -expressible classes of sparse diagonal τ_k -structures. To do so we use Lemma 13 and Lemma 14, whose proofs will be given in Subsection 5.2 and Subsection 5.3 respectively.

Let C be a class of finite structures. For the rest of this section, we denote by $a_C(n)$ the number of structures of size n in C . In particular, for languages L it holds that $a_L(n) = d_L(n)$.

Lemma 13. Let ϕ and ψ be $\mathbf{MSOL}(\rho_k)$ sentences. Let C be a class of sparse diagonal τ_k -structures which is $\langle \phi, \psi \rangle$ -expressible. There exist an alphabet Σ and a tuple $\bar{\alpha} = (\alpha_s : s \in \Sigma)$ of polynomials $\alpha_s(x) \in \mathbb{Q}[x]$ such that the counting sequence $a_C(n)$ has a PW-interpretation for $\bar{\alpha}$.

Lemma 14. Let Σ be an alphabet and let $\alpha_s(x)$ be polynomials in $\mathbb{Z}[x]$ for each $s \in \Sigma$. If $a(n)$ has a PW-interpretation for $\bar{\alpha} = (\alpha_s : s \in \Sigma)$, then there exist two classes C_1 and C_2 of sparse diagonal τ_k -structures which are both \mathbf{MSOL} -expressible such that $a(n) = a_{C_1}(n) - a_{C_2}(n)$.

Theorem 15. *Let $a(n)$ be a sequence of integers. The sequence $a(n)$ is P-recursive iff there exist two classes of sparse diagonal τ_k -structures C_1 and C_2 and a polynomial $r(x) \in \mathbb{Z}[x]$ such that*

$$a(n) = \frac{a_{C_1}(n) - a_{C_2}(n)}{\prod_{j=1}^n r(j)}. \quad (6)$$

Proof. Let $a(n)$ be a P-recursive sequence of integers. By Theorem 10 there exist an alphabet Σ , a regular language L and rational functions $\alpha_s(x) \in \mathbb{Q}(x)$ for every $s \in \Sigma$ such that $a(n)$ is a PW-interpretation for \bar{a} , i.e. $a(n) = d_{L, \bar{a}}(n)$. Let $p_s(x)$ and $r_s(x)$ be the numerator and the denominator, respectively, of $\alpha_s(x)$ for each $s \in \Sigma$. We may assume $p_s(x)$ and $r_s(x)$ are relatively prime and are polynomials over \mathbb{Z} . Let $p'_s(x) = p_s(x) \cdot \prod_{t \in \Sigma - \{s\}} r_t(x)$ and let $r(x) = \prod_{t \in \Sigma} r_t(x)$. It holds that $\alpha_s(x) = \frac{p'_s(x)}{r(x)}$. Let $b(n)$ be

$$b(n) = \sum_{w \in L, |w|=n} \prod_{j=1}^n p'_{w[j]}(j).$$

The sequence $b(n)$ has a PW-interpretation for $\bar{p}' = (p'_s(x) : s \in \Sigma)$ and it holds that

$$a(n) = \frac{1}{\prod_{j=1}^n r(j)} \cdot b(n).$$

Therefore, it follows from Lemma 14 that there exist two MSOL-expressible classes of sparse diagonal τ_k -structures, C_1 and C_2 , for which that Equation (6) holds.

Conversely, assume Equation (6) holds. By Lemma 13, there exist two alphabets Σ_1 and Σ_2 , two regular languages L_1 and L_2 over Σ_1 and Σ_2 respectively, and polynomials $\beta_s(x) \in \mathbb{Q}[x]$ for each $s \in \Sigma_1$ and $\gamma_t(x) \in \mathbb{Q}[x]$ for each $t \in \Sigma_2$, such that

$$a(n) = \frac{d_{L_1, \bar{\beta}}(n) - d_{L_2, \bar{\gamma}}(n)}{\prod_{j=1}^n r(j)}.$$

Let $\Sigma_2 = \{t_1, \dots, t_\mu\}$ and let Σ_3 be a disjoint copy of Σ_2 , $\Sigma_3 = \{\tilde{t}_1, \dots, \tilde{t}_\mu\}$. We may further assume $\Sigma_1 \cap (\Sigma_2 \cup \Sigma_3) = \emptyset$. Let $\Sigma = \Sigma_1 \cup \Sigma_2 \cup \Sigma_3$. Let $\tilde{L}_2 = \{w\tilde{t}_i \mid w t_i \in L_2\}$. Therefore $L_1 \cap \tilde{L}_2 = \emptyset$. Clearly it holds that $d_{L_2, \bar{\gamma}}(n) = d_{\tilde{L}_2, \bar{\tilde{\gamma}}}(n)$. We define $\delta_\sigma(x) \in \mathbb{Q}(x)$ for every $\sigma \in \Sigma$ as follows. Let $\delta_\sigma(x) = \beta_\sigma(x)$ if $\sigma \in \Sigma_1$, $\delta_\sigma(x) = \gamma_\sigma(x)$ if $\sigma \in \Sigma_2$, and $\delta_\sigma(x) = -\gamma_\sigma(x)$ if $\sigma \in \Sigma_3$. It holds that

$$a(n) = \frac{d_{L_1 \cup \tilde{L}_2, \bar{\delta}}(n)}{\prod_{j=1}^n r(j)}.$$

Letting $\delta'_\sigma(x) = \frac{\delta_\sigma(x)}{r(x)}$ for every $\sigma \in \Sigma$ we get that $a(n) = d_{L_1 \cup \tilde{L}_2, \bar{\delta}'}(n)$, i.e. $a(n)$ has a PW-interpretation for $\bar{\delta}'$, where $\delta'_\sigma(x)$ for every σ is a rational function over \mathbb{Q} . By Theorem 10, $a(n)$ is P-recursive. \square

Remark 16. *We want to show that both assumptions, diagonal and sparse are needed for Theorem 15.*

- (i) *Let C be a class of sparse diagonal τ_k -structures which is $\langle \phi, \psi \rangle$ -expressible with $\phi, \psi \in \text{MSOL}(\rho_k)$. It is easy to see that C is also $\text{MSOL}(\tau_k)$ -definable.*

- (ii) On the other hand, let \mathcal{D}_{func} be the class of τ_1 -structures such that R_1 is a function. Note that $a_{\mathcal{D}_{func}}(n) = n^n$. \mathcal{D}_{func} is **MSOL**(τ_k)-definable, but $a_{\mathcal{D}_{func}}(n) = n^n$ is not P-recursive [Ger04]. Notice \mathcal{D}_{func} is also not diagonal.
- (iii) Let the class \mathcal{D}_{diag} which consists of all diagonal τ_1 -structures. \mathcal{D}_{diag} is not sparse. Furthermore $a_{\mathcal{D}_{diag}}(n) = 2^{\binom{n+1}{2}}$. Again, the class \mathcal{D}_{diag} is easily seen to be **MSOL**-definable. However, by Lemma 2, the sequence $a_{\mathcal{D}_{diag}}(n)$ is not P-recursive.

5.2. Proof of Lemma 13

In this subsection we prove Lemma 13, which states:

Let C be a class of sparse diagonal τ_k -structures which is **MSOL**-expressible. There exist an alphabet Σ and a tuple $\vec{\alpha} = (\alpha_s : s \in \Sigma)$ of polynomials $\alpha_s(x) \in \mathbb{Q}[x]$ such that the counting sequence $a_C(n)$ has a PW-interpretation for $\vec{\alpha}$.

Let $\mathcal{M} = \langle [n], <_{nat}, R_1, \dots, R_k \rangle$ be a structure from C . As discussed before Theorem 4, we can regard words as structures which consist of a universe, a linear order on the universe and unary relations which partition the universe. Let $\omega(\mathcal{M})$ be a tuple of words $\omega(\mathcal{M}) = (w, u_1, u_2, \dots, u_n)$ over alphabet $[k]$ which satisfies:

- $w = \langle [n], R_1(1, -), \dots, R_k(1, -) \rangle$,
- for each i , $u_i = \langle [i], R_1(-, i), \dots, R_k(-, i) \rangle$, and
- $u_i[1] = w[i]$ for each i .

Let \mathcal{D} be given by

$$\mathcal{D} = \{\omega(\mathcal{M}) \mid \mathcal{M} \in C\}$$

Notice $a_C(n) = a_{\mathcal{D}}(n)$ since \mathcal{M} is determined uniquely by $\omega(\mathcal{M})$.

By Theorem 4 there exist regular languages L_1 and L_2 such that

$$\mathcal{D} = \{(w, u_1, \dots, u_{|w|}) \mid w \in L_1, \text{ and for every } i, |u_i| = i \text{ and } u_i \in L_2\}.$$

It is not hard to see that L_2 is a bounded language. Moreover, for every $\sigma \in \Sigma$, $L_{2,\sigma} = L_2 \cap (\sigma \cdot \Sigma^*)$ is regular and bounded.

We may rewrite $a_{\mathcal{D}}(n)$ as

$$a_{\mathcal{D}}(n) = \sum_{w \in L_1, |w|=n} \sum_{(u_1, \dots, u_n)} 1,$$

where the inner summation is over all tuples (u_1, \dots, u_n) such that for each i , $u_i \in L_{2,w[i]}$ and $|u_i| = i$. Since each $L_{2,\sigma}$ is bounded and regular, there exist for each $L_{2,\sigma}$ a natural number y_σ and polynomials $p_{(n \bmod y_\sigma), \sigma}(x) \in \mathbb{Q}[x]$ as guaranteed in Theorem 8. We may assume w.l.o.g that there exists y such that $y = y_\sigma$ for every $\sigma \in \Sigma$ by taking y to be the product of all the y_σ . Since the u_i do not depend on each other,

$$a_{\mathcal{D}}(n) = \sum_{w \in L_1, |w|=n} \prod_{j=1}^n p_{(n \bmod y_{w[j]}), w[j]}(j). \quad (7)$$

Let $\widehat{\Sigma}$ be the disjoint union of y disjoint copies of Σ ,

$$\widehat{\Sigma} = \bigcup_{z=1}^y \{\sigma^{(z)} \mid \sigma \in \Sigma\}.$$

Let h be the homomorphism $h : \widehat{\Sigma} \rightarrow \Sigma$ given by $h(\sigma^{(z)}) = \sigma$. The language

$$h^{-1}(L_1) = \{w \in \widehat{\Sigma}^* \mid h(w) \in L_1\}$$

is regular by the closure of regular languages to inverse homomorphisms. Let T be the language of the regular expression

$$\sum_{i=1}^y \left[\left(\left(\sum_{\sigma \in \Sigma} \sigma^{(1)} \right) \cdots \left(\sum_{\sigma \in \Sigma} \sigma^{(y)} \right) \right)^* \cdot \left(\sum_{\sigma \in \Sigma} \sigma^{(1)} \right) \cdots \left(\sum_{\sigma \in \Sigma} \sigma^{(i)} \right) \right].$$

The language T consists of all words $w = w[1] \cdots w[n]$ over $\widehat{\Sigma}$ where for each $i \in [n]$, there exists $\sigma \in \Sigma$ such that $w[i] = \sigma^{(\ell)}$, where $i \equiv \ell \pmod{y}$. The language $h^{-1}(L_1) \cap T$ is regular by the closure of regular languages to intersection. For each letter $\sigma^{(i)} \in \widehat{\Sigma}$, let $c_{\sigma^{(i)}}(x)$ be equal to the polynomial $p_{i,\sigma}(x)$.

From Equation (7) it follows that

$$a_{\mathcal{D}}(n) = \sum_{w \in h^{-1}(L_1) \cap T, |w|=n} \prod_{j=1}^n c_{w[j]}(j),$$

and the lemma follows.

5.3. Proof of Lemma 14

In this subsection we prove Lemma 14, which states:

Let $\Sigma = \{\sigma_1, \dots, \sigma_{\xi}\}$ be an alphabet and let $\alpha_s(x)$ be polynomials in $\mathbb{Z}[x]$ for each $s \in \Sigma$. If $a(n)$ has a PW-interpretation for $\bar{\alpha} = (\alpha_s : s \in \Sigma)$, then there exist two classes C_1 and C_2 of sparse diagonal τ_k -structures which are both **MSOL**-expressible such that $a(n) = a_{C_1}(n) - a_{C_2}(n)$.

The PW-interpretation is a sum of positive and negative integers, each of them obtained as a product. The general idea of the proof is to partition the integers so that the positive integers are counted by $a_{C_1}(n)$ and the negative by $a_{C_2}(n)$.

Let $\Sigma = \{\sigma_1, \dots, \sigma_{\xi}\}$ and let L_1 be a regular language over Σ such that $a(n) = d_{L_1, \bar{\alpha}}(n)$. Let $\beta_{\sigma_1}(x), \dots, \beta_{\sigma_{\xi}}(x) \in \mathbb{N}[x]$ and $\gamma_{\sigma_1}(x), \dots, \gamma_{\sigma_{\xi}}(x) \in \mathbb{N}[x]$ be polynomials over \mathbb{N} such that for each i , $\alpha_{\sigma_i}(x) = \beta_{\sigma_i}(x) - \gamma_{\sigma_i}(x)$. Let $\tilde{\Sigma} = \{\tilde{\sigma} \mid \sigma \in \Sigma\}$ and let $h : \Sigma \cup \tilde{\Sigma} \rightarrow \Sigma$ given by $h(\sigma) = \sigma$ and $h(\tilde{\sigma}) = \sigma$ for each $\sigma \in \Sigma$. Let L_2 be the language over $\Sigma \cup \tilde{\Sigma}$ obtained from L_1 as follows:

$$L_2 = \{w \mid h(w) \in L_1\}.$$

The language L_2 is regular by the closure of regular languages under inverse homomorphism. Thus, we may write $a(n)$ as

$$a(n) = \sum_{w \in L_2, |w|=n} \prod_{j:w[j] \in \Sigma} \beta_{w[j]}(j) \prod_{j:w[j] \in \tilde{\Sigma}} (-\gamma_{w[j]}(j)).$$

where the products range over all $j \in [n]$ where $w[j] \in \Sigma$ or $w[j] \in \tilde{\Sigma}$. Let L_{even} be the set of words $w \in (\Sigma \cup \tilde{\Sigma})^*$ such that the number of letters from $\tilde{\Sigma}$ in w is even. The language L_{odd} is defined similarly. Both L_{even} and L_{odd} are regular. So, $L_{2,\text{even}} = L_2 \cap L_{\text{even}}$ and $L_{2,\text{odd}} = L_2 \cap L_{\text{odd}}$ are regular by the closure of regular languages under intersection. Denote for any regular language L over $\Sigma \cup \tilde{\Sigma}$,

$$b_L(n) = \sum_{w \in L, |w|=n} \prod_{j: w[j] \in \Sigma} \beta_{w[j]}(j) \prod_{j: w[j] \in \tilde{\Sigma}} \gamma_{w[j]}(j) \quad (8)$$

It holds that $a(n)$ is the difference $a(n) = b_{L_{2,\text{even}}}(n) - b_{L_{2,\text{odd}}}(n)$. It remains to show that if L is regular then there exists a class of sparse diagonal structures \mathcal{C} which is **MSOL**-expressible and $b_L(n) = a_{\mathcal{C}}(n)$.

To do so, we first prove that for every polynomial $p(x) \in \mathbb{N}[x]$ with coefficients in \mathbb{N} there exists a bounded regular language S of a special type such that $a_S(n) = p(n)$. We say a bounded regular language S is *simple* if there exist letters t_1, \dots, t_e such that $S \subseteq t_1^* \cdots t_e^*$. Assume $p(n) = a_{S_1}(n) + a_{S_2}(n)$, where S_1 and S_2 are simple bounded regular languages, such that $S_1 \subseteq c_1^* \cdots c_{\mu_1}^*$ and $S_2 \subseteq d_1^* \cdots d_{\mu_2}^*$. We may assume $\{c_1, \dots, c_{\mu_1}\} \cap \{d_1, \dots, d_{\mu_2}\} = \emptyset$ and hence S_1 and S_2 are disjoint. It holds that $S_1 \cup S_2 \subseteq c_1^* \cdots c_{\mu_1}^* \cdot d_1^* \cdots d_{\mu_2}^*$ and $p(n) = a_{S_1 \cup S_2}(n)$. Now assume $p(n) = n^r$. Let $\Sigma_r = \wp([r])$ be the power set of $[r]$. Let S be the set of words $w \in \Sigma_r^*$ such that for every $i \neq j$, $w[i] \cap w[j] = \emptyset$ and $w[1] \cup \dots \cup w[n] = [r]$, where $n = |w|$. It is not hard to see that S is regular and that $a_S(n) = n^r$. Since S is sparse and regular, it is bounded. In fact, S is the disjoint union of a finite number of simple bounded regular languages with regular expressions of the form $\theta^* e_1 \theta^* \cdots \theta^* e_\eta \theta^*$, where e_1, \dots, e_η form a partition of $[r]$. This implies that there exists a language T which is a simple bounded regular language and for which $a_S(n) = a_T(n)$.

Hence, for every $p(x) \in \mathbb{N}[x]$ there exists a simple bounded regular language S with alphabet Γ such that $a_S(n) = p(n)$. Let $\Gamma' = \{\hat{\sigma} \mid \sigma \in \Gamma\}$ be a distinct copy of Γ . Let S' be the language over $\Gamma \cup \Gamma' \times \Gamma'$ given by

$$S' = \{\sigma_1 \cdots \sigma_{n-2} \cdot (\sigma_{n-1}, \hat{\sigma}_n) \mid \sigma_1 \cdots \sigma_n \in S\}.$$

Since S is regular, so is S' . It holds that $a_{S'}(n-1) = p(n)$ if $n \geq 2$.

Returning to Equation (8), for each $f \in \{\beta_\sigma, \gamma_\sigma \mid \sigma \in \Sigma\}$ let S_f be a simple bounded regular language such that $a_{S_f}(n-1) = f(n)$. We may assume that the languages S_f are over alphabets $\Gamma_f = \{\gamma_1, \dots, \gamma_{\theta_f}\}$ respectively, which are pairwise disjoint and each is also disjoint from the alphabet of L , Σ . We may assume $S_f \subseteq \gamma_1^* \cdots \gamma_{\theta_f}^*$. For each S_{β_σ} and S_{γ_σ} let $T_{\beta_\sigma} = \sigma \cdot S_{\beta_\sigma}$ and $T_{\gamma_\sigma} = \sigma \cdot S_{\gamma_\sigma}$ respectively. It holds that for each $f \in \{\beta_\sigma, \gamma_\sigma \mid \sigma \in \Sigma\}$, $a_{T_f}(n) = f(n)$. Notice the languages T_f are pairwise disjoint since their alphabets are disjoint.

Let $k = |\Sigma| + \sum_{\sigma \in \Sigma} |\Gamma_{\beta_\sigma}| + |\Gamma_{\gamma_\sigma}|$. Let g be a bijection from the alphabet $\Delta = \Sigma \cup \bigcup_{\sigma \in \Sigma} (\Gamma_{\beta_\sigma} \cup \Gamma_{\gamma_\sigma})$ to $[k]$ which preserves the order with respect to each Γ_f and for which $\{g(1), \dots, g(|\Sigma|)\} = \Sigma$. Let $T = \bigcup_{\sigma \in \Sigma} T_{\beta_\sigma} \cup \bigcup_{\sigma \in \Sigma} T_{\gamma_\sigma}$. The language T is a simple bounded regular language satisfying

$$T \subseteq g(1)^* \cdots g(k)^*.$$

Let \mathcal{U} be the class of all tuples t for which there exists n such that $t = (w, u_1, \dots, u_n) \in L \times \Delta \times \dots \times \Delta^n$, $|w| = n$ and for each i , $u_i[1] = w[i]$, $|u_i| = i$ and $u_i \in T$. The sequence $b_L(n)$ is the number of tuples (w, u_1, \dots, u_n) in \mathcal{U} , $b_L(n) = a_{\mathcal{U}}(n)$.

By Theorem 4, there exist **MSOL**(ρ_t) sentences, ϕ and ψ , such that ϕ defines L and ψ defines T . Let \mathcal{D} be the following class of τ_k -structures. The class \mathcal{D} consists of all structures

$$\mathcal{A}_{(w, u_1, \dots, u_{|w|})} = \langle [n], <_{\text{nat}}, R_1, \dots, R_k \rangle$$

such that

- (w, u_1, \dots, u_n) belongs to \mathcal{U} ,
- $(i, j) \in R_\ell$ iff $i \leq j$ and $g(u_j[i]) = \ell$.

The class \mathcal{D} is a class of diagonal sparse structures which is **MSOL**-expressible by ϕ and ψ and $a_{\mathcal{D}}(n) = a(n)$, as required.

6. Discussion and Open Problems

We studied combinatorial interpretations of counting functions by counting the number of relations definable in **MSOL** over linear orders on $[n]$.

We proved an analog of the Chomsky-Schützenberger Theorem for holonomic and q -holonomic sequences using positionally weighted structures with unary predicates (words), and for holonomic sequences also by counting binary relations, but without weights.

Some holonomic sequences can also be represented by counting relations over a fixed set (without a linear order). For example, the factorial sequence counts the number of linear orders on $[n]$, and the derangement numbers count the number of permutations without fixed points. Both these properties are definable in First Order Logic, and hence in **MSOL**. It follows from the results in [BS81, Spe88] that the central binomial coefficient $\binom{2n}{n}$ has no **MSOL**-definable combinatorial interpretation counting binary relations on $[n]$ without the order. In [BS81, Spe88] the focus is on modular recurrence relations, and it is noted that every SP-recursive sequence satisfies for every modulus m a linear recurrence relation with constant coefficients in \mathbb{Z}_m . The converse is not true because there are only countably many SP-recursive sequences, but uncountably many sequences satisfying linear modular recurrence relations with constant coefficients.

In [Spe88], Specker asks whether, given a sequence $a(n)$ with a combinatorial interpretation, there exists a definability criterion for the combinatorial interpretation which ensures that $a(n)$ is SP-recursive, and hence satisfies linear modular recurrence relations.

Specker's question leaves open whether we count relations on $[n]$ with or without fixed order. From the context of his paper it seems that he is interested mostly in the case without order.

This leaves us with some interesting open problems:

Problem 1 (Specker's Problem). *Is there a sufficient definability condition which ensures that integer sequences with a combinatorial interpretation over $[n]$ without order are SP-recursive?*

In [BS81, Spe88] a sufficient condition is given for $a(n)$ to satisfy linear modular recurrence relations, namely that the combinatorial interpretation is definable in Monadic Second Order Logic over $[n]$ without order.

Problem 2. *Does every holonomic sequence $a(n)$ of non-negative integers have a combinatorial interpretation definable in Second Order Logic over $[n]$ without order?*

The sequence of primes p_n is neither holonomic [FGS05] nor does it satisfy any modular linear recurrence relation, [Shi00].

Problem 3. *Does the sequence of primes p_n have a combinatorial interpretation over $[n]$, or over $[n]$ with order, definable in Second Order Logic?*

Note that it follows from [BS81, Spe88] that no combinatorial interpretation of the primes exists which is definable in Monadic Second Order Logic on $[n]$ without order and counts binary relations.

In [KM10] we have given a positive answer to Specker’s question in the case with order. However, the definability criterion given in [KM10] restricts the combinatorial interpretation to a very special class of lattice paths. Our Theorem 15 gives a more general definability criterion. However, in Lemma 13 the polynomials have rational coefficients. This only gives a characterization of P-recursive sequences where the leading polynomial is a constant $c \in \mathbb{N}$. Such sequences still satisfy an infinite set of modular linear recurrence relations, namely for each prime $p \geq c + 1$.

Problem 4. *Does every holonomic sequence $a(n)$ of non-negative integers which satisfies a linear recurrence $c \cdot a(n + r) = \sum_{i=0}^{r-1} p_i(n) \cdot a(n + i)$ with $c \in \mathbb{N}$, $c \neq 0$ satisfy linear modular recurrence relations for every modulus m ?*

Acknowledgements

We are indebted to I. Gessel, E. Fischer and E.V. Ravve for useful remarks. We would like to thank the anonymous referees of earlier conference versions of this paper for their encouraging remarks and extremely valuable suggestions, which have been incorporated into this new version.

References

- [AMS⁺97] S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. Gapped blast and psi-blast: a new generation of protein database search programs. *Nucleic Acids Res.*, 25:3389–3402, 1997.
- [BCL⁺06] C. Borgs, J. Chayes, L. Lovász, V.T. Sós, and K. Vesztegombi. Counting graph homomorphisms. In M. Klazar, J. Kratochvil, M. Loeb, J. Matousek, R. Thomas, and P. Valtr, editors, *Topics in Discrete Mathematics*, pages 315–371. Springer, 2006.
- [Ber09] F. Bergeron. *Algebraic combinatorics and coinvariant spaces*. CMS Treatises in Mathematics. Canadian Mathematical Society, 2009.
- [Bir30] G.D. Birkhoff. General theory of irregular difference equations. *Acta Mathematica*, 54:205–246, 1930.
- [BS81] C. Blatter and E. Specker. Le nombre de structures finies d’une th’eorie à caractère fin. *Sciences Mathématiques, Fonds Nationale de la recherche Scientifique, Bruxelles*, pages 41–44, 1981.
- [BT33] G.D. Birkhoff and W. J. Trjitzinsky. Analytic theory of singular difference equations. *Acta Mathematica*, 60:1–89, 1933.
- [CR92] William Y. C. Chen and Gian-Carlo Rota. q-analogs of the inclusion- exclusion principle and permutations with restricted position. *Discrete Mathematics*, 104(1), 1992.
- [CS63] N. Chomsky and M.P. Schützenberger. The algebraic theory of context free languages. In P. Brafford and D. Hirschberg, editors, *Computer Programming and Formal Systems*, pages 118–161. North Holland, 1963.
- [DIV06] Flavio D’Alessandro, Benedetto Intrigila, and Stefano Varricchio. On the structure of the counting function of sparse context-free languages. *Theor. Comput. Sci.*, 356(1-2):104–117, 2006.
- [EF95] H.D. Ebbinghaus and J. Flum. *Finite Model Theory*. Perspectives in Mathematical Logic. Springer, 1995.
- [EFT94] H.D. Ebbinghaus, J. Flum, and W. Thomas. *Mathematical Logic, 2nd edition*. Undergraduate Texts in Mathematics. Springer-Verlag, 1994.
- [EvPSW03] G. Everest, A. van Porten, I. Shparlinski, and T. Ward. *Recurrence Sequences*. Mathematical Surveys and Monographs, vol. 104. American Mathematical Society, 2003.
- [FGS05] P. Flajolet, S. Gerhold, and B. Salvy. On the non-holonomic character of logarithms, powers and the nth prime function. *Electron. J. Combin.*, 11:1–16, 2005.
- [FS09] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.
- [Ger04] S. Gerhold. On some non-holonomic sequences. *Electronic Journal of Combinatorics*, 11:1–7, 2004.
- [GKRS08] Pawel Gawrychowski, Dalia Krieger, Narad Rampersad, and Jeffrey Shallit. Finding the growth rate of a regular of context-free language in polynomial time. In *Developments in Language Theory*, pages 339–358, 2008.
- [KM10] T. Kotek and J.A. Makowsky. A representation theorem for holonomic sequences based on counting lattice paths. submitted, 2010.

- [Koo93] Tom H. Koornwinder. On Zeilberger's algorithm and its q -analogue. *Journal of Computational and Applied Mathematics*, 48:91–111, 1993.
- [Lib04] L. Libkin. *Elements of Finite Model Theory*. Springer, 2004.
- [NZ99] J. Noonan and D. Zeilberger. The Goulden-Jackson cluster method: Extensions, applications and implementations. *J. Differ. Equations Appl.*, 5 (4-5):355–377, 1999.
- [PWZ96] M. Petkovsek, H. Wilf, and D. Zeilberger. *A=B*. AK Peters, 1996.
- [Shi00] D.K.L. Shiu. Strings of congruent primes. *J. London Math. Soc.*, 61:359–373, 2000.
- [Spe88] E. Specker. Application of logic and combinatorics to enumeration problems. In E. Börger, editor, *Trends in Theoretical Computer Science*, pages 141–169. Computer Science Press, 1988. Reprinted in: Ernst Specker, *Selecta*, Birkhäuser 1990, pp. 324–350.
- [SS78] A. Salomaa and M. Soittola. *Automata-Theoretic Aspects of Formal Power Series*. Texts and Monographs in Computer Science. Springer, 1978.
- [SSGE82] G D Stormo, T D Schneider, L Gold, and A Ehrenfeucht. Use of the 'perceptron' algorithm to distinguish translational initiation sites in e. coli. *Nucleic Acid Research*, 10:2997–3012, 1982.
- [Sta80] R. P. Stanley. Differentiably finite power series. *European Journal of Combinatorics*, 1:175–188, 1980.
- [Tro81] V.I. Trofimov. Growth functions of some classes of languages. *Cybernetics*, 6:9–12, 1981.
- [WZ92] H. S. Wilf and D. Zeilberger. An algorithmic proof theory for hypergeometric (ordinary and “ q ”) multi-sum/integral identities. *Inventiones math.*, 108:557–633, 1992.
- [Zei90] D. Zeilberger. A holonomic systems approach to special functions identities. *J. of Computational and Applied Mathematics*, 32:321–368, 1990.